

Franc Janžekovič



MAKROEKOLOGIJA

Analiza biodiverzitetnih podatkov



Univerzitetna založba
Univerze v Mariboru



Univerza v Mariboru

Fakulteta za naravoslovje
in matematiko

MAKROEKOLOGIJA

Analiza biodiverzitetnih podatkov

Avtor

Franc Janžekovič

April 2023

| | | | |
|--|---|--|--|
| Naslov <i>Title</i> | Makroekologija <i>Macroecology</i> | | |
| Podnaslov <i>Subtitle</i> | Analiza biodiverzitetnih podatkov <i>Analysis of Biodiversity Data</i> | | |
| Avtor <i>Author</i> | Franc Janžekovič (Univerza v Mariboru, Fakulteta za naravoslovje in matematiko) | | |
| Recenzija <i>Review</i> | Dominik Benkovič (Univerza v Mariboru, Fakulteta za naravoslovje in matematiko) | Damijan Denac (Društvo za opazovanje in proučevanje ptic Slovenije) | |
| Jezikovni pregled <i>Language editing</i> | Mojca Garantini | | |
| Tehnični urednik <i>Technical editor</i> | Jan Perša (Univerza v Mariboru, Univerzitetna založba) | Oblikovanje ovitka <i>Cover designer</i> | Jan Perša (Univerza v Mariboru, Univerzitetna založba) |
| Grafike na ovitku in poglavjih <i>Cover graphics</i> | Maša Miklavc, 2023 | | Grafične priloge <i>Graphic material</i> |
| Avtorji fotografij <i>Photo authors</i> | Gregor Bernard: 2.1, 2.3, 2.6, 2.9, 2.11, 15.2, 15.9, 15.11, 15.12, 17.15; Marijan Govedič: 13.3.; Eva Horvat: 4.1.; Predrag Jakšić: 4.12.; Franc Janžekovič: 4.13, 9.4, 10.4, 10.5, 10.8, 11.5, 16.4, 17.2, 17.13.; Dušan Klenovšek: 7.1.; Peter Kozel: 3.3, 15.4.; Alenka Kryštufek: 17.5.; Boris Kryštufek: 13.8.; Nedko Nedvalkov: 12.1.; Tone Novak: 3.2.; Alen Ploj: 4.1. | | |
| Založnik <i>Published by</i> | Univerza v Mariboru Univerzitetna založba Sloški trg 15, 2000 Maribor, Slovenija https://press.um.si , zalozba@um.si | Izdajatelj <i>Issued by</i> | Univerza v Mariboru Fakulteta za naravoslovje in matematiko Koroška cesta 160, 2000 Maribor, Slovenija https://www.fnm.um.si , fnm@um.si |
| Izdaja <i>Edition</i> | Prva izdaja | Izdano <i>Published at</i> | Maribor, april 2023 |
| Vrsta publikacije <i>Publication type</i> | E-knjiga | Dostopno na <i>Available at</i> | https://press.um.si/index.php/ump/catalog/book/734 |

CIP - Kataložni zapis o publikaciji
Univerzitetna knjižnica Maribor

574.1/.3.087 (075.8) (0.034.2)

JANŽEKOVIČ, Franc, 1962-
Makroekologija [Elektronski vir] : analiza
biodiverzitetnih podatkov / avtor Franc
Janžekovič. - 1. izd. - E-učbenik. - Maribor :
Univerza v Mariboru, Univerzitetna založba,
2023

Način dostopa (URL) :
[https://press.um.si/index.php/ump/catalog/
book/734](https://press.um.si/index.php/ump/catalog/book/734)
ISBN 978-961-286-723-2 (PDF)
doi: 10.18690/um.fnm.2.2023
COBISS.SI-ID 147773443



© Univerza v Mariboru, Univerzitetna založba
/ University of Maribor, University Press

Besedilo / *Text* © Janžekovič, 2023

To delo je objavljeno pod licenco Creative Commons Priznanje avtorstva 4.0 Mednarodna. / *This work is licensed under the Creative Commons Attribution 4.0 International License.*

Uporabnikom je dovoljeno tako nekomercialno kot tudi komercialno reproduciranje, distribuiranje, dajanje v najem, javna priobčitev in predelava avtorskega dela, pod pogojem, da navedejo avtorja izvirnega dela.

Vsa gradiva tretjih oseb v tej knjigi so objavljena pod licenco Creative Commons, razen če to ni navedeno drugače. Če želite ponovno uporabiti gradivo tretjih oseb, ki ni zajeto v licenci Creative Commons, boste morali pridobiti dovoljenje neposredno od imetnika avtorskih pravic.

<https://creativecommons.org/licenses/by-nc-nd/4.0/>



Projekt sofinancirata Republika Slovenija, Ministrstvo za visoko šolstvo, znanost in inovacije, in Evropska unija – NextGenerationEU. Projekt se izvaja skladno z načrtom v okviru razvojnega področja Pametna, trajnostna in vključujoča rast, komponente Krepitev kompetenc, zlasti digitalnih in tistih, ki jih zahtevajo novi poklici in zeleni prehod (C3 K5), za ukrep investicija F. Izvajanje pilotnih projektov, katerih rezultati bodo podlaga za pripravo izhodišč za reformo visokega šolstva za zelen in odporen prehod v Družbo 5.0: projekt Pilotni projekti za prenovo visokega šolstva za zelen in odporen prehod. / *The project is co-financed by the Republic of Slovenia, the Ministry of Higher Education, Science and Innovation and the European Union – NextGenerationEU. The project is implemented in accordance with the Recovery and Resilience Plan (RRP) under the development area Smart, Sustainable and Inclusive Growth, component Strengthening competencies, in particular digital competencies and those required by the new professions and the green transition (C3 K5), for the investment measure Investment F. Implementation of pilot projects, the results of which will serve as a basis for the preparation of a roadmap for the reform of higher education for a green and resilient transition to a Society 5.0: project Pilot Projects for the Reform of Higher Education for a Green and Resilient Transition.*

ISBN 978-961-286-723-2 (pdf)
978-961-286-724-9 (trda vezava)

DOI <https://doi.org/10.18690/um.fnm.2.2023>

Cena
Price Brezplačni izvod

Odgovorna oseba založnika
For publisher prof. dr. Zdravko Kačič,
rektor Univerze v Mariboru

Citiranje
Attribution Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

Zahvala

V nastajanje knjige o analizi biodiverzitetnih podatkov je bilo tako ali drugače vključenih veliko oseb, fizičnih in pravnih iz službenega in zasebnega življenja.

Zahvalo začentam z družino – zahvaljujem se ženi Boži in otrokom Meti, Ani in Nejcu, ki mi dajejo zavetje doma, človeško toplino ter moč za delo in kreativnost. Prav tako hvala širši družini, ki me spodbuja, verjame vame in daje potrebno energijo za ustvarjanje: mami Marjeti ter sestram Eriki, Marjeti in Ani, ter bratoma Jožetu in Janezu in njihovim najdražjim. Globoko sem hvaležen prijateljem in sodelavcem za pomoč pri reševanju strokovnih dilem in njihovim prispevkom h končni obliki vsebine. Hvala prof. dr. Tonetu Novaku iz Fakultete za naravoslovje in matematiko Univerze v Mariboru, »najzahtevnejši bralec« in »prijateljski recenzent«, prijatelj in sodelavec Tone je zaslužen, da so misli natančneje izražene, dejstva pa korektno zapisana. Ob pripravi številnih znanstvenih člankov sva vedno znova ugotavljala, da je potrebno izvesti vse vidike statističnih vrednotenj, da se izkaže, kateri je dejansko najbližji za razlago dogajanja v naravi. Hvala prof. dr. Borisu Kryštufku iz Prirodoslovnega muzeja Slovenije, na »njegova ramena sem stopil«, da sem se lahko razgledal po goščavi biodiverzitetnih metodologij in z njegovim vodenjem prestopil visoke ovire. V dolgoletnem sodelovanju in prijateljevanju je Boris prispeval ideje in material za znanstvene objave, ter me spodbujal k učenju novih analitskih pristopov in osvajanju novih statističnih orodij. Z Borisom vsa nabrala nekaj čudovitih terenskih izkušenj, kjer sva prvi korak raziskovalnega dela – to je idejo, nadgradila z zbiranjem primarnih podatkov na terenu. Hvala Marijanu Govediču in Mladenu Kotarcu iz Centra za kartografijo

favne in flore, Marijan mi je odpiral nove perspektive organizacije in priprave biodiverzitetnih podatkov, prav tako je bistveno prispeval k izboljšanju teksta. Mladen je v številnih pogovorih odpiral nove možnosti in pristope k razumevanju in vrednotenju biodiverzitet tako v analitskem kot v družbenem kontekstu. Hvala sodelavcem prof. dr. Dušanu Devetaku, doc. dr. Tini Klenovšek, doc. dr. Vesni Klokočovnik, doc. dr. Janu Podlesniku, prof. dr. Saški Lipovšek in doc. dr. Petru Kozelu iz Fakultete za naravoslovje in matematiko Univerze v Mariboru. Z Dušanom, Tino, Vesno in Janom smo organizirali devet nepozabnih zooloških ekspedicij na področje Balkana in jih nadgradili v številne znanstvene članke, v katerih obravnavamo različne vidike biodiverzitet. Petru hvala za pogovore in pomoč pri usvajanju novega statističnega znanja, Saški hvala za vpogled v subcelično pestrost. Hvala prof. dr. Petru Trontlju iz Biotehniške fakultete Univerze v Ljubljani in prof. dr. Eleni Bužan iz FAMNIT iz Univerze na Primorskem, ki sta izboljšala zgodnje verzije nekaterih poglavij. Hvala recenzentoma prof. dr. Dominiku Benkoviču iz Oddelka za matematiko in računalništvo na Fakulteti za naravoslovje in matematiko Univerze v Mariboru in doc. dr. Damijanu Dencu iz Društva za opazovanje in proučevanje ptic Slovenije DOPPS. Dominik in Damijan sta vsak iz vidika svoje stroke, matematike in statistike ter biologije in biodiverzitet prispevala svoj delež, da je vsebina konsistentna in strokovno neoporečna. Hvala lektorici Mojci Garantini za skrben pregled in popravke besedila. Hvala Janu Perši urednik na Univerzitetni založbi Univerze v Mariboru, Jan je poskrbel za oblikovanje končne verzije knjige, ki jo imate pred seboj. Napake, ki so preživele vse skrbne bralce, so posledica samo mojega

premalo zavzetega dela. Hvaležen vam bom, da napake ali predloge za izboljšanje posredujete na moj naslov.

Posebno zahvalo namenjam Maši Miklavc. Maša je mlada umetnica, slikarka na začetku kariere. Prepričala me je s svojim talentom in zavzetostjo za delo. Maša je izdelala risbe za naslovnico knjige in za začetke poglavij. Pristrčna je bila njena ugotovitev, več energije in dela kot vložiti v risbo, boljši je izdelek!

Na koncu se želim prikloniti tudi organizacijam, pravzaprav ljudem, ki delujejo v njih, v katerih tudi sam delujem, ali s katerimi sem sodeloval, ter so tako ali drugače pomembno prispevale k razvoju in nastanku knjige o Makroekologiji oz. analizi biodiverzitetnih podatkov. Fakulteta za naravoslovje in matematiko Univerze v Mariboru, kjer sem dobil ogromno znanja, je

ustanova, ki mi je dajala formalni okvir za ustvarjanje in kjer sem prebil večino ustvarjalnega obdobja. Zahvala Biotehniški fakulteti Univerze v Ljubljani, kjer sem dobil večino znanja; Prirodoslovnemu muzeju Slovenije, ki je »arhiv« prirodnin in materialno izhodišče ter navdih za nadaljnje raziskave biodiverzitete; Društvu za opazovanje in proučevanje ptic Slovenije, ki mi je omogočalo strokovno rast, me povežalo s čudovitimi ljudmi in mi odprlo oči naravovarstvene problematike; Centru za kartografijo favne in flore, kjer so me uvajali v organizacijo prostorskih in biodiverzitetnih podatkov; Društvu biologov Slovenije za izmenjavo strokovnih pogledov; Javni agenciji za raziskovalno dejavnost Republike Slovenije, ki je sicer mačehovsko, pa vendarle finančno omogočila izvedbo nekaterih raziskav in vzorčenje biodiverzitete na terenu.

Franc Janžekovič

Kazalo

| | |
|---|------------|
| Predgovor | 1 |
| 1. poglavje: MAKROEKOLOGIJA IN BIODIVERZITETA | 5 |
| 1.1 Makroekologija..... | 7 |
| 1.2 Biodiverzitetna..... | 8 |
| Zaključki | 13 |
| 2. poglavje: TELESNA VELIKOST | 15 |
| 2.1 Frekvenčna porazdelitev telesne velikosti..... | 20 |
| 2.2 Odnos med telesno velikostjo in abundanco | 21 |
| Zaključki | 28 |
| 3. poglavje: VRSTNO BOGASTVO | 31 |
| 3.1 Prostorska komponenta vrstnega bogastva..... | 34 |
| 3.1.1 Alfa diverzitetna in gama diverzitetna..... | 36 |
| 3.1.2 Beta diverzitetna..... | 37 |
| 3.1.3 Zeta diverzitetna..... | 44 |
| 3.2 Ocena vrstnega bogastva..... | 51 |
| 3.3 Redčenje..... | 57 |
| 3.4 Temna diverzitetna | 61 |
| 3.5 Prikrita diverzitetna..... | 65 |
| Zaključki | 66 |
| 4. poglavje: VRSTNA PESTROST | 69 |
| 4.1 Cenilke vrstne pestrosti..... | 72 |
| 4.2 Hillova diverzitetna števila in prava diverzitetna..... | 80 |
| 4.3 Vrstna poravnost | 82 |
| 4.4 Primerjave vrstne pestrosti med združbami | 84 |
| 4.5 Relativna vrstna abundanca..... | 91 |
| Zaključki | 100 |
| 5. poglavje: FUNKCIONALNA PESTROST | 103 |
| 5.1 Na surovih podatkih temelječe analize..... | 109 |
| 5.2 Na razdaljah temelječe analize | 113 |
| 5.2.1 Funkcionalna razpršenost FDis..... | 113 |
| 5.3 Na poligonih temelječe analize | 126 |
| 5.3.1 Funkcionalno razhajanje FDiv | 126 |
| 5.3.2 Funkcionalno bogastvo FRic..... | 129 |
| 5.3.3 Funkcionalna poravnost FEve | 131 |
| Zaključki | 134 |

| | |
|---|------------|
| 6. poglavje: FILOGENETSKA PESTROST | 137 |
| 6.1 Filogenetska pestrost..... | 139 |
| 6.1.1 Dimenzija bogastva | 140 |
| 6.1.2 Dimenzija raznolikost..... | 141 |
| 6.1.3 Dimenzija poravnosti..... | 142 |
| 6.2 Taksonomska pestrost | 144 |
| Zaključki | 151 |
| | |
| 7. poglavje: ANALIZA EKOLOŠKIH ZDRUŽB | 153 |
| 7.1 Ekološka podobnost – Q analiza, primerjave med združbami..... | 160 |
| 7.1.1 Koeficienti podobnosti..... | 161 |
| 7.1.2 Koeficienti razdalj..... | 167 |
| 7.2 Ekološka podobnost – R analiza – primerjave med vrstami..... | 172 |
| 7.2.1 Podatki o prisotnosti oz. odsotnosti vrst..... | 172 |
| 7.2.2 Podatki o abundanci vrst..... | 173 |
| 7.3 Problem dvojne ničle in paradoks abundance..... | 176 |
| 7.4 Transformacije podatkov o sestavi združb..... | 178 |
| 7.5 Predstavitev odnosov med združbami | 180 |
| 7.5.1 Razvrščanje v skupine..... | 181 |
| 7.5.2 Multidimenzionalno skaliranje..... | 185 |
| Zaključki | 189 |
| | |
| 8. poglavje: ODNOS MED VRSTNIM BOGASTVOM IN POVRŠINO..... | 191 |
| Zaključki | 203 |
| | |
| 9. poglavje: VRSTNO BOGASTVO V GRADIENTU GEOGRAFSKE ŠIRINE | 205 |
| Zaključki | 224 |
| | |
| 10. poglavje: VRSTNO BOGASTVO V GRADIENTU NADMORSKE VIŠINE..... | 227 |
| Zaključki | 240 |
| | |
| 11. poglavje: FENOTIPSKA PLASTIČNOST | 243 |
| 11.1 Opis morfotipa in opisna statistika | 246 |
| 11.2 Analiza morfotipa..... | 246 |
| 11.2.1 Preverjanje razlik med skupinami – testna statistika | 247 |
| 11.2.2 Redukcija razsežnosti – analiza glavnih komponent | 247 |
| 11.3 Alometrija: statična, ontogenetska, evolucijska | 248 |
| 11.4 Fenotipska plastičnost..... | 249 |
| Zaključki | 273 |
| | |
| 12. poglavje: NUMERIČNA TAKSONOMIJA | 275 |
| 12.1 Analiza glavnih komponent | 280 |
| 12.2 Metoda voditeljev | 283 |
| 12.3 Diskriminantna analiza | 287 |
| Zaključki | 292 |
| | |
| 13. poglavje: FENETIKA IN KLADISTIKA..... | 293 |
| 13.1 Fenetika..... | 297 |
| 13.1.1 Fenetska analiza na kvalitativnih morfoloških znakih..... | 298 |
| 13.1.2 Fenetska analiza na kvantitativnih morfoloških znakih | 306 |
| 13.1.3 Tradicionalna morfometrija in izločitev velikosti | 308 |
| 13.1.4 Geometrijska morfometrija..... | 315 |
| 13.1.5 Metoda združevanja sosedov | 319 |

| | | |
|---|---|-----|
| 13.1.6 | UPGMA metoda | 327 |
| 13.2 | Kladistika | 331 |
| 13.2.1 | Metoda največje varčnosti | 336 |
| Zaključki | | 340 |
| 14. poglavje: FILOGENETIKA | 343 | |
| 14.1 | Substitucijski modeli in genetske razdalje | 351 |
| 14.2 | Genske banke | 360 |
| 14.3 | Metoda največjega verjetja | 361 |
| 14.4 | Bayesovo sklepanje | 366 |
| 14.4.1 | Bayesov izrek | 367 |
| 14.4.2 | Monte Carlo z markovskimi verigami | 369 |
| 14.4.3 | Metropolis Monte Carlo z markovskimi verigami | 371 |
| 14.5 | Filogenetska drevesa | 371 |
| 14.6 | Vrednotenje dreves | 377 |
| 14.7 | Filogenetski signal | 385 |
| Zaključki | | 388 |
| 15. poglavje: MONITORING | 391 | |
| Zaključki | | 403 |
| 16. poglavje: VZORČENJE | 407 | |
| 16.1 | Vzorec | 413 |
| 16.2 | Način izbire lokalitet v biodiverzitetni raziskavi | 413 |
| 16.3 | Načrt vzorčenja | 414 |
| 16.4 | Organizacija podatkov | 418 |
| 16.5 | Politike odprte znanosti in ravnanje z raziskovalnimi podatki | 420 |
| 16.5.1 | Tipi in formati podatkov | 421 |
| 16.5.2 | Načelo FAIR | 422 |
| 16.6 | Zemljevid | 423 |
| 16.7 | Naravoslovne zbirke | 425 |
| Zaključki | | 426 |
| 17. poglavje: RAZLAGANJE STRUKTURE BIODIVERZITETNIH PODATKOV | 429 | |
| 17.1 | Enostavna linearna regresija | 432 |
| 17.1.1 | Enostavna linearna regresija – Model I | 434 |
| 17.1.2 | Enostavna linearna regresija – Model II | 437 |
| 17.2 | Multipla linearna regresija | 442 |
| 17.3 | Polinomska regresija | 456 |
| 17.4 | Primerjava matrik Mantelov test | 458 |
| Zaključki | | 460 |
| 18. poglavje: STATISTIČNO SKLEPANJE | 463 | |
| 18.1 | Velikost vzorca | 469 |
| 18.1.1 | Velikost vzorca za oceno vrstnega bogastva | 469 |
| 18.1.2 | Velikost vzorca za oceno statističnega parametra | 470 |
| 18.2 | Velikost učinka | 473 |
| 18.2.1 | Mere standardiziranih razlik | 474 |
| 18.2.2 | Mere povezanosti | 475 |
| 18.3 | Analiza moči | 476 |
| Zaključki | | 480 |

| | |
|----------------------------|------------|
| DODATKI | 483 |
| Stvarno kazalo..... | 509 |

Predgovor

Makroekologija je veda o globalnih vzorcih in procesih biodiverzitete v prostoru in času v velikem merilu ter o odzivu biodiverzitete na spreminjajoče se okolje. *Biodiverziteteta* je najvrednejša lastnost Zemlje, je bogastvo in pestrost živih bitij. Raziskovalci na področju makroekologije za razlago številčnosti, porazdelitve in variabilnosti biodiverzitete uporabljamo statistično in numerično metodologijo in terminologijo.

Pomen in zanimanje javnosti je biodiverziteteta začela dobivati s prebujanjem naravovarstvene zavesti v šestdesetih in dobila pravi zagon v sedemdesetih letih 20. stoletja. Mejniki za bistveno večjo družbeno čuječnost za ekološka vprašanja je postavila Rachel Carson z leta 1962 izdano knjigo *Nema pomlad*. Od osemdesetih let naprej se je naravovarstveno gibanje, z opozarjanjem na okoljske netrajnostne prakse in ekološke nesreče, samo še stopnjevalo. Leta 1980 je izšla ena od temeljnih monografij *Conservation biology*, z njo je avtor Thomas Lovejoy utemeljil področje *varstvene biologije* kot znanstveno disciplino in biodiverziteteto uvedel v področje znanosti.

Besedno zvezo *biotska pestrost*, s katero označujemo pestrost in variabilnost živih bitij, je prvi zapisal Raymond F. Dasmann leta 1968. Termin je nastal iz potrebe po vrednotenju pestrosti naravnega okolja in živih bitij bodisi v lokalnem okolju bodisi globalno v hitro se razvijajoči zahodni družbi. Njena enobesedna različica *biodiverziteteta* je bila uvedena leta 1988 v zborniku z naslovom *Biodiversity*, objava tega zbornika je pomenila popularizacijo izraza v strokovni in laični javnosti.

Sledila so zaskrbljujoča strokovna in znanstvena poročila o zmanjševanju številčnosti populacij prostoživečih živali, izumiranju vrst, onesnaževanju okolja, ter degradaciji in uničevanju habitatov. Kopičenje znanstvenih spoznanj in visoka angažiranost civilne

družbe je svetovno politiko pripeljalo do sprejetja ključnega biodiverzitetnega dokumenta. Leta 1992 je Organizacija Združenih Narodov OZN sprejela *Konvencijo o biološki raznovrstnosti*, Slovenija je to mednarodno pogodbo ratificirala leta 1996. Konvencija je dobila razširitev v dveh dodatkih. *Kartagenski protokol* iz leta 2000, ki ureja promet oz. trgovanje z gensko spremenjenimi organizmi in tveganja pri uporabi gensko spremenjenih organizmov. Ter leta 2014 sprejeti *Nagojski protokol*, s katerim je urejena dostopnost in pravična delitev koristi iz genetskih virov. Aktualnost biodiverzitetne problematike se je stopnjevala v razglasitvi OZN leta 2010 za *Mednarodno leto biodiverzitetne* in obdobje 2011–2020 za *Desetletje biodiverzitetne*. 22. maj je razglašen za *Mednarodni dan biodiverzitetne*. Strokovna in znanstvena javnost je spoznala, da je Zemlja vstopila v proces šestega masovnega izumiranja vrst. Geološko obdobje, v katerem sedaj živimo, je poimenovano *antropocen*, zaradi posledic delovanja ljudi, ki ima planetarne razsežnosti in globalno vpliva na Zemljo. Biodiverzitetna tematika je uvrščena v ključne strateške in operativne dokumente Evropske unije EU, izpostavljamo *Strategijo EU za biotsko raznovrstnost*. Omeniti moramo tudi odziv EU na zdravstveno krizo COVID-19, ki je sprejela finančni mehanizem *Načrt za okrevanje in odpornost*, katerega sestavni del je tudi vpliv ukrepov na ohranjanje biodiverzitetne. Vse navedeno pomeni, da je biodiverzitetna kriza ob globalnih klimatskih spremembah ključni dejavnik zagotavljanja trajnostnega načina življenja in bivanja v globalnem planetarnem obsegu.

Ukrepi, s katerimi globalna družba, regionalni centri politične in vojaške moči, posamezne države in tudi lokalne skupnosti naslavljajo biodiverzitetno krizo, morajo imeti trdno oporo v biološki stroki in v znanosti različnih disciplin. Biodiverzitetna kriza je globalna in se dotika vseh vidikov našega delovanja in življenja, zato je nujen multidisciplinarni in interdisciplinarni pristop v iskanju rešitev in predlaganju ukrepov za njeno blaženje. V tej knjigi naslavljamo samo en vidik raziskav, to je področje zbiranja, organizacije, analiziranja in vrednotenja podatkov o biodiverziteti. Knjiga prinaša temeljne strokovne vsebine s področja metodologije ter numerične in statistične analize biotske pestrosti, tako na lokalnem kot v globalnem merilu. S to knjigo v slovenskem jeziku dobivamo strnjeni pregled znanja o analiziranju lokalnih in globalnih vzorcev in procesov biodiverzitetne v prostoru in času, to je tudi prispevek k razvoju novega znanstvenega področja imenovanega *makroekologija*.

Vsebina knjige je razdeljena v 18 poglavij. Obravnavam vse za današnji čas aktualne vidike v analizi biodiverzitetnih podatkov. V predstavitev sem vključil numerične in statistične postopke za analiziranje in vrednotenje biodiverzitetne. Definicija makroekologije in večplastnost koncepta biodiverzitetne ter predstavitev sodobnega pogleda na filogenijo živih bitij so predstavljeni v prvem poglavju. V drugem poglavju je v kontekstu biodiverzitetnih raziskav predstavljena telesna velikost. Tretje, četrto, peto in šesto poglavje sem namenil posameznim vidikom analize biodiverzitetne analizi vrstne, funkcionalne in filogenetske pestrosti. Analiza ekoloških združb je v posebnem, sedmem

poglavju. V poglavjih osem, devet in 10 sem predstavil ključne biodiverzitetne trende v prostoru, to je odnos med površino in biodiverzitetno, biodiverzitetna v gradientu geografske širine ter biodiverzitetna v gradientu nadmorske višine. V poglavju 11 naslavljam vzroke in analitske pristope fenotipske variabilnosti živali. Sledi poglavje 12 o numerični taksonomiji. Filogenetske rekonstrukcije so predstavljene v dveh poglavjih. V 13. obravnavam fenetski in kladistični pristop. Filogenetski pristop na dednini je predstavljen v poglavju 14. Poglavje 15 je namenjeno predstavitvi monitoringa, ključnega elementa v spremljanju sprememb biodiverzitetne v času. V 16. poglavju predstavljam načrtovanje vzorčenja biodiverzitetne in organizacijo biodiverzitetnih podatkov. Delo zaključujeta dve poglavji s statističnimi vsebinami, v njih sem predstavil metodološka orodja za analizo biodiverzitetnih trendov in pristop k statističnemu vrednotenju. V 17. poglavju predstavljam osnovne tipe regresijske analize, nepogrešljive v biodiverzitetnih raziskavah. V 18. poglavju sem predstavil metodološke zakonitosti statističnega sklepanja v raziskavah biodiverzitetne. Vseh razvijajočih se trendov v analizi biodiverzitetne se nisem dotaknil, prav tako nisem vključil vseh metodoloških pristopov. Knjiga je pisana na dveh nivojih, ki se izmenjujeta in dopolnjujeta. Prvi nivo je podajanje teoretičnih zakonitosti, razlaga pojmov in dejstev ter metodoloških pristopov za njihovo vrednotenje. Sledi drugi nivo, kjer na praktičnih primerih prikažem organizacijo podatkov, postopek računanja, ocenjevanja, interpretacije in predstavljanja rezultatov. Primere, s katerimi dopolnjujem teorijo in predstavljam analizo biodiverzitetnih podatkov sem jemal iz lastnih raziskovalnih izkušenj. Objekti, s katerimi so predstavljeni primeri, so iz živalskega sveta, večinoma vretenčarji, pa tudi suhe južine in metulji.

Knjiga je zasnovana kot univerzitetni učbenik, s pregledom temeljnih zakonitosti v obravnavi makroekoloških vsebin in analizi biodiverzitetne. Osnovni razmislek pri snovanju knjige je bil, kako razmejiti vsebino med podajanjem teorije in prikazom praktičnih rešitev. Predstavljeni postopki analiz segajo od elementarnih in najenostavnejših izračunov do najzahtevnejših postopkov analize biodiverzitetne, ki so razviti do današnjega časa. Praktični primeri analize so namenjeni prikazu uporabe teoretičnih zakonitosti. Angažirani bralec bo lahko hitreje napredoval s statističnimi postopki. Zato dokaj nazorno predstavljam kombiniranje različnih programskih orodij, začevši z MS Excel in Access kot temeljnima orodjema za organizacijo podatkov. Nadaljujem z uporabo splošnih statističnih programov, kot sta npr. SPSS in Statistica ter z uporabo specialnih programov za reševanje posameznih analiz, kot so npr. PAST4, MorphoJ in EstimateS. Večino rešitev, sploh najnaprednejših, sem predstavil s programskimi kodami v programskem okolju R.

Ideja za ta univerzitetni učbenik, predvsem njegova vsebina, nista nastali iz trenutnega navdiha, ideja je zorela, vsebina pa nastajala dobro desetletje. Vsebinska je rastla skupaj z razvojem študijskih programov biologije na Fakulteti za naravoslovje in matematiko

Univerze v Mariboru. Prečiščeno obliko vsebine sem oblikoval z objavami znanstvenih člankov v uglednih svetovnih revijah s področja biodiverzitete, zoologije in ekologije.

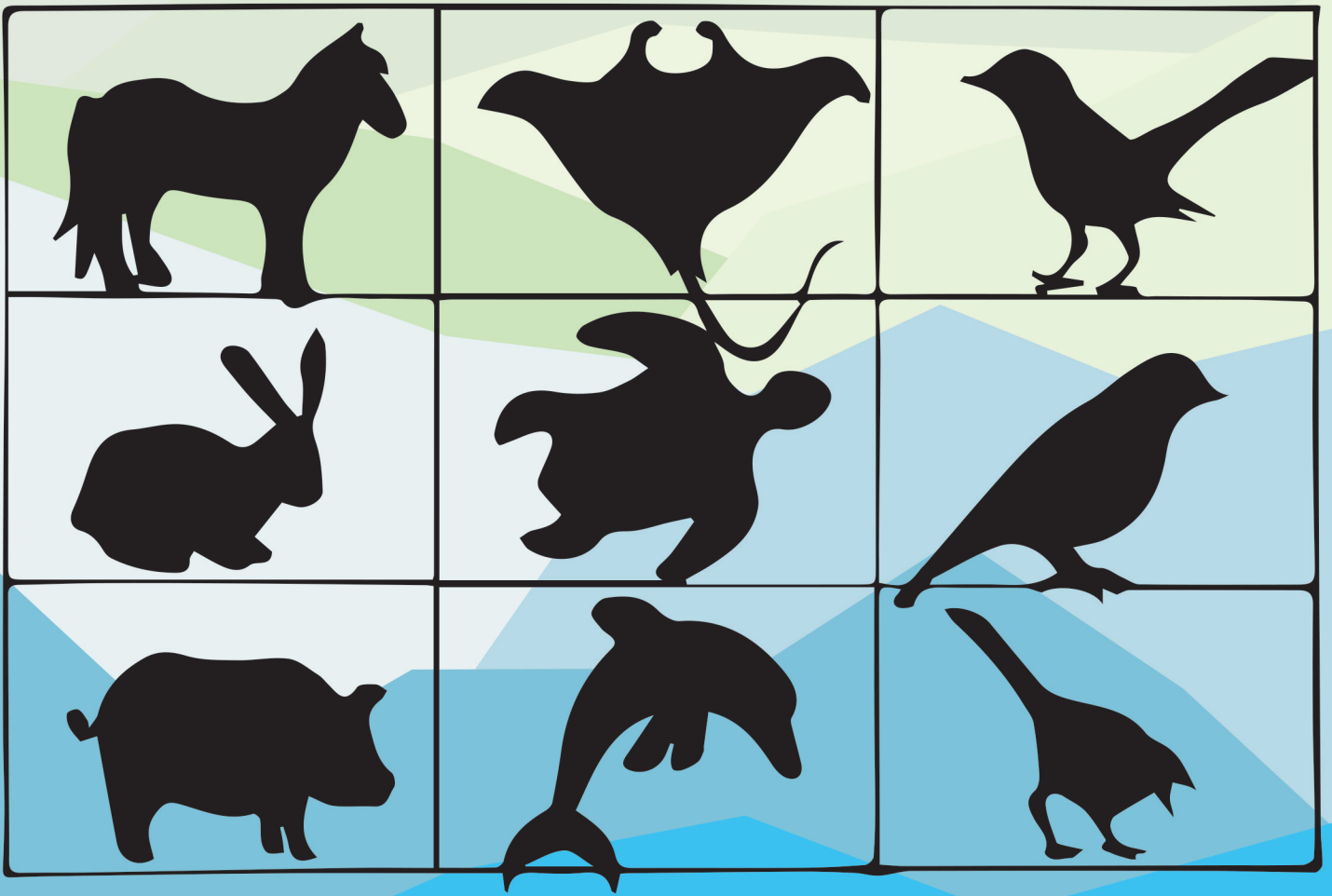
Učbenik je v prvi vrsti namenjen poučevanju oz. študentom biologije, ekologije, naravovarstva in vsebinsko sorodnih naravoslovnih vsebin. Sistematično zbrano znanje bo v oporo tudi strokovnjakom, ki se v poklicnem življenju srečujejo z biodiverzitetnimi vsebinami. Študij biologije ni pogoj za uporabo, uporaben je tako za načrtovalce prostora, kmetijsko proizvodnjo in druge poklicne skupine, ki se na tak ali drugačen način srečujejo z živimi bitji v naravnem okolju.

Učbenik o makroekologiji je tudi prispevek k biodiverzitetni pismenosti. To je sposobnosti prepoznavanja in vrednotenja pomena živih bitij v naravi in se ustrezno odzivati na dogajanje z njimi.

Septembra 2022 na Vurbergu

60 let po Nemi pomladi in v mojem 60. letu starosti

1. poglavje



MAKROEKOLOGIJA IN BIODIVERZITETA

Povzetek

1. poglavje: Makroekologija in biodiverzitet

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

Makroekologija je naravoslovna veda o živih bitjih, o porazdelitvi in vzorcih biodiverzitet v prostoru in času v velikem merilu ter o odzivanju biodiverzitet na spremembe okolja. Metodologija analize biodiverzitet temelji na integraciji različnih kategorij podatkov in njihovem statističnem in numeričnem vrednotenju. Področje makroekološke ekspertize so kompleksni ekološki sistemi, ki zahtevajo povezovanje znanja naravoslovnih disciplin.

Makroekologija predstavlja naravoslovni pristop k raziskovanju vprašanj o stanju in spreminjanju globalne biodiverzitet, tudi kot posledico globalnega spreminjanja okolja z razumevanjem večplastnih dejavnikov s sistemskim vplivom. Z makroekološkimi raziskavami analiziramo in razlagamo tudi biodiverzitetno krizo. V interdisciplinarnem pristopu makroekologija prispeva ključne elemente za napovedovanje prihodnjih scenarijev razvoja porazdelitve živih bitij in omogoča strateško ukrepanje za ohranjanje biodiverzitet. Izvajanje naravovarstvenih ukrepov za ohranitev biodiverzitet je uspešnejše z interdisciplinarno perspektivo, v kateri se dopolnjujejo tako naravoslovne kot družboslovne vede. Kombiniranje znanj in orodij različnih strok omogoča širši pogled na biodiverzitetno krizo, predvsem velike zbirke podatkov v kombinaciji z naprednimi analitskimi orodji omogočajo reševanje kompleksnih problemov in postavljanje naravovarstvenih prioritet.

Biodiverzitet je pestrost in variabilnost živih bitij na Zemlji. Biodiverzitetno prepoznavamo na treh nivojih: genskem, vrstnem in

ekosistemskem. *Organizacija združenih narodov* je pomen biodiverzitet za dobrobit človeštva in življenja na Zemlji leta 1992 prepoznala s sprejemom *Konvencije o biodiverziteti*. Slovenija jo je ratificirala leta 1996. Konvencijo so dopolnili s *Kartagenskim* in *Nagoyskim protokolom*.

Koncept biodiverzitet je večplasten in vključuje tri strukturne vidike: vrstni, funkcionalni in filogenetski vidik, ki jih vrednotimo po treh dimenzijah: bogastvo, raznolikost in poravnost.

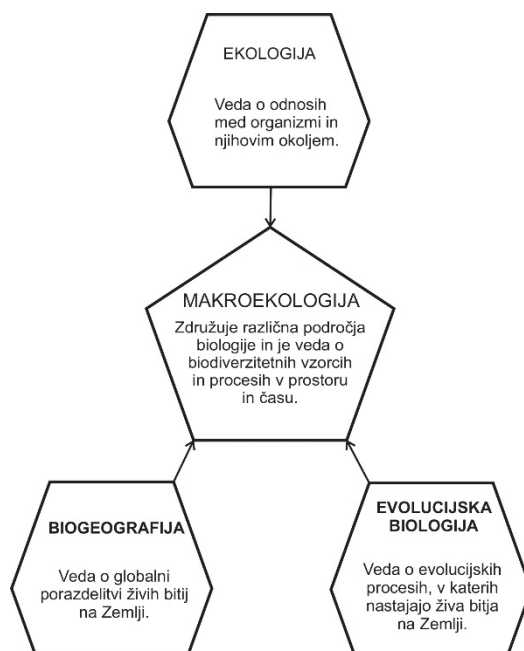
Biodiverzitet ni enakomerna razporejena po planetu. Na razporeditev kopenske biodiverzitet vplivajo: temperatura, padavine, nadmorska višina, zgradba tal in medvrstni odnosi med organizmi. Najvišja pestrost organizmov je v tropskih območjih in se zmanjšuje proti poloma. Evidentirana planetarna biodiverzitet je okrog 1,2 milijona vrst, napovedana pestrost vrst pa znaša 8,7 milijona vrst, od tega 2,2 milijona morskih.

Sodobna sistematika živih bitij klasificira organizme po načelih kladistike in z monofiletskim konceptom. Osnovna delitev živih bitij temelji na izoblikovanosti celičnega jedra. Delimo jih na prokariote Prokaryota s skupinama: arheje Archaea in bakterije Bacteria ter evkariote Eucaryota s petimi skupinami: protozoji Protozoa, rastline Plantae, glive Fungi, kromisti Chromista in živali Animalia. Razdelitev skupin še ni stabilna, nove ugotovitve spreminjajo odnose na nivoju enoceličnih evkariotov. Viruse pogojno uvrščamo med živa bitja, nedvomno so rezultat evolucije, ne izkazujejo pa vseh znakov živih bitij.

1.1 Makroekologija

Brown in Maurer sta leta 1989 uvedla **makroekologijo** (ang. Macroecology) (Brown, 1995) in jo definirala kot vedo o porazdelitvi in vzorcih biodiverzitet v prostoru in času v velikem merilu in z obravnavo številnih vrst. Makroekologija sodi v naravoslovje med ekološke znanosti. Združuje vidike biogeografije, ekologije in evlucijske biologije (Slika 1.1). Metodološko se opira na integracijo različnih tipov podatkov in na mnogotere statistične in numerične pristope pri vrednotenju variabilnosti biodiverzitet v prostoru in času in njenem odzivanju na biotske in abiotske dejavnike. Področje makroekološke obravnave so kompleksni ekološki sistemi, ki zahtevajo znanja vseh naravoslovnih disciplin. Makroekološki pristop vključuje evlucijski in ekološki pristop obravnave, zato je osrednjega pomena pri prepoznavanju vzrokov in mehanizmov, ki določajo odziv vrst, združb in ekosistemov v spreminjajočem se okolju. Zaradi svojega osrednjega položaja v naboru ekološke in evlucijske znanosti predstavlja makroekologija celostni pristop k raziskovanju zapletenih vprašanj o spreminjanju globalne biodiverzitet, tudi kot posledico globalnega spreminjanja okolja. Biodiverzitetno krizo, v kateri se nahajamo s šestim globalnim izumiranjem in razumevanje večplastnih dejavnikov s sistemskim vplivom, ki jo povzročajo, lahko analiziramo in razlagamo z makroekološkimi raziskavami. V interdisciplinarnem pristopu, ki ga zahtevajo izzivi biodiverzitetne krize, makroekologija prispeva ključne elemente za napovedovanje prihodnjih scenarijev v razvoju biodiverzitet in znanje za strateško ukrepanje pri ohranjanju biodiverzitet. Z vključevanjem preteklih evlucijskih dogodkov v makroekološke raziskave, s pomočjo filogenetskih informacij, in z rekonstrukcijo človekovega vpliva na naravo lahko izboljšamo razumevanje procesov, ki določajo vzorce in procese biodiverzitet. Napovedovanje prihodnje porazdelitve biodiverzitet, glede na projekcije razvoja klime, rabe prostora in izčrpavanja naravnih virov, je pomembno orodje za oceno posledic na biodiverzitet. Izvajanje naravovarstvenih ukrepov za ohranitev biodiverzitet je uspešneje z interdisciplinarno perspektivo, v kateri se dopolnjujejo tako naravoslovne kot družboslovne vede.

Interdisciplinarnost je definirana kot sodelovanje dveh ali več disciplin pri reševanju izzivov, ki ga ena sama disciplina ne more rešiti. Transdisciplinarni pristop vključuje sodelovanje med znanstveniki in praktiki. V primerih, ko neka disciplina postane samo servis drugi (npr. ponudnik storitev za daljinsko zaznavanje podnebja ali vegetacije) multidisciplinarnost ni pravi interdisciplinarni pristop. Čeprav vsa makroekološka vprašanja ne potrebujejo multidisciplinarnega pristopa, pa kombiniranje znanj in orodij iz različnih strok omogoča širši pogled na znanstvene izzive. To postaja pomembno v sedanjih dobi velikih zbirk podatkov (ang. Big Data Set) in njihov obseg hitro narašča. Velike zbirke podatkov v kombinaciji z naprednimi analitskimi orodji omogočajo znanstvenikom reševanje kompleksnih problemov, npr. satelitsko spremljanje globalne kopenske biodiverzitet in na tej podlagi postavljanje naravovarstvenih prioritet.



Slika 1.1 Makroekologija je veda o biodiverzitetnih vzorcih in procesih v globalnem merilu in združuje spoznanja ekologije, biogeografije in evolucijske biologije.

1.2 Biodiverziteta

Z besedo **biodiverziteta** opredeljujemo pestrost in variabilnost živega sveta na Zemlji (Gaston in Spicer, 2004). *Okoljski program Združenih narodov* (ang. United Nations Environment Programme UNEP) biodiverziteto razlikuje na treh hierarhičnih nivojih: genskem, vrstnem in ekosistemskem nivoju. Več o razvoju področja in njegovih naravoslovno-družboslovnih razsežnostih v globalnem kontekstu lahko preberete na [www Biodiversity](http://www.Biodiversity) (2018) in v regionalnem merilu v *Nacionalnem programu varstva okolja* (2019) in Kryštufek (1999).

Koncept biodiverzitete je večplasten (ang. Multifaceted) in vključuje tri strukturne vidike: vrstno, funkcionalno in filogenetsko pestrost, ki jih vrednotimo po treh dimenzijah: bogastvo, raznolikost in poravnianost (Preglednica 1.1). **Vrstni vidik pestrosti** obsega variabilnost v vrstni sestavi, tako na nivoju vrstne sestave lokalne združbe organizmov, spreminjanje vrstne sestave med združbami in vrstno sestavo na nivoju pokrajine. **Funkcionalni vidik pestrosti** obsega variabilnost ekomorfoloških, reproduktivnih in socio-etoloških lastnosti vrst. Omogoča primerjave med nabori vrst in mehanistično povezavo z ekosistemsko funkcionalnostjo vrst, njihovo odpornostjo in zmožnostjo okrevanja. **Filogenetski vidik pestrosti** obsega pestrost dednine. Ocenjevanje evolucijskih razdalj med vrstami, od njihove ločitve od skupnega prednika in oceno v dednini ohranjenih ekoloških in morfoloških razlik med vrstami. Nabor vrst obravnavamo po treh vidikih diverzitete (vrstni, funkcionalni, filogenetski) in potem vsak posamezni vidik vrednotimo po treh dimenzijah (bogastvo, raznolikost, poravnianost). **Dimenzija bogastva** (ang. Richness) obsega število vrst, število stanj funkcionalnih znakov in število

haplotipov. **Dimenzija raznolikosti** (ang. Divergence) obsega pestrost vrst, pestrost funkcionalnih znakov in pestrost haplotipov. **Dimenzija poravnosti** (ang. Regularity) obsega poravnost v vrstni sestavi, poravnost v funkcionalnih znakih in poravnost v evolucijskih linijah. Vsak vidik diverzitet je obdelan v svojem poglavju, kjer so tudi podrobneje predstavljeni analitski pristopi vrednotenja posamezne dimenzije.

Preglednica 1.1 Večplasten značaj biodiverzitet po treh vidikih in treh dimenzijah. Vsaka kombinacija vidika in dimenzija ima lasten nabor analiz, navedene so oznake za samo eno cenilko (celoten nabor cenilk in podrobna razlaga je pri posameznem poglavju)

| | | Vidik pestrosti | | |
|---------------------|-----------|-------------------------|------------------------------|-------------------------------|
| | | Vrstni | Funkcionalni | Filogenetski |
| Dimenzija pestrosti | Bogastvo | Vrstno bogastvo, S | Funkcionalno bogastvo, FRic | Filogenetsko bogastvo, PD |
| | Pestrost | Vrstna pestrost, H' | Funkcionalna pestrost, FDis | Filogenetska pestrost, avPD |
| | Poravnost | Vrstna poravnost, H_a | Funkcionalna poravnost, FEve | Filogenetska poravnost, I_c |

Vrstni, funkcionalni in filogenetski vidik diverzitet v kombinaciji z dimenzijo analiziramo tudi po komponentah prostora in časa. Pri tem je osnovna enota analiziranja nabor vrst (ang. Species Assembly) (sin. vrstna sestava), ki smo jih zajeli z vzorčenjem; enota vzorčenja pa je vzorec. Glede na prostorsko ali časovno komponento razlikujemo alfa, beta in gama diverziteti. V gradientu prostora obravnavamo procese: razmerje med površino ozemlja in diverziteti, odziv diverzitet vzdolž gradienta geografske širine in odziv diverzitet vzdolž gradienta nadmorske višine. V časovni komponenti raziskujemo dnevno variabilnost (dnevno nočni ritem), sezonsko variabilnost (med letnimi časi), variabilnost v dalj časa trajajočih obdobjih (več let ali desetletij) in evolucijsko dimenzijo variabilnosti (zemeljska časovna obdobja), spreminjanje diverzitet v zemeljskih časovnih obdobjih.

Variabilnost posameznih vidikov biodiverzitet obravnavamo tudi v kontekstu abiotskih in biotskih gradientov, kjer sta vidika obravnave odziv biodiverzitet na spreminjajoče se dejavnike okolja in medvrstne interakcije. V kontekst ekoloških gradientov je pomemben dejavnik filtriranje vrst (mobilnost, okoljski filtri, biotski filtri).

Vidike biodiverzitet obravnavamo tudi na različnih hierarhičnih nivojih telesne organizacije. V vrstni vidik vključujemo osebke vrst, torej njihovo prisotnost ali njihovo število oz. delež. V funkcionalni vidik obravnave vključujemo telesne dimenzije, prehranske kategorije, razmnoževalne lastnosti, socialno-etološke lastnosti in druge ekološke lastnosti. V obravnavo filogenetskega vidika pestrosti vključimo filogenetsko drevo, ki je narejeno na nukleotidnem zaporedju izbrane regije genoma.

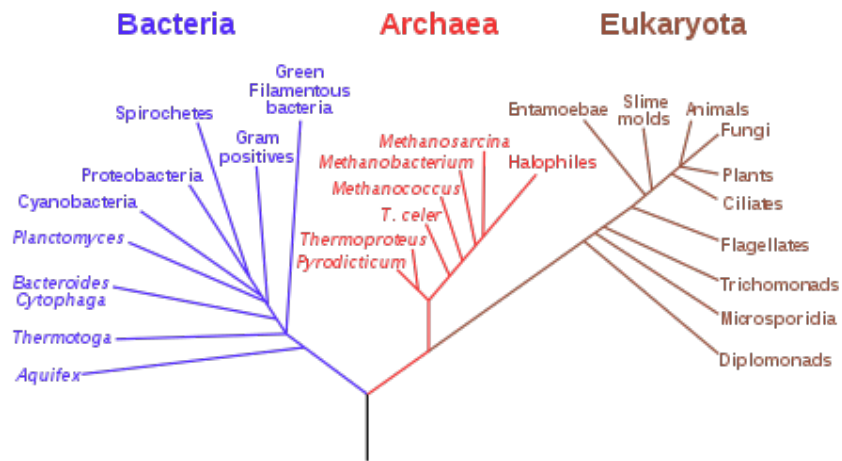
Zavedanje in pomen biodiverzitete za dobrobit človeštva in življenja na Zemlji je bil leta 1992 prepoznano v Združenih narodih in formalizirano v *Konvenciji o biodiverziteti* (ang. Convention on Biological Diversity CBD) (1996), ki jo je ratificirala tudi Slovenija. Konvencija je bila leta 2000 dopolnjena s *Kartagenskim protokolom* (2002) in leta 2014 z *Nagoyskim protokolom* (2014).

Biodiverziteta ni enakomerna razporejena po planetu. Na razporeditev kopenske biodiverzitete vplivajo predvsem: temperatura, padavine, nadmorska višina, zgradba tal in medvrstni odnosi med organizmi. Najvišja pestrost organizmov je praviloma v tropskih območjih in se zmanjšuje proti poloma. Diverzitetna kopenskih organizmov je približno 25 krat višja od morske biodiverzitete. Evidentirana planetarna biodiverziteta je okrog 1,2 milijona vrst, napovedana pestrost vrst znaša 8,7 milijona vrst, od tega 2,2 milijona morskih (Preglednica 1.2). Inventar vrst vodijo v projektu Katalog življenja (ang. Catalogue of Life), (www.sp2000.org).

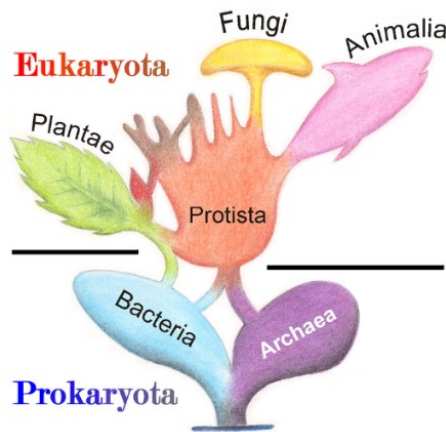
Sistematika živih bitij je predmet živahnih strokovnih polemik. Sodobna klasifikacija organizmov po načelih moderne kladistike in z monofiletskim konceptom prepoznava tri domene (Slika 1.2). Taksonomska kategorija kraljestvo izgublja pomen, saj rang kraljestvo ne sledi monofiletskemu konceptu. Osnovna delitev živih bitij temelji na izoblikovanosti celičnega jedra. Delimo jih na prokariote Prokaryota in evkariote Eukaryota (Slika 1.3). Skupina protistov Protista je postala »odlagališče« za vse evkariote, ki niso živali, rastline ali glive. Zato so na Mednarodnem združenju protistologov (International Society of Protistologists) predlagali klasifikacijo evkariotov v več "superskupin" (Slika 1.4), kar pa formalno taksonomsko še ni uveljavljeno.

Preglednica 1.2 Katalogizirano in napovedano število vrst na kopnem in v oceanih
(povzeto iz Mora in sod. 2011)

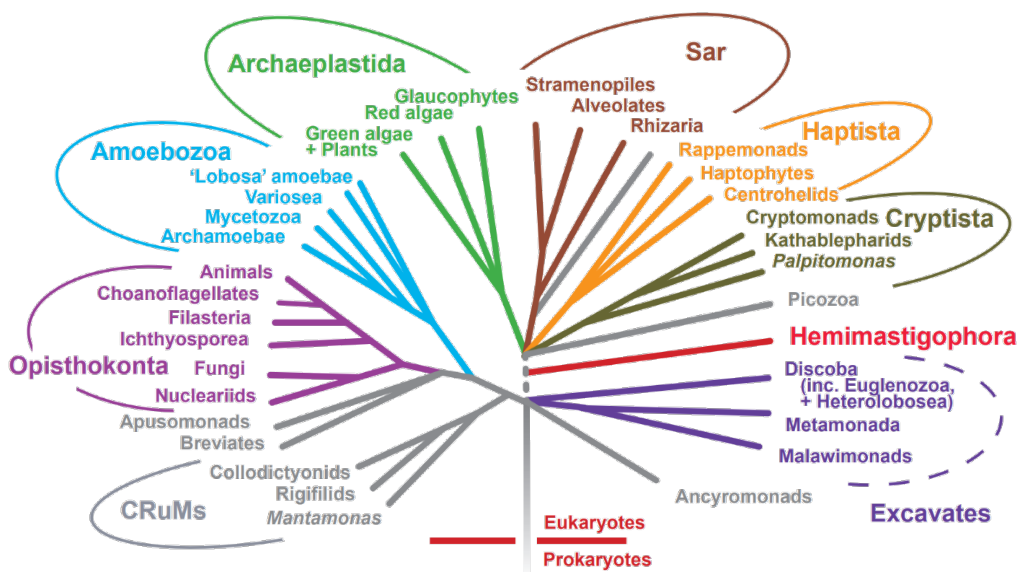
| Organizmi | Kopno | | Oceani | |
|-------------------------------|------------------|------------------|----------------|------------------|
| | Katalogizirano | Napovedano | Katalogizirano | Napovedano |
| Evkarionti, Eukaryota | | | | |
| Živali, Animalia | 953.434 | 7.770.000 | 171.082 | 2.150.000 |
| Kromisti, Chromista | 13.033 | 27.500 | 4.859 | 7.400 |
| Glive, Fungi | 43.271 | 611.000 | 1.097 | 5.320 |
| Rastline, Plantae | 215.644 | 298.000 | 8.600 | 16.600 |
| Protozoji, Protozoa | 8.118 | 36.400 | 8.118 | 36.400 |
| <i>Skupaj</i> | 1.233.500 | 8.740.000 | 193.756 | 2.210.000 |
| Prokarioti, Prokaryota | | | | |
| Arheje, Archaea | 502 | 455 | 1 | 1 |
| Bakterije, Bacteria | 10.358 | 9.680 | 652 | 1.320 |
| <i>Skupaj</i> | 10.860 | 10.100 | 653 | 1.320 |
| <i>Skupaj skupaj</i> | 1.244.360 | 8.750.000 | 194.409 | 2.210.000 |



Slika 1.2: Filogenetsko drevo zasnovano na rRNA tri domene so razdeljene v številna kraljestva. Vzeto iz Kingdom Biology (2022).



Slika 1.3: Filogenetsko in simbiogenetsko drevo živih organizmov. Vzeto iz Kingdom Biology (2022).



Slika 1.4 Hipoteza o filogenetskih razmerjih med evkarionti, modificirano po Simpson in Roger 2004. Vzeto iz Kingdom Biology (2022).

Pri raziskovanju biodiverzitete se soočamo z različnimi omejitvami pri ocenjevanju oz. vrednotenju števila vrst. Na oceno števila vrst vplivajo:

- Definicija vrste – različni taksonomski strokovnjaki (zoologi, botaniki, mikrobiologi, bakteriologi, mikologi ...) uporabljajo različne definicije vrste, prav tako tudi različne nivoje diferenciacije za prepoznavanje vrst. Posledica različnih pristopov je, da niso možne neposredne primerjave med pestrostjo skupin.
- Spremembe višjih taksonov – število višjih taksonov se pogosto spreminja zaradi novih spoznanj, posledica so razdružitve ali združevanja višjih taksonov iz fenetskih v filogenetske skupine, istočasno prihaja do sinonimije (različno poimenovanje istih vrst).
- Taksonomsko delo – število aktivnih raziskovalcev in strokovnjakov za posamezne skupine živih bitij (taksonomi) se močno razlikuje, od tega pa je odvisna stopnja raziskanosti in opisovanje novih vrst.

Statistični pristop. Preučevanje porazdelitve in vzorcev biodiverzitete poteka z numeričnimi in statističnimi analitskimi orodji. Makroekologija je eksaktna naravoslovna veda, ki temelji na meritvah vrstnega bogastva in razlaga vzorce biodiverzitete z okoljskimi dejavniki. Iskanje povezav med velikostjo populacij, vrstno pestrostjo in okoljskimi gradienti omogočajo statistične metode, predvsem regresijske analize. Analiza združb in vrednotenje bogastva vključuje različne ordinacijske in klastrske metode.

Makroekologija pomeni tudi kvantifikacijo biodiverzitete in analizo njene variabilnosti z numeričnimi in statističnimi metodami. Viri variabilnosti biodiverzitete so:

- Celoten spekter naravnih okoljskih gradientov in umetnih gradientov, ki jih povzročamo ljudje, vključno s polucijo in degradacijo habitatov.
- Prostorske dimenzije življenjskih okolij in površine habitatov, vključno z zmanjševanjem in fragmentacijo naravnih habitatov zaradi širjenja kmetijstva in infrastrukture.
- Medvrstni odnosi, vključno z ne trajnostno rabo populacij s strani ljudi.
- Znotraj vrstni odnosi, vključno s kompeticijo za iste naravne vire.
- Evolucijski procesi z nastajanjem novih vrst in naravnim izumiranjem vrst, vključno z naravovarstvenimi aktivnostmi za ohranjanje vrst in istočasnim iztrebljanjem vrst zaradi vseprisotne in vseobsegajoče aktivnosti ljudi.

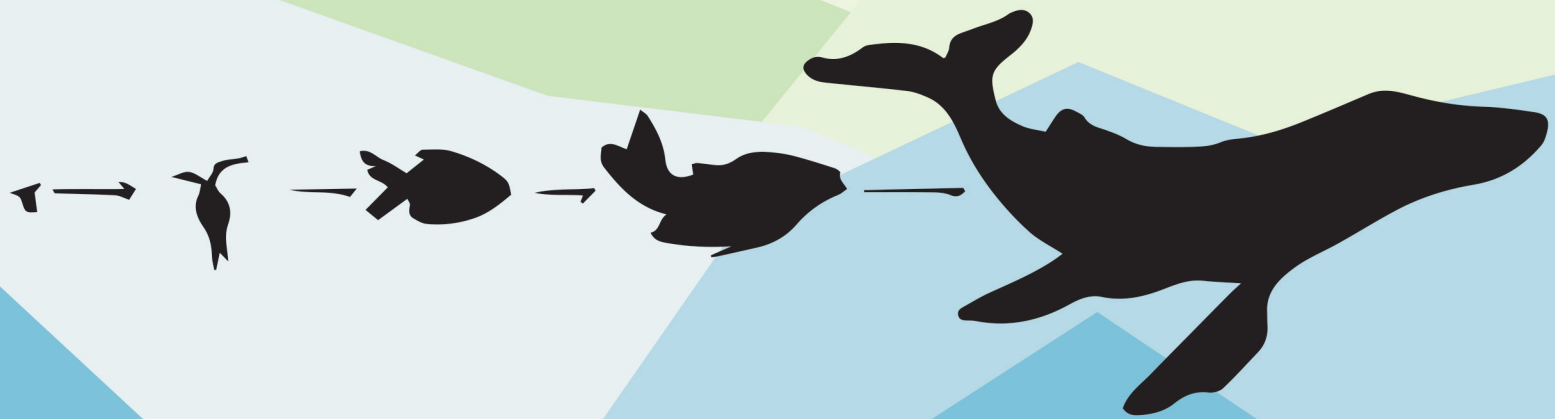
Zaključki

- Makroekologija je naravoslovna veda o porazdelitvi in vzorcih biodiverzitet v prostoru in času, v velikem merilu in sočasno obravnava mnogo vrst.
- Makroekološka metodologija temelji na integraciji različnih kategorij podatkov in njihovem statističnem in numeričnem vrednotenju. Področje obravnave so kompleksni ekološki sistemi, ki zahtevajo znanja različnih disciplin.
- Makroekologija predstavlja celostni pristop k raziskovanju spreminjanja globalne biodiverzitet, tudi kot posledico globalnega spreminjanja okolja z razumevanjem večplastnih dejavnikov s sistemskim vplivom.
- Z makroekološkimi raziskavami analiziramo in razlagamo biodiverzitetno krizo, 6. globalno izumiranje.
- Makroekologija z interdisciplinarnim pristopom prispeva ključne elemente za napovedovanje prihodnjih scenarijev in strateške ukrepe za ohranjanje biodiverzitet.
- Znanja različnih strok zagotavljajo širši pogled na biodiverzitetno krizo, predvsem velike zbirke podatkov v kombinaciji z naprednimi analitskimi orodji omogočajo reševanje kompleksnih problemov in postavljanje naravovarstvenih priorit.
- Biodiverzitet je pestrost in variabilnost živega sveta na Zemlji.
- Biodiverzitetno prepoznavamo na treh nivojih: genetskem, vrstnem in ekosistemskem.
- Organizacija združenih narodov je leta 1992 sprejela Konvencijo o biodiverziteti, Slovenija jo je ratificirala 1996. Konvencija je bila dopolnjena s Kartagenskim in Nagoyskim protokolom.
- Koncept biodiverzitet je večplasten in vključuje tri strukturne vidike: vrstni, funkcionalni in filogenetski vidik pestrost, ki jih vrednotimo po treh dimenzijah: bogastvo, raznolikost in poravnost.
- Biodiverzitet ni enakomerna razporejena po planetu. Na razporeditev kopenske biodiverzitet vplivajo: temperatura, padavine, nadmorska višina, zgradba tal in medvrstni odnosi med organizmi. Najvišja pestrost organizmov je v tropskih območjih in se zmanjšuje proti poloma.
- Evidentirana planetarna biodiverzitet je 1,2 milijona vrst, napovedana pestrost vrst znaša 8,7 milijona vrst, od tega 2,2 milijona morskih.
- Sistematika živih bitij klasificira organizme po načelih kladistike in z monofiletskim konceptom.
- Osnovna delitev živih bitij temelji na izoblikovanosti celičnega jedra, delimo jih na prokariote Procaryota s skupinama: arheje Archaea in bakterije Bacteria ter evkariote Eucaryota s petimi skupinami: protozoi Protozoa, rastline Plantae, glive Fungi, kromisti Chromista in živali Animalia.
- Viruse pogojno uvrščamo med živa bitja, nedvomno so rezultat evolucije, ne izkazujejo pa vseh znakov živih bitij.

Literatura

- Bebber D.P., Marriott F.H.C., Gaston K.J., Harris S.A., Scotland R.W. 2007. Predicting unknown species numbers using discovery curves. *Proc. R. Soc. B*, 274: 1651–1658.
- Biodiversity. 2018. Wikipedia. Preneseno 2. 10. 2018, iz <https://en.wikipedia.org/wiki/Biodiversity>
- Brown J.H. 1995. *Macroecology*. The University of Chicago Press. Chicago.
- CBD. 2018. Convention on Biological Diversity. Preneseno 2. 10. 2018, iz <https://www.cbd.int/>.
- Gaston K.J., Spicer J.I. 2004. *Biodiversity: An Introduction*. Second Edition. Blackwell Publishing. Oxford.
- Kartagenski protokol. 2002. Uradni list RS - Mednarodne pogodbe št. 23/02. Zakon o ratifikaciji Kartagenskega protokola o biološki varnosti h Konvenciji o biološki raznovrstnosti. Preneseno 2. 10. 2018, iz <http://pisrs.si/Pis.web/pregledPredpisa?id=ZAKO3465>.
- Kingdom (Biology). 2022. Wikipedija. Preneseno 25. 10. 2022, iz [https://en.wikipedia.org/wiki/Kingdom_\(biology\)](https://en.wikipedia.org/wiki/Kingdom_(biology)).
- Konvencija o biološki raznovrstnosti. 1996. Zakon o ratifikaciji konvencije o biološki raznovrstnosti. UL RS – Mednarodne pogodbe št. 7/96. Preneseno 2. 10. 2018, iz <http://pisrs.si/Pis.web/pregledPredpisa?id=ZAKO785>.
- Kryštufek B. 1999. *Osnove varstvene biologije*. Tehniška založba Slovenije. Ljubljana.
- Leitão P.J., Andrew C.J., Engelhardt E.K., Graham C.H., Martinez-Almoyna C., Mimet A., Pinkert S., Schröder B., Voskamp A., Hof C., Fritz S.A. 2020. Macroecology as a hub between research disciplines: Opportunities, challenges and possible ways forward. *Journal of Biogeography*, 47: 13–15.
- Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. 2011. How Many Species Are There on Earth and in the Ocean? *PLoS Biol*, 9(8): e1001127. doi:10.1371/journal.pbio.1001127.
- Nacionalni program varstva okolja. 2019. Priprava nacionalnega programa varstva okolja 2030. Ministrstvo za okolje in prostor. Preneseno 1 8. 2019, iz http://www.mop.gov.si/si/priprava_nacionalnega_programa_varstva_okolja_2030/.
- Nagojski protokol. 2014. Uredba (EU) št. 511/2014 Evropskega parlamenta in Sveta z dne 16. aprila 2014 o ukrepih za izpolnjevanje obveznosti uporabnikov iz Nagojskega protokola o dostopu do genskih virov ter pošteni in pravični delitvi koristi, ki izhajajo iz njihove uporabe, v Uniji. Preneseno 2. 10. 2018, iz https://eur-lex.europa.eu/legal-content/SL/ALL/?uri=uriserv:OJ.L_.2014.150.01.0059.01.SLV.
- Presley S., Willing M. 2022. Dimensions of Biodiversity. Preneseno 29. 5. 2022, iz: <https://hydrodictyon.eeb.uconn.edu/people/willig/Research/Dimensions%20of%20Biodiversity.html>.

2. poglavje



TELESNA VELIKOST

Povzetek

2. poglavje: Telesna velikost

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

Velikost telesa je ena od osnovnih lastnosti organizma. Z njo so povezane druge biološke lastnosti, npr.: življenjska doba, velikost domačega okolja, razmerja med telesnimi organi oz. organskimi sistemi, način premikanja in številni drugi vidiki življenja in ekologije. Odnos med telesno velikostjo in abundanco organizmov oblikuje povezavo med nivojem osebka in nivojem populacije. Slednje vpliva na strukturo vrstne sestave in na dinamiko ekološke združbe. Telesna velikost je ključna za presnovne procese in posledično tudi za porabo naravnih virov (hrane, prostora ...), zato odnos med velikostjo in abundanco razkriva porazdelitev naravnih virov v ekološkem sistemu.

Frekvenčna porazdelitev telesne velikosti organizmov v naboru vrst je posledica več energetske-evolucijskih mehanizmov, in sicer stopnje speciacije, stopnje izumiranja, stopnje disperzije, medvrstne kompeticije in energetskih omejitev. Vzorec porazdelitev je odvisen tudi od merila v katerem opazujemo pojav. V lokalnih razmerah, npr. habitat ali posamezna združba, je porazdelitev praviloma zelo desno asimetrična, kar pomeni prevladujoče število vrst, ki so manjše od povprečja. V globalnem merilu na nivoju kontinentov se porazdelitev telesne velikosti približuje normalni porazdelitvi.

Odnos med telesno velikostjo organizmov in njihovo abundanco prepoznavamo na štiri nivojih. 1) Globalni odnos med velikostjo in gostoto je monotono spadajoč in se v povprečju dobro ujema z regresijsko funkcijo s smernim koeficientom $-0,75$. 2) Lokalni odnos med velikostjo in gostoto izkazuje bistveno nižjo stopnjo povezanosti, smerni koeficient regresijske premice ima vrednost okrog $-0,25$. Projekcija vrednosti gostote na maso oblikuje trikotnik, kar je posledica velike variabilnosti gostote majhnih do srednje velikih vrst. 3) Individualna porazdelitev velikosti je frekvenčna porazdelitev telesne velikosti osebkov v združbi ne glede na vrstno pripadnost. Osebkki so razporejeni v ustrezni velikostni razred in posamezen razred lahko vsebuje predstavnike različnih vrst. Ena vrsta je lahko zastopana v več velikostnih razredih. 4) Odnos med združbami lahko opišemo kod samoredčenje, znan je predvsem pri sesilnih združbah. Rast posameznega organizma pusti manj in manj prostora za druge posameznike. Gre za negativen odnos med povprečno velikostjo organizma in celokupno abundanco združbe.

Velikost telesa (sin. telesna velikost, ang. Body Size) je ena od osnovnih lastnosti vsakega organizma. S telesno velikostjo so povezane druge biološke lastnosti, npr.: življenjska doba, velikost domačega okolja, razmerja med telesnimi organi oz. organskimi sistemi, način premikanja in številni drugi vidiki življenja in ekologije.

Posebna pozornost ekologov in evolucionistov je namenjena odnosu med telesno velikostjo in abundanco organizmov. Odnos oblikuje ključno povezavo med nivojem osebka in nivojem populacije v ekološki ali evolucijski obravnavi vrste. Slednje se izkazuje v strukturi vrstne sestave in v dinamiki ekološke združbe. Analiza telesne velikosti je tudi ena od ključnih vsebin makroekoloških analiz. Še več, telesna velikost pogojuje presnovne procese in posledično porabo naravnih virov (hrane, prostora ...). Odnos med velikostjo in abundanco zato razkriva tudi porazdelitev virov znotraj ekološkega sistema.

Porazdelitev telesne velikosti v okviru nabora vrst je posledica več energetske evolucijskih mehanizmov: stopnja speciacije, stopnja izumiranja, stopnja disperzije, interspecifična kompeticija in energetske omejitve (Brown in Nicoletto 1991, Gaston in sod. 1997, Maurer in sod. 1992).

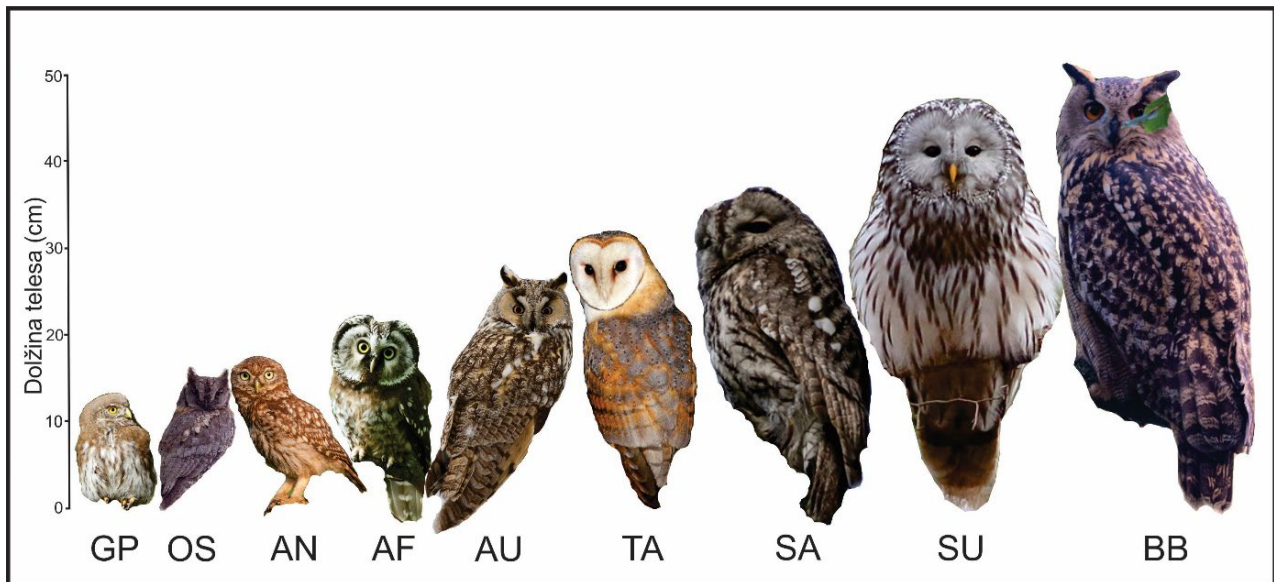
Stopnja speciacije – Majhni organizmi imajo večinoma krajši generacijski čas, hitreje spolno dozori in se prej razmnožujejo. Te lastnosti pospešujejo speciacijo, saj se ob hitrejšem metabolizmu zvišata število mutacij in število selekcijskih dogodkov in molekularna evolucija je hitrejša. Bakterije npr. razvijejo rezistenco na antibiotike v zelo kratkem času, saj »evolucijska ura bije hitreje« pri manjših organizmih. Kakorkoli že, tovrstnih dokazov ni pri homeotermih, npr. pri sesalcih in pticah.

Stopnja izumiranja – Tveganje za izumrtje vrste je v neposredni povezavi z velikostjo njene populacije/ij. Majhne populacije izumirajo pogosteje kot velike. Telesno velike vrste potrebujejo več naravnih virov, da zadostijo dnevne presnovne potrebe, kar posledično pomeni manjše število telesno velikih osebkov na enoto površine. Zato veliki organizmi praviloma naseljujejo večje območje, kar poveča velikost populacije in zmanjša tveganje za izumrtje. Z drugimi besedami, večji areal vrste pomeni nižjo tveganje za njeno izumrtje. Kadar je vse ostalo enako, bo telesno velika vrsta z majhnim arealom slej ko prej izumrla.

Stopnja disperzije – Disperzija je pomemben mehanizem genskega pretoka, njeno razumevanje pa ne pomaga k razlagi vzorca odnosa med telesno velikostjo in vrstnim bogastvom. Dosedanje raziskave ne potrjujejo povezanosti med telesno velikostjo in sposobnostjo disperzije.

Interspecifična kompeticija – Vrste, ki na podoben način izkoriščajo naravne vire, so praviloma kompetitorji (tekmeci). Če sta kompetitorni vrsti enake velikosti, to še poveča tekmovalnost. Predstavniki istega ceha (ang. Guild), ki na podoben način izkoriščajo iste

vire, se praviloma kategorično razlikujejo v telesni velikosti. Primer je porazdelitev telesne velikosti sov, ki v nočnem času plenijo male sesalce in tudi večje žuželke in druge vrste manjših ptičev. Ko sove razporedimo v naraščajočo vrsto telesne velikosti (Preglednica 2.1), je posamezna vrsta od predhodne večja v povprečju za 1/3 (Preglednica 2.1, Slika 2.1).



Slika 2.1 Devet vrst sov, ki živijo v osrednjem delu evropskega kontinenta, razporejene po naraščajoči vrednosti telesne velikosti. GP – mali skovik *Glaucidium passerinum* (povprečna dolžina telesa 16,5 cm), OS – veliki skovik *Otus scops* (19,5 cm), foto Gregor Bernard; AN – navadni čuk *Athene noctua* (22 cm), AF – koconogi čuk *Aegolius funereus* (25 cm), AU – mala uharica *Asio otus* (36 cm), foto Gregor Bernard; TA – pegasta sova *Tyto alba* (34 cm), SA – lesna sova *Strix aluco* (38 cm), foto Gregor Bernard; SU – kozača *Strix uralensis* (61 cm), foto Gregor Bernard; BB – velika uharica *Bubo bubo* (67,5 cm).

Fotografije prenesene iz spletne strani <https://galerija.foto-narava.com/>.

Energetske omejitve – Telesno majhne živali imajo večjo možnost, da so številčne, kot telesno velike živali. Pri tem lahko izhajamo iz enostavne geometrijske zakonitosti, v isto prostornino gre več majhnih kot velikih elementov. Vendar so v ekoloških sistem ob prostoru bistvene tudi druge omejitve. Drugi naravni viri, predvsem hrana so ključni omejujoči dejavnik za število osebkov. Po drugi strani je za telesno majhne živali na razpolago bistveno več ekoloških niš kot za telesno velike živali. Telesno velike živali potrebujejo večjo količino hrane kot majhne živali, po drugi strani potrebujejo relativno manjšo količino energije na enoto telesne mase. Telesno velike živali se praviloma prehranjujejo z energetsko revnejšo hrano, v primerjavi s telesno majhnimi vrstami, za katere sta kakovost hrane in njen energetski izkoristek zelo pomembna.

Kakorkoli že, v skladu z metabolno teorijo v ekologiji, s katero razlagamo makroekološki vzorec v telesni velikosti, je relativni energetski tok skozi vse velikostne razrede živali izenačen. Slednje vpliva na število osebkov v posameznem velikostnem razredu in

potencialno tudi na število vrst. Zakonitost znotraj posameznega velikostnega razreda je za odnos med vrstnim bogastvom S in telesno velikostjo L (telesna dolžina), posplošena v enačbi:

$$S = L^{-2}$$

Zakonitost razmerja med vrstnim bogastvom in telesno velikostjo, kjer bogastvo upada s kvadratom naraščanja telesne dolžine, so potrdili pri različnih kopenskih živalih, med njimi pri sesalcih, pticah, hroščih in metuljih (May 1988).

Preglednica 2.1 Pregled 13 vrst evropskih sov v naraščajoči vrsti njihove povprečne mase. Povprečno povečanje glede na predhodno vrsto in delež povečanja. Podatki o povprečni telesni masi vrst so povzeti iz funkcionalnih znakov (Storchová, Hořák 2018)

| Vrsta | Vrsta | Povprečna masa (g) | Povečanje mase glede na predhodno vrsto (g) | Delež povečanja mase glede na predhodno vrsto (%) |
|-------------------|------------------------------|--------------------|---|---|
| Mali skovik | <i>Glaucidium passerinum</i> | 62,5 | | |
| Veliki skovik | <i>Otus scops</i> | 90,5 | 28 | 30,9 |
| Koconogi čuk | <i>Aegolius funereus</i> | 135 | 44,5 | 33,0 |
| Čuk | <i>Athene noctua</i> | 153 | 18 | 11,8 |
| Mala uharica | <i>Asio otus</i> | 255,5 | 102,5 | 40,1 |
| Močvirska uharica | <i>Asio flammeus</i> | 295 | 39,5 | 13,4 |
| Skobčja sova | <i>Surnia ulula</i> | 308 | 13 | 4,2 |
| Pegasta sova | <i>Tyto alba</i> | 354 | 46 | 13,0 |
| Lesna sova | <i>Strix aluco</i> | 495,5 | 141,5 | 28,6 |
| Kozača | <i>Strix uralensis</i> | 765 | 269,5 | 35,2 |
| Brada sova | <i>Strix nebulosa</i> | 891,5 | 126,5 | 14,2 |
| Snežna sova | <i>Bubo scandiacus</i> | 1.925 | 1.033,5 | 53,7 |
| Velika uharica | <i>Bubo bubo</i> | 2.288 | 363 | 15,9 |

Telesno velikost organizma lahko podajamo v absolutnih vrednosti (npr. z dolžino telesa ali maso) ali relativno glede na primerjave z drugimi organizmi. Npr. siva podgana je velika v primerjavi s hišno mišjo, vendar je manjša od kune belice. V raziskovalni praksi merimo posamezne količine, ki so povezane z velikostjo njihovega telesa. Najpogostejši oceni velikosti sta telesna masa, ki jo izrazimo v g ali kg , in dolžina telesa, ki jo najpogosteje izrazimo v mm , cm ali m . Meritve telesa se nanašajo na bolj ali manj standardizirane mere posameznih telesnih segmentov. Npr. pri malih sesalcih merimo: telesno maso, dolžino telesa, dolžino repa, dolžino zadnjega stopala, dolžino uhlja, dolžino lobanje ipd. Pri pticah merimo: telesno maso, dolžino telesa, dolžino peruti, dolžino repa, dolžino kljuna, dolžino kraka (tarzo-metatarzusa) ipd. V analizo telesne velikosti najpogosteje vključimo bodisi telesno maso bodisi dolžino telesa, nobena od teh vrednosti pa v celoti ne podaja informacije o dejanski velikosti organizma. V primerih, ko razpolagamo z meritvami različnih segmentov telesa, lahko skupno informacijo o telesni velikosti izluščimo iz niza

podatkov s pomočjo multivariatne metode Analiza glavnih komponent. V njej 1. glavna komponenta razlaga variabilnost v telesni velikosti, podrobnejšo razlago in postopek glej v Poglavjih 11 in 12. To pomeni, da v nadaljnje analize in vrednotenje telesne velikosti ne vključimo meritev posameznih telesnih segmentov, temveč v nadaljnje analize vključimo prvo glavno komponento, ki je nosilka informacije o velikosti organizma.

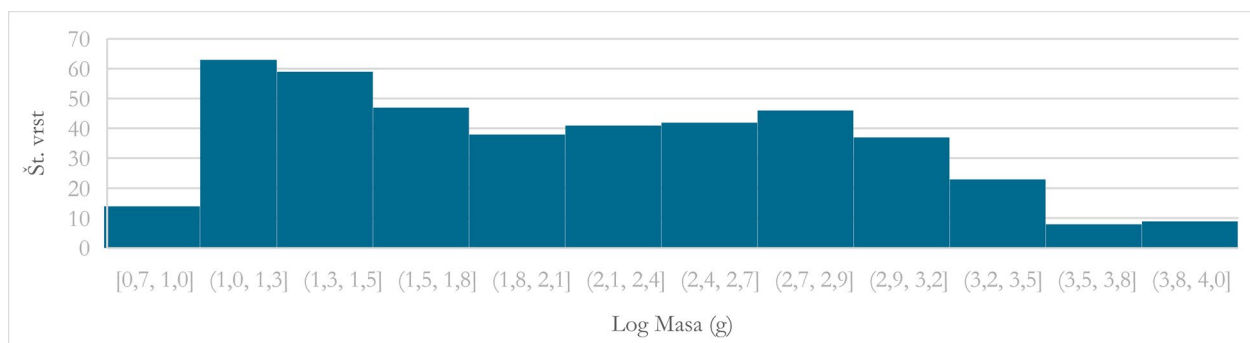
Vsebina je razdeljena v dve poglavji:

- Frekvenčna porazdelitev telesne velikosti.
- Odnos med telesno velikostjo in abundanco.

2.1 Frekvenčna porazdelitev telesne velikosti

Porazdelitev frekvence mase kopenskih vretenčarjev je odvisna od merila, v katerem fenomen opazujemo. V lokalnih razmerah (npr. v istem habitatu ali posamezni združbi) je porazdelitev praviloma močno desno asimetrična in z bistveno prevladujočim vrstnim bogastvom vrst lažjih od povprečja. V globalnem merilu na nivoju kontinentov se frekvenčna krivulja približuje normalni porazdelitvi. Vzorci porazdelitve velikosti se nekoliko razlikujejo med posameznimi skupinami vretenčarjev. Npr. pri ptičih je povprečna masa evropske gnezdilke 653 g, manjšo maso od povprečja ima 324 vrst oz. 75,7 % vrst ptičev, le 24,3 % vrst je težjih od povprečja. Frekvenčna porazdelitev mase ptičev izkazuje vzorec z dvema vrhoma, s prvim vrhom okrog vrednosti 20 g in drugim pri približno 636 g telesne mase (Slika 2.2).

Po podatkih Evropske okoljske agencije (European Environment Agency 2021) je v obdobju 2013 do 2018 v Evropi gneznilo 428 vrst ptičev. Njihova povprečna masa je bila v razponu od 5,2 g pri rdečeglavem kraljičku do 11.150 g pri labodu grbcu, s povprečjem 635,5 g, kar je povprečna masa kvakača.



Slika 2.2 Frekvenčna porazdelitev logaritmiranih vrednosti telesne mase 428 vrst evropskih ptičev gnezdilcev. Porazdelitev izkazuje dva vrhova, prvega pri masi okrog 20 g Log 1,2 (npr. kmečka lastovka *Hirundo rustica*) in drugega pri masi okrog 636 g Log 2,7 (npr. kvakač *Nycticorax nycticorax*). Podatki za izdelavo grafa so kombinacija poročila o evropskih gnezdilkah (European Environment Agency 2021) in podatkov o funkcionalnih znakih ptičev (Storchová, Hořák 2018).



Slika 2.3 Najmanjša in največja vrsta ptiča v slovenski ornitofavni. Razmerje med fotografijama ustreza dejanskemu razmerju v velikosti. Rumenooglavi kraljiček *Regulus regulus* (povprečna telesna dolžina 90 mm) zgoraj desno in labod grbec *Cygnus olor* (1.500 mm) z mladiči.

Foto: Gregor Bernard.

2.2 Odnos med telesno velikostjo in abundanco

Pomen odnosa med telesno velikostjo organizmov in njihovo abundanco je v ekologiji in evoluciji dobro prepoznan fenomen. Dejansko ne gre le za en, ampak za štiri različne kombinacije procesov, ki jih moramo analizirati ločeno. Da bi bolje razumeli, kako in zakaj sta abundanca in telesna velikost povezani, so v nadaljevanju predstavljeni vsi štirje tipi odnosa.

Globalni odnos med velikostjo in populacijsko gostoto GSDR (ang. Global Size-density Relationship) je najbolj prepoznaven vzorec odnosa med telesno velikostjo in gostoto osebkov (številom živali na površino, npr. število gnezdečih parov na km²). V globalnem merilu se odnos med povprečno maso vrste M_{sp} in njeno povprečno gostoto N_{sp} v povprečju zelo dobro prilagaja regresijski funkciji ($r^2 > 80\%$; Slika 2.4). Odnos je monotono padajoč, smerni koeficient regresijske premice je približno $\alpha = -0,75$; regresijska enačba ima obliko:

$$N_{sp} = cM_{sp}^{-\frac{3}{4}}.$$

Pri vrednosti c regresijska premica seka Y os. Opisan odnos je znan kot Damuthovo pravilo. To pravilo je za GSDR še posebej zanimivo, ker odnos med gostoto populacije in telesno velikostjo lahko zapišemo kot $N_{sp} \propto M_{sp}^{-3/4}$. Ta odnos je v proporcionalnem nasprotju glede na stopnjo metabolizma, ki je: $B \propto M_{sp}^{3/4}$. Rezultat tega odnosa je Damuthovo pravilo, po katerem so energetske potrebe populacije ($B * N_{sp}$) nasprotno sorazmerne njeni masi:

$$M_{sp}^{3/4} * M_{sp}^{-3/4} = M_{sp}^0 = 1.$$

To je tako imenovano »energetsko ekvivalentno pravilo« EER (ang. Energetic Equivalence Rule), ki navdušuje ekologe, zaradi vzpostavljanja zakonitosti med ekološkimi in fiziološkimi procesi.

Naslednja lastnost, ki podpira energetsko pravilo v GSDR, je vrednost koeficienta c . Slednji se značilno razlikuje med skupinami ektotermnih in endotermnih živali. Nižja vrednost c pri ektotermih je v povprečju proporcionalna prenosu energije med trofičnimi nivoji in istočasno inverzno proporcionalna razliki v stopnji metabolizma (White in sod. 2007).

Lokalni odnos med velikostjo in gostoto LSDR (ang. Local Size-density Relationship) izkazuje drugačen odnos kot GSDR. Gostota in telesna velikost izkazuje na lokalnem nivoju bistveno nižjo stopnjo povezanosti ($r^2 < 15\%$) kot na globalnem, smerni koeficient regresijske premice ima vrednost okrog $z \approx -0,25$. Projekcija vrednosti gostote na maso oblikuje razpršitev vrednosti v obliki trikotnika (Slika 2.4). Gostota telesno majhnih do srednje velikih vrst lahko močno variira, gostota telesno velikih vrst pa je vedno nizka. Zgornja vrednost odnosa LSDR je monotono omejena na odnos primerljiv z GSDR (Slika 2.4). Razliko v vzorcu porazdelitve med GSDR in LSDR lahko razložimo s tem, da GSDR predstavlja maksimalno zgornjo mejo lokalne abundance in privzema, da je gostota osebkov homogeno razporejena po celotnem prostoru. Lokalna variabilnost okolja ne omogoča homogene gostote, zato prihaja do znatnih razlik v gostoti izbrane vrste med nahajališči. Posledica nehomogene gostote vrst, ki je posledica ekoloških razmer v lokalnem okolju, oblikuje porazdelitev odnosa v obliki trikotnika, kot je predstavljeno na Sliki 2.4.

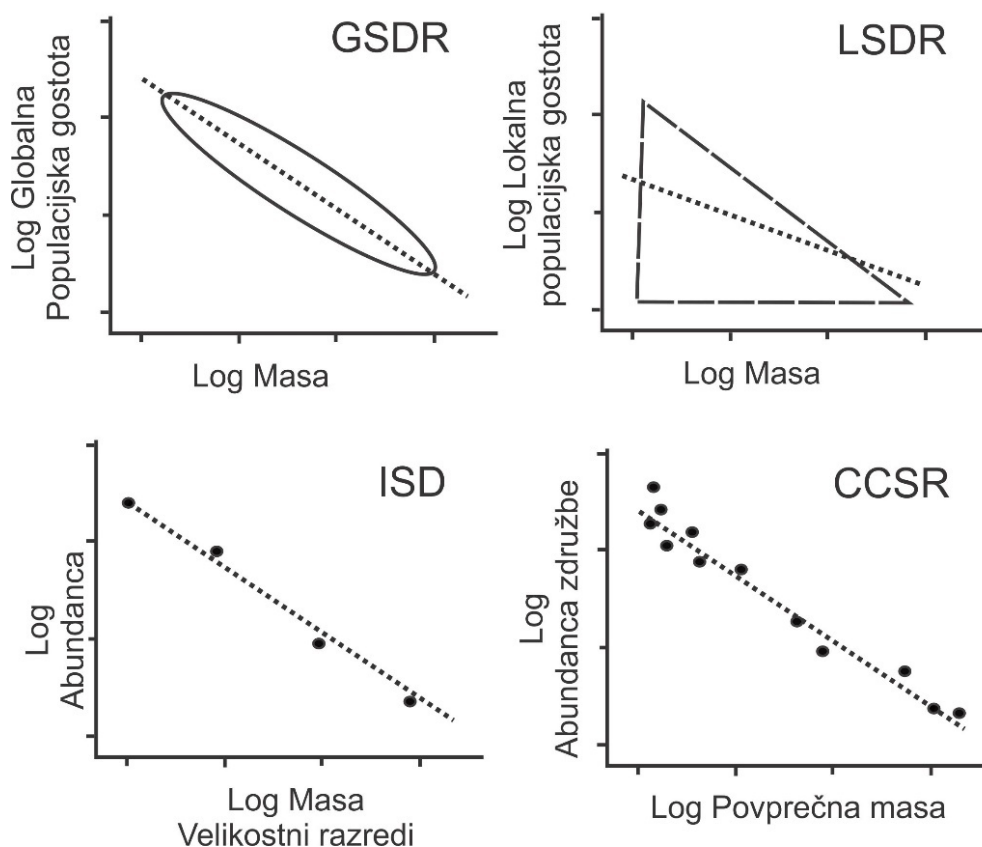
Individualne porazdelitve velikosti oz. velikostni spekter ISD (ang. Individual Size Distributions or Size Spectra) je frekvenčna porazdelitev telesne velikosti osebkov v združbi ne glede na njihovo vrstno pripadnost. Torej, n_i število osebkov razporedimo v j število velikostnih razredov s povprečno maso razreda \overline{m}_j . Posamezni razred lahko vsebuje predstavnike različnih vrst, prav tako je ena vrsta lahko zastopana v več velikostnih razredih. Odnos med abundanco in velikostjo je negativen, večje kot so živali manjša je abundanca (Slika 2.4). Takšen tip odnosa med velikostjo in abundanco se uporablja predvsem za vodne organizme, npr. ribe, ki rastejo celotno življenje zato ne moremo določiti natančne ocene velikosti odraslega osebkov.

Vzorec individualne porazdelitve nastopa v treh oblikah: 1. monotoni upad, kjer je najmanjši velikostni razred najbolj številčen, z vsakim naslednjim velikostnim razredom pa število osebkov linearno upada; 2. unimodalna porazdelitev, pri kateri je eden od vmesnih razredov najštevilčnejši; 3. multimodalna porazdelitev, porazdelitev ima več vrhov.

Odnos med združbami CCSR (ang. Cross-community Scaling Relationship) lahko opišemo kod zakonitost samo redčenja (ang. Self-thinning Rule). Znan je predvsem iz združb rastlin in intersticijalnih združb. Pri raziskovanju sesilnih združb lahko opazujemo, da z rastjo posameznega organizma ostaja vse manj prostora za druge posameznike. To lahko zapišemo kot negativni odnos med povprečno velikostjo organizma \overline{m}_{ind} in celotno abundanco združbe \overline{n}_{tot} (Slika 2.4). Analizo izvedemo bodisi z dolgoročnim spremljanjem združbe ali s primerjavo več združb v različnih starostnih stopnjah. Na splošno lahko odnos CCSR zapišemo v obliki zakonitosti:

$$n_{tot} \propto \overline{m}_{ind}^{-z}.$$

Eksponent z je podoben inverzni vrednosti eksponenta pri stopnji metabolizma. Odziv CCSR pri različnih sistemih oz. taksonih ima zelo podoben vzorec. Kjer je razpoložljivost naravnih virov oz. življenjskega prostora v posameznih lokalitetah konstantna, povprečna vrednost virov na enak način determinira število osebkov v populaciji.



Slika 2.4 Štirje tipi odnosa med telesno velikostjo in abundanco. GSDR – Globalni odnos med velikostjo in gostoto, LSDR – Lokalni odnos med velikostjo in gostoto, ISD – Individualna porazdelitev velikosti oz. velikostni spekter, CCSR – Odnos med združbami. Črtna črta predstavlja regresijsko premico.

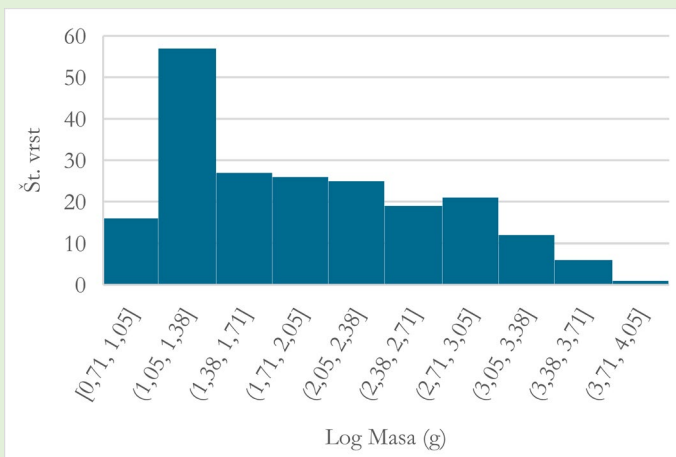


PRIMER 2.1

Frekvenčna porazdelitev telesne velikosti gnezdilc Slovenije

Iz Gnezditvenega atlasa Slovenije (Mihelič in sod. 2019) smo povzeli nabor 210 vrst ptičev, ki gnezdijo na ozemlju Slovenije. Za cenilko telesne velikosti smo privzeli povprečno maso vrste, kot jo med funkcionalnimi znaki navajata Storchová in Hořák (2018). Najlažja gnezdilka je rdečeglavi kraljiček s povprečno maso 5,2 g in najtežja labod grbec s povprečno maso 11.150 g (Slika 2.3), jerebica *Perdix perdix* z maso 383,4 g je povprečne velikosti.

Frekvenčna porazdelitev v velikostne razrede izkazuje močno desno asimetrično porazdelitev. Z vrstami najbogatejši razred je v razponu od 11 do 24 gramov (s povprečno maso 17 g), v katerega je uvrščenih 56 vrst (Slika 2.5). Povprečno maso 17 g ima taščica *Eirithacus rubecula* (Slika 2.6).



Slika 2.5 Frekvenčna porazdelitev telesne mase 210 vrst ptičev gnezdilcev v Sloveniji.



Slika 2.6 Taščica *Eirithacus rubecula* s povprečno maso 17,5 g je v razporeditvi telesne mase slovenskih gnezdilcev uvrščena v najštevilčnejši razred s povprečjem 17 g.

Foto. Gregor Bernard.



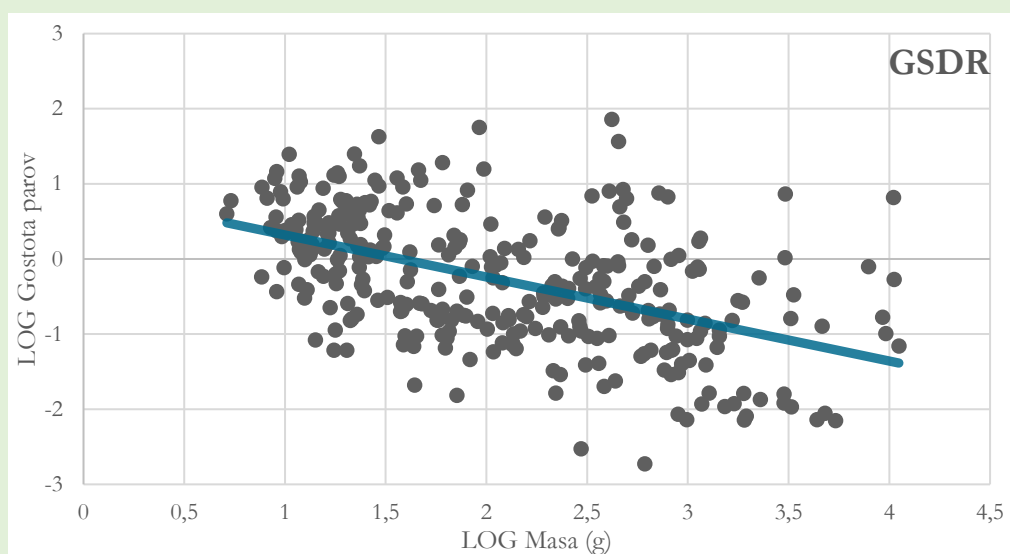


PRIMER 2.2

Odnos med telesno velikostjo in številom parov na površino – globalni vzorec

Iz nabora 427 vrst evropskih gnezdilk (European Environment Agency 2021) smo povzeli vrste, ki gnezdiijo na celinskem območju Evrope v geografski širini od 30 do 50 stopinj (Hagemeijer in Blait 1997). Seznam obsega 323 vrst. Podatke o velikosti geografskega območja (km²) in gostoti (število parov / km²) smo prav tako povzeli po Evropski okoljski agenciji (European Environment Agency 2021) in jih kombinirali s povprečno maso vrste, ki smo jo povzeli kot funkcionalni znak iz Storchová in Hořák (2018).

Odnos, ki ga opišeta povprečna gnezditvena gostota in povprečna masa, je linearen in monotono padajoč s smernim koeficientom $z = -0,57$ (Slika 2.7).



Slika 2.7 Primer odnosa med gostoto gnezdečih parov in povprečno maso vrste v globalnem merilu. Globalni odnos med abundanco in maso vrste GSDR na primeru evropskih gnezdilk.

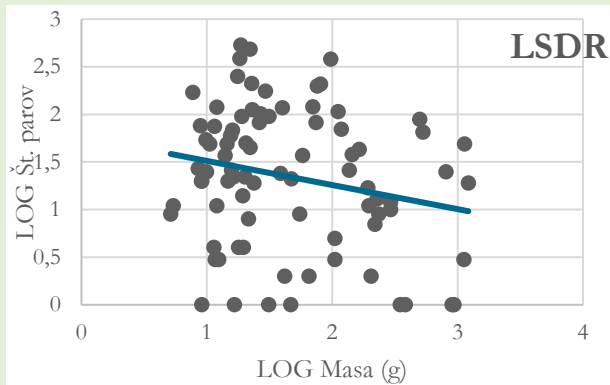


PRIMER 2.3

Odnos med telesno velikostjo in številom parov na površino – lokalni vzorec

Iz članka Kmecl in sod. (2014) smo povzeli podatke o številu gnezdečih parov ptic na območju Kozjanskega parka in jih kombinirali s povprečno maso vrste, ki smo jo povzeli kot funkcionalni znak iz Storchová in Hořák (2018).

Odnos, ki ga opišeta gostota in masa je na projekciji v obliki trapezoida (Slika 2.8), (zaradi odsotnosti največjih vrst manjka desna stran trikotnika). To je primer lokalnega vzorca odnosa med abundanco in maso živali, v katerem je variabilnost majhnih in srednje velikih vrst zelo velika. Zgornja meja abundance je omejena z linearno padajočo vrednostjo. Vrsta z najvišjo abundanco je črnoglavka *Sylvia atricapilla* (Slika 2.9).



Slika 2.8 Primer odnosa med abundanco in maso vrste v lokalnem merilu, LSDR. Lokalni odnos med abundanco in maso LSDR na primeru gnezdilk Kozjanskega parka.



Slika 2.9 Črnoglavka *Sylvia atricapilla* je na območju Kozjanskega parka vrsta z najvišjo gnezditveno abundanco; njena povprečna masa znaša 18,6 g.

Foto. Gregor Bernard.



PRIMER 2.4

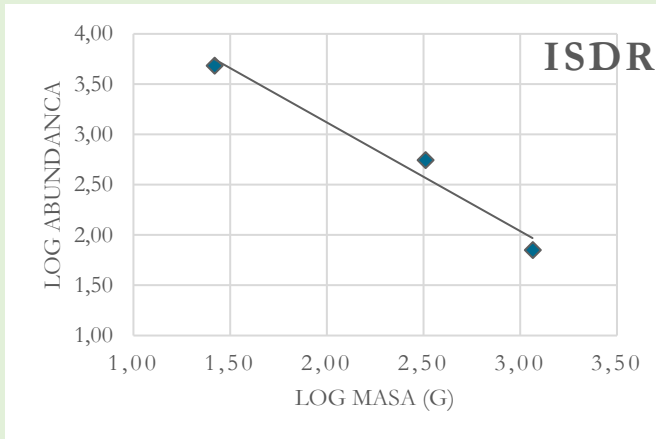
Odnos med telesno velikostjo in številom osebkov – individualni pristop

Podatke o številu gnezdečih parov na območju Kozjanskega parka (Kmecl in sod. 2014) smo kombinirali s povprečno maso vrste, ki smo jo povzeli iz Storchová in Hořák (2018).

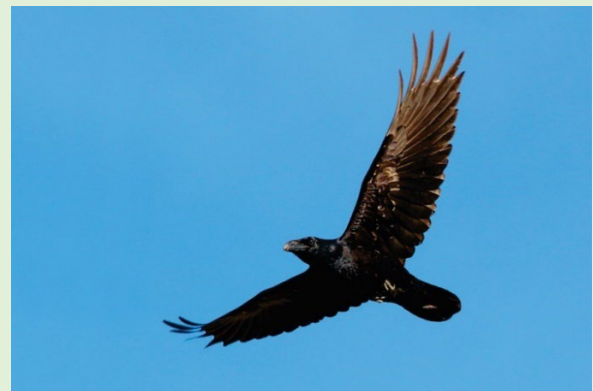
Posamezno vrsto smo uvrstili v enega od treh v naprej definiranih velikostni razredov (Preglednica 2.2). Število osebkov za posamezni velikostni razred smo izračunali kot vsoto vseh parov za vrste, ki so uvrščene v konkreten velikostni razred (Preglednica 2.2). Vrsta z najvišjo maso je krokar *Corvus corax* (Slika 2.11). Odnos med individualnimi frekvencami vrst in njihovo povprečno maso opiše linearna in monotono padajoča funkcija s smernim koeficientom $z = -1,1$ (Slika 2.10).

Preglednica 2.2 Razporeditev gnezdljk Kozjanskega parka v tri velikostne razrede, število vrst in skupno število osebkov v posameznem razredu

| Masa-razred | Povprečna masa (g) | LOG_Masa (g) | N-vrst | N-osebkov | LOG_N-osebkov |
|------------------------------|--------------------|--------------|--------|-----------|---------------|
| Lahki (do 99 g) | 26,3 | 1,42 | 58 | 4859 | 3,69 |
| Srednji (od 100 do 999 g) | 324 | 2,51 | 22 | 556 | 2,75 |
| Težki (nad 1.000 g) | 1157 | 3,06 | 3 | 71 | 1,85 |



Slika 2.10: Individualna obravnava osebkov v odnosu med abundanco in maso ISDR na primeru gnezdljk Kozjanskega parka.



Slika 2.11 Krokav *Corvus corax* je s povprečno maso 1.217 g najtežja vrsta ptiča na območju Kozjanskega parka.
Foto. Gregor Bernard.



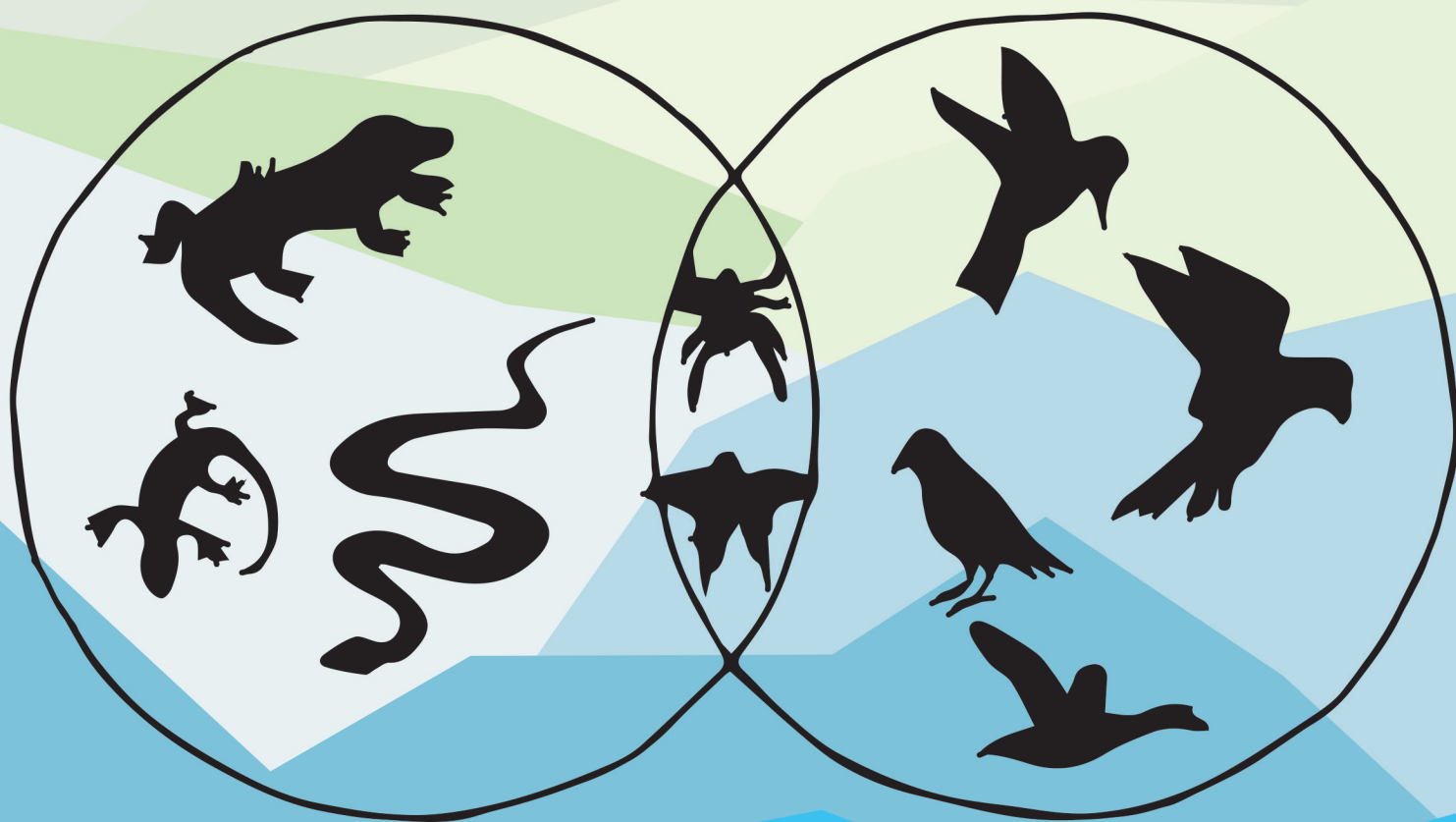
Zaključki

- Velikost telesa je ena od osnovnih lastnosti organizma. S telesno velikostjo so povezane druge biološke lastnosti, npr.: življenjska doba, velikost domačega okoliša, razmerja med telesnimi organi oz. organskimi sistemi, način premikanja in drugi vidiki življenja in ekologije.
- Odnos med telesno velikostjo in abundanco organizmov izraža povezavo med nivojem osebka in nivojem populacije. Slednje vpliva na strukturo vrstne sestave in na dinamiko ekološke združbe.
- Telesna velikost je ključna za presnovne procese in posledično tudi za porabo naravnih virov (hrane, prostora ...), zato odnos med velikostjo in abundanco razkriva porazdelitev virov znotraj ekološkega sistema.
- Frekvenčna porazdelitev telesne velikosti je posledica več energetske evolucijskih mehanizmov: stopnja speciacije, stopnja izumiranja, stopnja disperzije, interspecifična kompeticija in energetske omejitve. Vzorec porazdelitev je odvisen tudi od merila, v katerem ga opazujemo. V lokalnih razmerah (npr. habitat ali posamezna združba) je porazdelitev praviloma zelo desno asimetrična in z bistveno prevladujočim vrstnim bogastvom vrst lažjih od povprečja. V globalnem merilu na nivoju kontinentov pa se približuje normalni porazdelitvi.
- Odnos med telesno velikostjo organizmov in njihovo abundanco prepoznavamo na štiri nivojih: GSDR, LSDR, ISD, CCSR.
- Globalni odnos med velikostjo in gostoto GSDR je monotono padajoč in se v povprečju dobro ujema z regresijsko funkcijo s smernim koeficientom $-0,75$.
- Lokalni odnos med velikostjo in gostoto LSDR izkazuje bistveno nižjo stopnjo povezanosti, smerni koeficient regresijske premice ima vrednost okrog $-0,25$. Projekcija vrednosti gostote na maso oblikuje trikotnik, kar je posledica velike variabilnosti gostote majhnih do srednje velikih vrst.
- Individualna porazdelitev velikosti ISD je frekvenčna porazdelitev telesne velikosti osebkov v združbi, ne glede na vrstno pripadnost. Osebki so razporejeni v ustrezen velikostni razred, posamezni razred lahko vsebuje predstavnike različnih vrst, prav tako je ena vrsta lahko zastopana v več velikostnih razredih.
- Odnos med združbami CCSR lahko opišemo kod samoredčenje, znan je predvsem pri sesilnih združbah. Rast posameznega organizma povzroči vse manj prostora za druge posameznike. Gre za negativni odnos med povprečno velikostjo organizma in celokupno abundanco združbe.

Literatura

- Brown J.H., Nicoletto P.F. 1991. Spatial scaling of the species composition - body masses of North-American land mammals. *American Naturalist*, 138(6): 1478–1512.
- European Environment Agency. 2021. Status and trends of bird populations: datasets from Article 12, Birds Directive 2009/147/EC reporting. Prezeno 8. 6. 2022, iz: <https://www.eea.europa.eu/data-and-maps/data/article-12-database-birds-directive-2009-147-ec-1>.
- Gaston K.J., Blackburn T.M., Lawton J.H. 1997. Interspecific abundance-range size relationships: an appraisal of mechanisms. *Journal of Animal Ecology*, 66: 579–601.
- Hagemeijer E.J.M., Blair M.J. 1997. *The EBCC Atlas of European Breeding Birds: Their Distribution and Abundance*. T&AD Poyser. London.
- Kmecl P., Jančar T., Mihelič T. 2014. Spremembe v avifauni Kozjanskega parka med letoma 1999 in 2010: velik upad števila travniških ptic. *Acrocephalus*, 35(162/163): 125–138.
- Maurer B.A., Brown J.H., Rusler R.D. 1992. The micro and macro in body size evolution. *Evolution*, 46(4): 939–953.
- May R. 1988. How many species are there on earth? *Science*, 241(4872): 1441–1449.
- Mihelič T., Kmecl P., Denac K., Koče U., Vrezec A., Denac D. (eds.) (2019). *Atlas ptic Slovenije. Popis gnezdilk 2002–2017*. DOPPS. Ljubljana.
- Storchová L., Hořák D. 2018. Life-history characteristics of European birds. *Global Ecology and Biogeography*, 27(4) DOI: 10.1111/geb.12709.
- White E.P., Ernest S.K.M., Kerckhoff J.A., Enquist J.B. 2007. Relationships between body size and abundance in ecology. *Trends in Ecology and Evolution*, 22(6): 323–330.

3. poglavje



VRSTNO BOGASTVO

Povzetek

3. poglavje: Vrsto bogastvo

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

Vrstno bogastvo je skupno število vrst v vzorcu, v združbi ali v regiji. Vrsto bogastvo izmerimo z ugotavljanjem prisotnosti vrst. Z večanjem vloženega raziskovalnega dela zmanjšamo pristransko podcenjenost števila vrst. Raziskanost združbe ocenimo z akumulacijsko krivuljo vrst. Ko krivulja doseže asimptoto, je raziskanost vrstnega bogastva ustrezna. Primerjave vrstnega bogastva med združbami, ki so raziskane z različnim vložkom raziskovalnega dela, izvajamo z redčenjem. To je z oceno števila vrst posameznih združb, ki ga standardiziramo na enako količino raziskovalnega dela. Standardizacija raziskanosti je lahko izvedena glede na število osebkov ali glede na vloženo delo (število vzorčenj, čas vzorčenja). Število vzorčenj in število osebkov sta tesno povezani. Pri popisovanju biote z enako intenzivnim in konstantnim vzorčenjem, kumulativno število prešteti osebkov linearno narašča skupaj s kumulativnim številom vzorcev. Združbo raziskujemo z vidika lokalne – alfa in regionalne – gama biodiverzitete. Znotraj ene regije je prisotnih več lokalnih alfa diverzitet, ki skupaj sestavljajo gama diverzitetu. Biodiverzitetu v prostoru in času ni konstantna, v prostorskem in časovnem gradientu se spreminja tako število vrst, kot njihova sestava. Prostorsko ali časovno variabilnost v vrstni sestavi znotraj regije (oz. spreminjanje v številčnosti in sestavi vrst med lokalnimi vzorci) imenujemo beta diverzitetu, vzroki zanjo so okoljski in biotski. Primerjave med posameznimi alfa diverzitetami ocenjujemo

s cenilkami parnih primerjav, ob hkratni primerjavi treh ali več lokalnih diverzitet jih vrednotimo s cenilkami mnogoterih primerjav. Alternativa parnim primerjavam je analiza zeta diverzitete, kjer istočasno analiziramo spremembo v vrstni sestavi med mnogo lokalitetami. Vzorčenje biote z namenom evidentiranja vrstnega bogastva oz. biodiverzitete izbrane skupine organizmov načrtujemo v skladu z optimalnim vložkom raziskovalnega dela, da ujamemo ravnotežje med ustrezno raziskanostjo ter časovno in materialno investicijo v izvedbo vzorčenja. V analizi biodiverzitete smo osredotočeni na prisotnost vrst, pomembno informacijo pridobimo tudi iz manjkajočih vrst. Pričakovane, vendar manjkajoče vrste v združbi, imenujemo temna diverzitetu. Koncept temne diverzitete temelji na manjkajočih in pričakovanih vrstah in vključuje: nabor vrst, tarčno združbo in filtriranje vrst. Temno diverzitetu izrazimo kot razliko med pričakovanim in opazovanim vrstnim bogastvom, njeno relativno vrednost pa z indeksom popolnosti združbe. Z Bealovim indeksom podamo verjetnost prisotnosti vrste na lokaliteti. Prikrute vrste so skupina sorodnih organizmov, ki so si fenotipsko zelo podobni, morfološke razlike med njimi niso zaznavne ali pa so zelo slabo izražene. Po navadi so prikrute vrste monofiletska skupina v zgodnji fazi speciacije, lahko pa gre tudi za vrste, ki so ločene že dolgo časovno obdobje, vendar pri njih ni prišlo do morfološke diferenciacije.

Vrstno bogastvo (ang. Species Richness) je skupno število vrst v vzorcu, v združbi ali v regiji. **Vzorec** (ang. Sample) v biodiverzitetni raziskavi je enkratni popis združbe izbranih organizmov na definiranem območju in v definiranem času. Vzorec lahko spremljajo meritve drugih biotskih in abiotskih dejavnikov, povezanih z vzorcem, praviloma gre za spremenljivke, s katerimi želimo razložiti lastnosti in variabilnost v združbi. V biodiverzitetnih raziskavah se po navadi usmerimo v izbrano skupino organizmov in jo poimenujemo **združba** (ang. Community). V biodiverzitetni raziskavi se osredotočimo na združbo iz istega trofičnega nivoja, ki predstavlja manjši del življenjske združbe (biocenoze) oz. samo del celotne skupnosti živih organizmov. Ob terenskem popisovanju izbrane združbe organizmov v njihovem habitatu (npr. malih sesalcev v bukovem gozdu, prezimujočih ptic na jezeru ali metuljev na tradicionalno košenem travniku), geografski prostor na katerem poteka raziskava združbe, imenujemo območje. **Območje** je v biodiverzitetni raziskavi geografski prostor, na katerem analiziramo lastnosti združbe, lahko je dobro prepoznavno in definirano z biotopskimi in krajinskimi elementi, in tako razmejeno od drugih območij (npr. Pohorje). Lahko je razmejeno s političnimi mejami (npr. območje občine Duplek ali območje Evropske skupnosti). **Čas** je obvezna spremenljivka v biodiverzitetni raziskavi, navajamo ga z datumom (dd.mm.yyyy) ali časovnim obdobjem v katerem je potekalo vzorčenje.

Vrstno bogastvo določimo z ugotavljanjem prisotnosti (prezenci), tj. s prešteti vrst v vzorcu, pri čemer ne upoštevamo njihove številčnosti (abundance) oz. njihovih deležev (relativne abundance). Štetje vrst je po navadi pristransko, vrednost je podcenjena glede na dejansko število vrst v združbi. S povečanjem raziskovalnega dela (s preštevanje večjega števila osebkov, pregledovanjem večjega števila vzorčenih enot ali z vzorčenjem na večjem območju) se namreč večja število odkritih vrst.

V analizi biodiverzitete se v glavnem osredotočamo na prisotnost vrst, manjkajočim vrstam pa ne posvečamo posebne pozornosti. Z drugimi besedami ali manjkajoče vrste dopolnjujejo informacijo o biodiverziteti? **Temna diverziteteta** (ang. Dark Diversity) je nabor vrst, ki na lokaliteti manjkajo, vendar jih potencialno pričakujemo. Besedna zveza »temna diverziteteta« je nastala po analogiji s »temno snovjo« v fiziki. Temne snovi ne vidimo, lahko pa izmerimo njen vpliv.

Analiza vrstnega bogastva obsega različne vidike obravnave, ki so deloma predstavljeni v tem poglavju, deloma v 4. poglavje Vrstna pestrost in v 7. poglavju Analiza združbe.

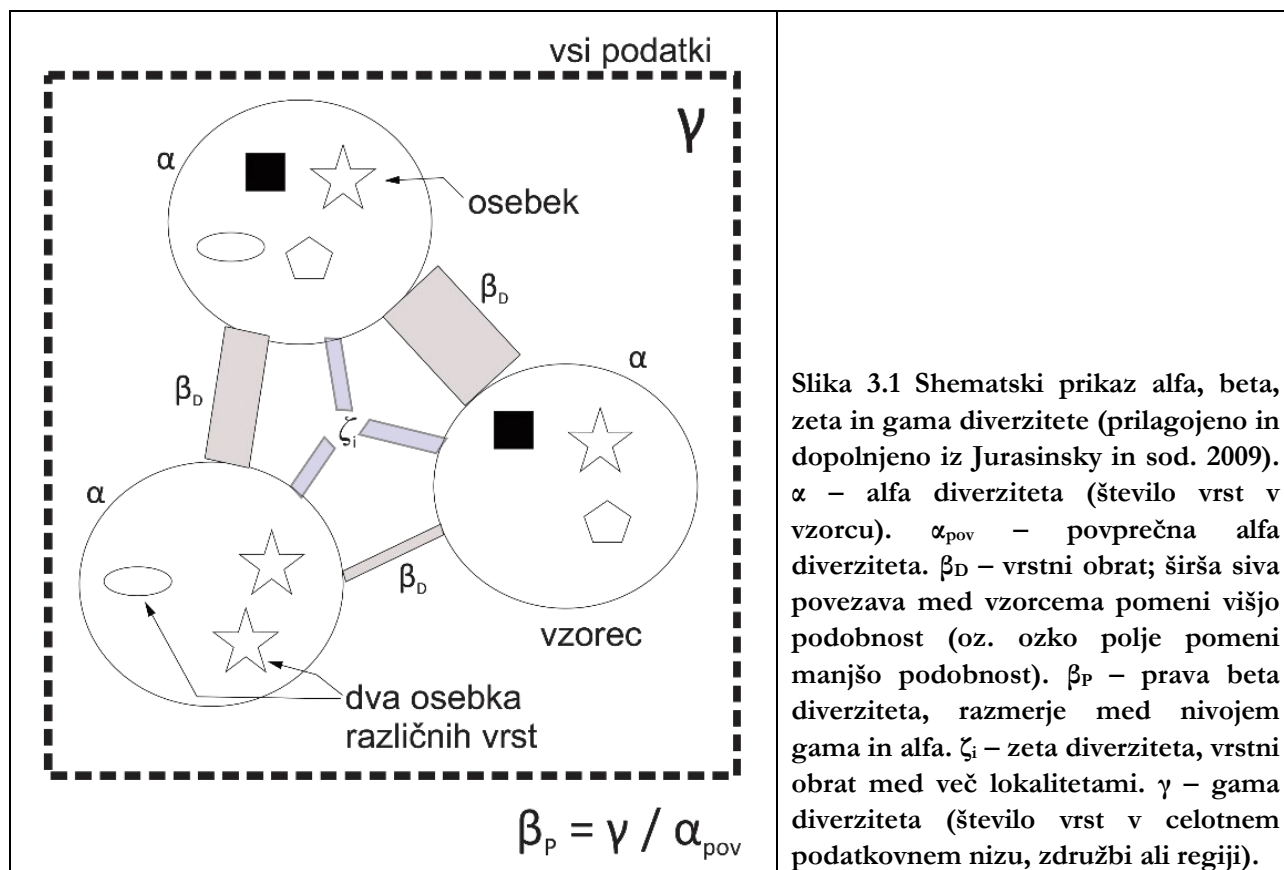
Literatura s tega področja je obsežna in se nanaša na teorijo bogastva vrst kot tudi na analitične pristope in konkretna raziskovalna poročila. V seznamu literature je navedenih nekaj temeljnih virov. V tem poglavju so predstavljene naslednje vsebine:

- Prostorska komponenta vrstnega bogastva.

- α alfa diverziteta.
- β beta diverziteta.
- γ gama diverziteta.
- vrstni obrat in ugnezdenost.
- ζ zeta diverziteta.
- Ocena vrstnega bogastva.
 - Akumulacijska krivulja vrst in redčenje.
- Temna diverziteta in prikrita diverziteta.

3.1 Prostorska komponenta vrstnega bogastva

Kljub navidezni razumljivosti koncepta vrstnega bogastva sta njegova interpretacija in analiza večplastni. Vrstno bogastvo razumemo kot število vrst v združbi ali na določenem območju (lokalni nivo), kar imenujemo **točkovna diverziteta** oz. α diverziteta (alfa diverziteta) (Tome 2006). Različno sestavo vrst med združbami imenujemo **vrstni obrat**. To sta β diverziteta (beta diverziteta) in ζ diverziteta (zeta diverziteta). Tretji nivo je skupno **število vrst v celotni regiji** oz. v skupini obravnavanih združb (regionalni nivo), kar imenujemo γ diverziteta (gama diverziteta). V primeru, da v vrednoteni združbi ni podvzorcev (razpolagamo s samo enim vzorcem), v tem primeru γ diverziteta ustreza α diverziteti (Slika 3.1). Analiziranje vrstnega bogastva je predstavljeno na raziskavah suhih južin v visokogorskih habitatih, pristop je podrobneje predstavljen v Projektu 3.1.



Slika 3.1 Shematski prikaz alfa, beta, zeta in gama diverzitete (prilagojeno in dopolnjeno iz Jurasinsky in sod. 2009). α – alfa diverziteta (število vrst v vzorcu). α_{pov} – povprečna alfa diverziteta. β_D – vrstni obrat; širša siva povezava med vzorcema pomeni višjo podobnost (oz. ozko polje pomeni manjšo podobnost). β_P – prava beta diverziteta, razmerje med nivojem gama in alfa. ζ_i – zeta diverziteta, vrstni obrat med več lokalitetami. γ – gama diverziteta (število vrst v celotnem podatkovnem nizu, združbi ali regiji).



PROJEKT 3.1

Vrstno bogastvo suhih južin Opiliones v alpskih visokogorskih habitatnih tipih

Novak in sodelavci (2017) so izvedli popise suhih južin (Slika 3.3) v petih habitatnih tipih (HT) v Kamniško-Savinjskih Alpah (Preglednica 3.1). Prezenco in abundanco vrst na posameznih lokacijah (tj. vzorčenih enotah) so vnesli v Excelovo preglednico. Vzorčenje je potekalo zaradi ugotavljanja, katere habitatne tipe poseljujejo posamezne vrste suhih južin. **Metode zbiranja podatkov** so bile razdeljene na terensko delo, laboratorijsko delo in delo z računalnikom. **Na terenu** je več raziskovalcev z opazovanjem iskalo suhe južine z delom ca. 70 minut/oseba/vzorčeno enoto (Slika 3.2). V **laboratoriju** so s pomočjo literature determinirali osebke do vrste. Zbrane podatke, ločeno za prezenco in abundanco posameznih vrst v posameznih vzorčenih enotah, so uredili v Excelovo tabelo **za računalnikom**. Ti dve preglednici omogočata številne vidike analize biodiverzitete, na njih so predstavljeni principi in načini obdelave podatkov o vrstnem bogastvu.



Slika 3.2 Vzorčenje suhih južin v visokogorskih habitatih Kamniško–Savinjskih Alp. Pri opazovanju so pregledovali vegetacijo, jo odstirali, obračali kamne in pregledovali mikrohabitate (skalne stene, grušč, gola tla itd.). Kjer je bil na voljo drobir oz. sipek substrat, so ga približno 5 l presejali skozi mrežo z odprtini 1,5 x 1,5 mm. V sejancu so iskali suhe južine, ki živijo v substratu.

Foto. Tone Novak.



Slika 3.3 Suha južina vrste navadni širokookec *Platybunus bucephalus*.

Foto. Peter Kozel.



3.1.1 Alfa diverziteta in gama diverziteta.

Z α diverzitetu podamo število vrst na definiranem **območju** (v eni vzorčni enoti), z γ diverzitetu podamo število vrst v regiji, ki praviloma zajema več vzorčenih enot (oz. območij). V Primerih 3.1 in 3.2 je predstavljena alfa in gama diverziteta.



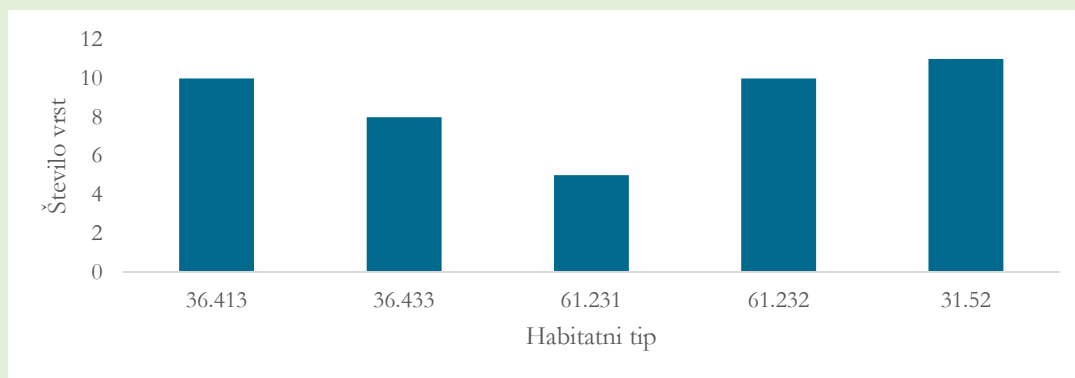
PRIMER 3.1

Suhe južine v visokogorskih habitatih

V posameznem habitatnem tipu v visokogorju so našli od 5 do 11 vrst suhih južin. Tabelarično (Preglednica 3.1) in grafično (Slika 3.4) je predstavljeno število vrst suhih južin Opiliones v petih habitatnih tipih HT. Oznaka HT in njihovo poimenovanje je dogovorjena z EU klasifikacijo habitatnih tipov, predstavljeno na [www: https://eunis.eea.europa.eu/index.jsp](https://eunis.eea.europa.eu/index.jsp).

Preglednica 3.1 Število vrst suhih južin – alfa diverziteta za posamezni habitatni tip HT

| Oznaka HT | Habitatni tip | Št. vrst |
|-----------|--|----------|
| 31.52 | Ruševje zunanjih severnih in jugovzhodnih Alp | 11 |
| 36.413 | Alpiska in subalpiska travišča na karbonatni podlagi v Južnih Apneniških Alpah | 10 |
| 36.433 | Blazinasta travišča čvrstega šaša | 8 |
| 61.231 | Montanska do subalpiska karbonatna melišča | 5 |
| 61.232 | Alpiska karbonatna melišča | 10 |



Slika 3.4 Alfa diverziteta – število vrst suhih južin v posameznem habitatnem tipu. Poimenovanje habitatnih tipov je v Preglednici 3.1.

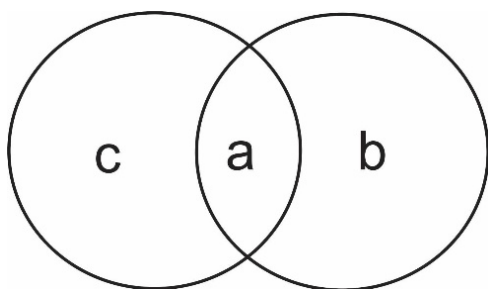
Vrednost gama diverzitete enačimo s celokupnim vrstnim bogastvom, ki v tem primeru znaša 14 vrst.



3.1.2 Beta diverziteteta

Z β diverzitetu vrednotimo razlike v vrstni sestavi med dvema območjema ali istočasno med več območji. Beta β diverziteteta je posledica razlik v določenem ekološkem dejavniku oz. njihovi kombinaciji (fizikalne, kemijske ali biotske narave) ali pa vzorčenju v različnih habitatih. Računamo jo kot indeks raznolikosti med dvema območjema (lokalnima združbama) znotraj iste regije. V literaturi je znanih več kot 20 cenilk za izračun beta diverziteteta. Primerjave razlik med območji lahko izvajamo v parih – **parne primerjave** (ang. Pairwise Dissimilarities) ali pa primerjamo več območij istočasno – **mnogotere primerjave** (ang. Multiple-site Dissimilarities). V osnovi računamo β diverzitetu s primerjanjem vrstne sestave dveh lokalnih vzorcev, pri čemer upoštevamo prisotnost/odsotnost – PA (ang. Presence / Absence) vsake vrste. Cenilke s katerimi primerjamo podobnost ali različnost med dvema združbama, računamo iz PA matrike. Pri primerjanju dveh območij eno opredelimo kot referenčno območje in njegovo vrstno sestavo primerjamo z vrstno sestavo sosednjega območja. Dobljene vrednosti označimo s črkami a , b in c , pri čemer vrednost a predstavlja število vrst, ki je skupno obema območjema, vrednost b število vrst, prisotnih le v sosednjem območju, in c število vrst, prisotnih le v referenčnem območju (Slika 3.5). Predstavitev analize beta diverziteteta je predstavljen v Primerih 3.2 in 3.3.

Poglavje o beta diverziteti temelji na virih Baselga 2010, Beta diversity 2019, Koleff in sod. 2003, Oksanen 2022 in Oksanen in sod. 2019.



Prostorska porazdelitev vrst na dveh primerjanih območjih. Komponenta a predstavlja število vrst, skupnih obema območjema; komponenta b predstavlja število vrst, opazovanih samo na sosednjem območju; komponenta c predstavlja število vrst, opazovanih samo na referenčnem območju.

Slika 3.5 Vennov diagram prostorske razporeditve vrst na dveh primerjanih območjih.

Cenilke parnih primerjav. S cenilkami parnih primerjav (ang. Pair-wise Measures) primerjamo vrstni obrat med dvema območjema. Glede na odziv posamezne cenilke na variabilnost posamezne komponente Vennovega diagrama (a , b in c) razlikujemo štiri skupine cenilk beta diverziteteta:

1. cenilke kontinuitete in izgube: β_{rlb} ;

2. cenilke gradientov vrstnega bogastva: β_{gl} ;
3. cenilke kontinuitete: β_j , β_{sor} , β_w , β_{-1} , β_c , β_t , β_{hk} , β_z , β_m ;
4. cenilke naraščanja in izgube: β_{co} , β_r , β_l , β_e , β_{rs} , β_{-2} , β_{-3} , β_{sim} .

Vseh cenilk je 24, nekaj jih je med seboj identičnih, različnih je dejansko 18, identične cenilke so označene.

Cenilke kontinuitete in izgube – od vseh 24 cenilk je samo β_{rlb} odvisen le od vrednosti a in c . Cenilka ima vrednost nič, 0, (popolna različnost), ko si dve območji ne delita nobene skupne vrste ($a = 0$), in vrednost ena, 1 (popolna podobnost), ko sta vrednost $c = b = 0$. Cenilko računamo po formuli:

$$\beta_{rlb} = \frac{a}{a+c}.$$

Cenilke gradienta vrstnega bogastva – vrednost β_{gl} je odvisna od razlike v vrstnem bogastvu med dvema območjema. Cenilka β_{gl} ima najnižjo vrednost, kadar si dve območji ne delita nobene vrste ($a = 0$), komponenti c in b pa sta izenačeni ($b = c$). Vrednost cenilke narašča z zmanjševanjem deleža a in b komponente. Cenilko računamo po formuli:

$$\beta_{gl} = \frac{2|b-c|}{2a+b+c}.$$

Cenilke kontinuitete – izkazujejo vrstni obrat v »širšem pomenu«, poudarjeno imajo raznolikost v sestavi združbe, zato bolj izkazujejo gradient v sestavi, zapostavljen pa je relativni pomen naraščanja ali izgube vrst. Razlikujemo dve skupini cenilk, neposredno sta odvisni od komponente a , to je skupnih vrst obema območjema. V prvo skupino cenilk kontinuitete uvrščamo β_j in β_{sor} , njuni vrednosti sta v pozitivni povezavi s komponento a . Ko komponenta a narašča, naraščata tudi vrednosti cenilk, zato se tudi imenujeta cenilki podobnosti. Visoka vrednost cenilke pomeni nizko beta diverziteteto oz. visoko podobnost med združbama in obratno. (V to skupino cenilk uvrščamo tudi identična $\beta_g = \beta_{cc}$, zanju velja, da je $\beta_j = 1 - \beta_{cc}$). Cenilki računamo po formulah:

$$\beta_j = \frac{a}{a+b+c},$$

$$\beta_{sor} = \frac{2a}{2a+b+c}.$$

V drugo skupino cenilk kontinuitete uvrščamo β_w , β_{-1} , β_c , β_{wb} , β_l , β_t , β_{hk} , β_z , β_{me} , β_m . Njihove vrednosti so v negativni povezavi s komponento a . Ko komponenta a narašča, se vrednosti cenilk zmanjšujejo. Cenilko β_w najpogosteje uporabimo za prikaz beta diverzitetete. Nekaj opomb: $\beta_w = \beta_{-1} = \beta_t$, $\beta_c = \beta_l$, $\beta_t = \beta_{hk} = \beta_{me}$. Cenilke računamo po formulah:

$$\beta_w = \frac{b + c}{2a + b + c}$$

$$\beta_c = \frac{b + c}{2}$$

$$\beta_{wb} = b + c$$

$$\beta_z = \frac{\log 2 - \log(2a + b + c) + \log(a + b + c)}{\log 2}$$

$$\beta_m = \frac{(2a + b + c)(b + c)}{a + b + c}$$

Cenilke naraščanja in izgube – v to skupino uvrščamo β_{co} , β_r , β_l , β_e , β_{rs} , β_{-2} , β_{-3} in β_{sim} (β_{rs} je v R->vegan označen z »19«), vrednosti teh cenilk so odvisne od komponente a in relativne velikosti komponent b in c . Ta skupina cenilk izkazuje vrstni obrat v »ožjem pomenu«, pri izračunih je poudarek na razlikah v vrstni sestavi, zanemarjene pa so razlike v vrstnem bogastvu obravnavanih združb. Vrednosti vseh teh cenilk naraščajo z zmanjševanjem komponente a , vrh pa dosežejo z izenačevanjem deležev komponent b in c ; ter nasprotno, vrednosti cenilk se zmanjšujejo ob naraščanju komponente a , in hkratnem povečevanju neenakosti med komponentama b in c . Cenilke računamo po formulah:

$$\beta_{co} = \frac{ac + ab + 2bc}{2(a + b)(a + c)}$$

$$\beta_r = \frac{2bc}{(a + b + c)^2 - 2bc}$$

$$\beta_l = \log(2a + b + c) - \frac{2a \log 2}{2a + b + c} - \frac{(a + b)(\log(a + b) + (a + c)\log(a + c))}{2a + b + c}$$

$$\beta_e = e^{\log(2a + b + c)} - \frac{2a \log 2}{2a + b + c} - \frac{(a + b)\log(a + b) + (a + c)\log(a + c)}{(2a + b + c) - 1}$$

$$\beta_r = \frac{\frac{2(bc + 1)}{a + b + c}}{a + b + c - 1}$$

$$\beta_{-2} = \frac{\min(b, c)}{\max(b, c) + a}$$

$$\beta_{-3} = \frac{\min(b, c)}{a + b + c}$$

$$\beta_{sim} = \frac{\min(b, c)}{\min(b, c) + a}$$

S **cenilkami mnogoterih primerjav** (ang. Multiple-site Measures) hkrati vrednotimo vrstni obrat na nivoju celotne regije, tj. med tremi ali več vzorčenimi območji. Izvirna cenilka za oceno vrstnega obrata na nivoju regije se imenuje **prava diverziteta β** (ang. True β Diversity) in predstavlja razmerje med skupnim številom vrst v regiji (γ diverziteta) in povprečnim številom vrst na vseh lokalitetah (povprečna a diverziteta). Vrstni obrat na nivoju regije je uvedel Whittaker, zato ga imenujemo **Whittakerjeva diverziteta** in ga izračunamo po formulah (Whittaker je predlagal dve obliki izračuna):

$$\beta_W = \frac{\gamma}{\alpha_p} \quad \text{in} \quad \beta_W = \frac{\gamma}{\alpha_p} - 1,$$

kjer so β_W – Whittakerjeva β diverziteta, γ – gama diverziteta, α_p – povprečna vrednost diverzitete $a = \sum \alpha / n$; n – število združb.

Sørensenov indeks mnogoterih raznolikosti (β_{SOR}) in **Simpsonov indeks mnogoterih raznolikosti (β_{SIM})**:

$$\beta_{SOR} = \frac{[\sum_{i < j} \min(b_{ij}, b_{ji})] + [\sum_{i < j} \max(b_{ij}, b_{ji})]}{2[\sum_i S_i - S_T] + [\sum_{i < j} \min(b_{ij}, b_{ji})] + [\sum_{i < j} \max(b_{ij}, b_{ji})]},$$

$$\beta_{SIM} = \frac{[\sum_{i < j} \min(b_{ij}, b_{ji})]}{[\sum_i S_i - S_T] + [\sum_{i < j} \min(b_{ij}, b_{ji})]},$$

kjer so β_{SOR} – Sørensenov indeks mnogoterih raznolikosti, β_{SIM} – Simpsonov indeks mnogoterih raznolikosti, b_{ij} – število skupnih vrst v vzorcu i in j , ko ju medsebojno primerjamo, S_i – število vseh vrst v vzorcu i , S_T – skupno število vseh vrst iz vseh vzorcev.

Prostorski obrat in ugnezenost – β diverziteta vključuje dve komponenti, komponento **prostorskega obrata** (ang. Spatial Turnover), ki ga povzema Simpsonov indeks raznolikosti (β_{sim}), in komponento **ugnezdenosti** (ang. Nestedness), ki ga povzema indeks ugnezdene raznolikosti, (β_{nes}) (ang. Nestedness Dissimilarity). Komponenta prostorskega obrata se nanaša na vrednotenje vrstnega bogastva in primerjave v različnosti vrstne sestavi združb dveh ali več vzorcev. Vrednotimo razliko v številu različnih vrst, prisotnih v prvem in drugem vzorcu. To so vrste, ki niso prisotne v prvem vzorcu, v drugem vzorcu pa so prisotne in nadomestijo manjkajoče vrste iz prvega vzorca; zato govorimo o **vrstnem obratu** (ang. Species Turnover). Komponenta ugnezdenih vrst se nanaša na razlike med vzorci zaradi manjkajočih vrst. To so vrste, ki so v prvem vzorcu prisotne, v drugem pa manjkajo. Komponenta ugnezdenih vrst je pomembna predvsem pri primerjanju več vzorcev, ko izvajamo večkratne parne primerjave in nas zanimajo vzorci z manjkajočimi vrstami. V konceptu ugnezdenosti interpretiramo vrstno revnejša območja kot podskupino območja z večjim vrstnim bogastvom, oz. v kontekstu nabora vrst celotnega območja (Slika 3.6).

Obe komponenti β diverzitete, tako prostorski obrat kot ugnezenost, sta zajeti v Sørensenovem indeksu raznolikosti (β_{sor}), zato velja:

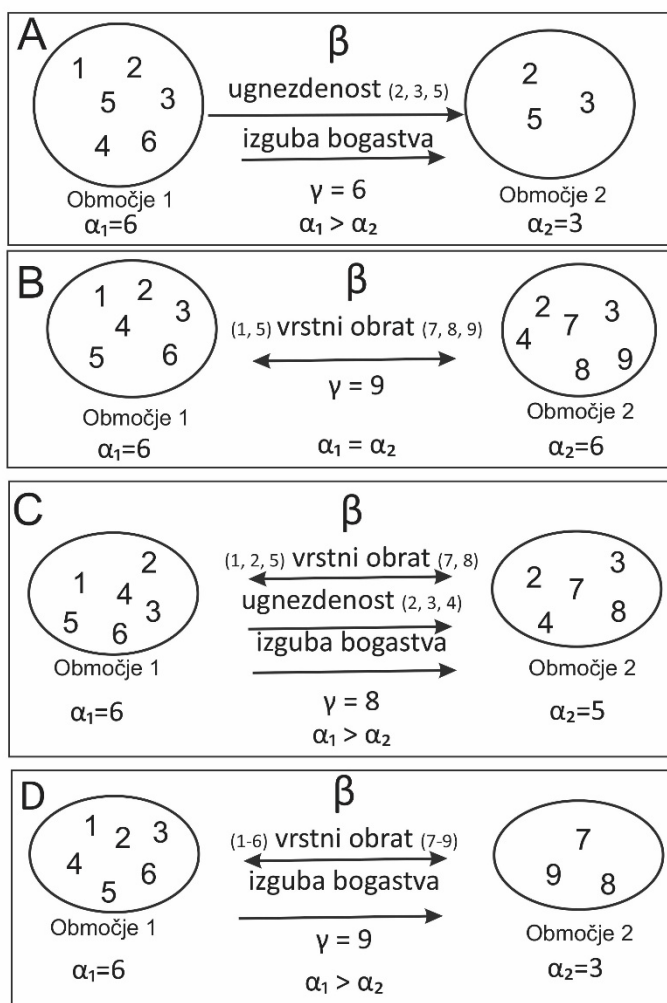
$$\beta_{\text{sor}} = \beta_{\text{sim}} + \beta_{\text{nes}}$$

S preureditvijo enačbe lahko izračunamo komponento ugnezenosti, oz. **indeks ugnezdene raznolikosti**, β_{nes} :

$$\beta_{\text{nes}} = \beta_{\text{sor}} - \beta_{\text{sim}} .$$

Enako velja za mnogotere primerjave, s tem da v izračun vključimo Sørensenov indeks mnogoterih raznolikosti (β_{SOR}) in Simpsonov indeks mnogoterih raznolikosti (β_{SIM}), iz katerih dobimo **indeks ugnezdene mnogotere raznolikosti**, β_{NES} :

$$\beta_{\text{NES}} = \beta_{\text{SOR}} - \beta_{\text{SIM}} .$$



Slika 3.6 Komponenti beta diverzitete: vrstni obrat in ugnezenost. Na primeru dveh območij in z devetimi vrstami (1–9) so predstavljeni procesi: vrstni obrat, ugnezenost in izguba vrstnega bogastva; α – alfa diverziteta, β – beta diverziteta, γ – gama diverziteta.

Hipotetični primeri (A-D) primerjav med dvema območjema (Območje 1 in 2). Vrsto bogastvo v primeru A je primer ugnezenosti. Vrsteni obrat v primeru B je podmnožica bogatejšega območja 1.

V primeru B gre za vrsteni obrat, vrstno bogastvo obeh območij je izenačeno, nekaj vrst je identičnih: 2, 3, 4, nekaj je različnih: 1, 5, 7, 8 in 9.

V primeru C so predstavljeni procesi: vrsteni obrat (območji imata nekaj skupnih in nekaj različnih vrst), ugnezenost (območje 2 je delna podmnožica območja 1) in izguba vrstnega bogastva.

V primeru D območji ne delita skupnih vrst, prisotna sta procesa vrstni obrat in izguba vrst.

RAZLAGA

PA podatki (ang. Presence–absence Data) so na dihomen način zapisani **podatki o prisotnosti** (prezenci) oz. odsotnosti vrst v vzorčni enoti. Prisotni vrsti pripišemo vrednost 1, njeno odsotnost opredelimo z vrednostjo 0. Rezultat je PA matrika (z vrednostnima 0 ali 1), kjer so v vrsticah popisi posameznega vzorca in v stolpcih vrste. Drugi način priprave preglednice je zapis **podatkov o številčnosti (abundanci)**, tj. o številu osebkov določene vrste na izbrani vzorčni enoti.

Vzorčena enota SU (ang. Sampling Unit) ali vzorčni popis je enkratni popis v določenem času in na definiranim območju (npr. na vzorčni oz. popisni ploskvi, vzorčni lokaliteti, kvadratu, v linijskem transektu, definiranim volumnu vode, tal ali drugega medija, v ekološki združbi, v regiji), ali definirana z določeno količino vložene dela v vzorčenje (ura/človek/vzorčena enota) ali po drugih kriterijih.



PRIMER 3.2

Suhe južine v visokogorskih habitatih – pregled vrst

Preglednico s PA podatki (Preglednica 3.2) s prisotnostjo vrst v posameznem habitatnem tipu (HT, razlaga v preglednici 3.1) orientiramo z vrstami v vrsticah in HT v stolpcih. Vrednost 1 pomeni, da smo vrsto potrdili v habitatnem tipu, vrednost 0 pa, da vrste v tem HT nismo našli. Z barvami so označene posamezne kategorije vrstnega bogastva; vrste potrjene v vseh HT, edinstvene – vrste potrjene v samo enem HT, dvojne – vrste potrjene v dveh HT. Število skupnih vrst ter število vrst samo v referenčnem oz. samo v sosednjem vzorcu za posamezni par habitatov je predstavljeno v Preglednici 3.3.

Preglednica 3.2 Pregled vrst suhih južin Opiliones in njihova prisotnost v petih habitatnih tipih (HT) na območju Kamniško-Savinjskih Alp. Vrste, prisotne v vseh območjih, so označene z modro, edinstvene z rdečo in dvojne z zeleno

| Koda | Vrsta | / HT | 36.413 | 36.433 | 61.231 | 61.232 | 31.52 |
|--------|--|------|--------|--------|--------|--------|-------|
| HISDEN | <i>Histicostoma dentipalpe</i> | | 0 | 1 | 0 | 1 | 1 |
| TROROS | <i>Trogulus rostratus</i> | | 0 | 1 | 1 | 1 | 0 |
| TRONEP | <i>Trogulus nepaeformis</i> | | 1 | 0 | 1 | 0 | 1 |
| TROTIN | <i>Trogulus tingiformis</i> | | 1 | 0 | 0 | 1 | 1 |
| TROFAL | <i>Trogulus falcipenis</i> | | 1 | 0 | 0 | 0 | 0 |
| LOPPAL | <i>Lophopilio palpinalis</i> | | 1 | 0 | 0 | 1 | 1 |
| PLABUC | <i>Platybunus bucephalus</i> | | 1 | 1 | 0 | 0 | 1 |
| MITMOR | <i>Mitopus morio</i> | | 1 | 1 | 1 | 1 | 1 |
| MITALP | <i>Mitostoma alpinum</i> | | 1 | 1 | 1 | 1 | 0 |
| GYAANN | <i>Gyas annulatus</i> | | 0 | 0 | 0 | 0 | 1 |
| NEMTRI | <i>Nemastoma triste</i> | | 1 | 1 | 1 | 1 | 1 |
| NEMBBI | <i>Nemastoma bidentatum bidentatum</i> | | 1 | 1 | 0 | 1 | 1 |
| DICGAS | <i>Dicranopalpus gasteinensis</i> | | 1 | 1 | 0 | 1 | 1 |
| MITCHR | <i>Mitostoma chrysomelas</i> | | 0 | 0 | 0 | 1 | 1 |

Preglednica 3.3 Vrednosti *a*, *b* in *c* Vennovega diagrama (Slika 3.5) pri primerjanju vrstne pestrosti med petimi habitatnimi tipi

| a | 36.413 | 36.433 | 61.231 | 61.232 | | b | 36.413 | 36.433 | 61.231 | 61.232 |
|--------|--------|--------|--------|--------|--|--------|--------|--------|--------|--------|
| 36.433 | 6 | | | | | 36.433 | 2 | | | |
| 61.231 | 4 | 4 | | | | 61.231 | 1 | 1 | | |
| 61.232 | 7 | 7 | 4 | | | 61.232 | 3 | 3 | 6 | |
| 31.52 | 8 | 6 | 3 | 8 | | 31.52 | 3 | 5 | 8 | 3 |
| c | 36.413 | 36.433 | 61.231 | 61.232 | R > package = »vegan« # Izračun vrednosti a, b, c Vennovega diagrama abc <- betadiver(data) abc | | | | | |
| 36.433 | 4 | | | | | | | | | |
| 61.231 | 6 | 4 | | | | | | | | |
| 61.232 | 3 | 1 | 1 | | | | | | | |
| 31.52 | 2 | 2 | 2 | 2 | | | | | | |



PRIMER 3.3

Suhe južine v visokogorskih habitatih – beta diverziteta

Vrednosti Simpsonovega indeksa raznolikosti so nizke ($< 0,5$), to pomeni, da prostorski obrat vrst ni izrazit, še najvišjo vrednost $\beta_{sim} = 0,4$ izkazuje primerjava med habitatnima tipoma 61.231 in 31.52 (Preglednica 3.4).

Vrednosti Sørensenovega indeksa so pri večini primerjav večje od 0,5. To pomeni, da imajo primerjani habitatni tipi večino vrst skupnih, najvišjo vrednost dosega primerjava med tipoma 63.433 in 61.232 $\beta_{sor} = 0,778$ (Preglednica 3.4), ki imata 7 skupnih vrst.

Najvišjo ugnezenost, to je največ manjkajočih vrst, ima habitat 31.52 v primerjavi s habitatom 61.231, $\beta_{nes} = 0,025$ (Preglednica 3.4), skupne vrste so le 3, habitat 61.231 ima 8 vrst drugačnih kot habitat 31.52.

Komponente mnogoterih primerjav Whittakerjeva β diverziteta ima vrednost $\beta_w = 0,591$. Sørensenov indeks mnogoterih raznolikosti ima vrednost $\beta_{SOR} = 0,508$, Simpsonov indeks mnogoterih raznolikosti ima vrednost $\beta_{SIM} = 0,362$ in Indeks ugnezenosti v mnogoterih primerjavah ima vrednost $\beta_{NES} = 0,146$. Vrednost Simpsonovega indeksa je zmerna in manjša od 0,5, kar razlagamo s tem, da je povprečni prostorski obrat med habitatni zmeren

Preglednica 3.4 Parne primerjave prostorske komponente v sestavi favne opilionov med habitatnimi tipi s pomočjo Sørensenovega – β_{sor} , Simpsonovega indeksa raznolikosti – β_{sim} ter indeksa ugnezdenosti – β_{nes} . Iz rezultatov razberemo, da višji je indeks, tem večji je vrstni obrat, in obratno

| β_{sim} | 36.413 | 36.433 | 61.231 | 61.232 |
|----------------------|--------|--------|--------|--------|
| 36.433 | 0,250 | | | |
| 61.231 | 0,200 | 0,200 | | |
| 61.232 | 0,300 | 0,125 | 0,200 | |
| 31.52 | 0,200 | 0,250 | 0,400 | 0,200 |

| β_{sor} | 36.413 | 36.433 | 61.231 | 61.232 |
|----------------------|--------|--------|--------|--------|
| 36.433 | 0,667 | | | |
| 61.231 | 0,533 | 0,615 | | |
| 61.232 | 0,700 | 0,778 | 0,533 | |
| 31.52 | 0,762 | 0,632 | 0,375 | 0,762 |

| β_{nes} | 36.413 | 36.433 | 61.231 | 61.232 |
|----------------------|--------|--------|--------|--------|
| 36.433 | 0,417 | | | |
| 61.231 | 0,333 | 0,415 | | |
| 61.232 | 0,400 | 0,653 | 0,333 | |
| 31.52 | 0,562 | 0,382 | 0,025 | 0,562 |

```
R > package = »vegan«
# Indeksi parnih primerjav
betadiver(help=TRUE)
# Simpsonov indeks
sim <- betadiver(data, "sim")
sim
# Sørensenov indeks
sor <- betadiver(data, "sor")
sor
# Indeks ugnezdenosti
nes <- sor - sim
nes
abc <- betadiver(data)
abc
```

Sørensenov indeks je blizu 0,5, kar pomeni da je povprečno število skupnih vrst vsem habitatom zmerno. Indeks ugnezdenosti je nizek in blizu 0, ker pomeni, da se hitro povečuje povprečno število manjkajočih vrst v primerjavi med habitatami.

```
R > package = »betapart«
# Indeksi mnogoterih primerjav
sor.beta <- beta.multi(data, index.family="sorensen")
sor.beta
abc <- betadiver(data)
abc1
```



3.1.3 Zeta diverziteta

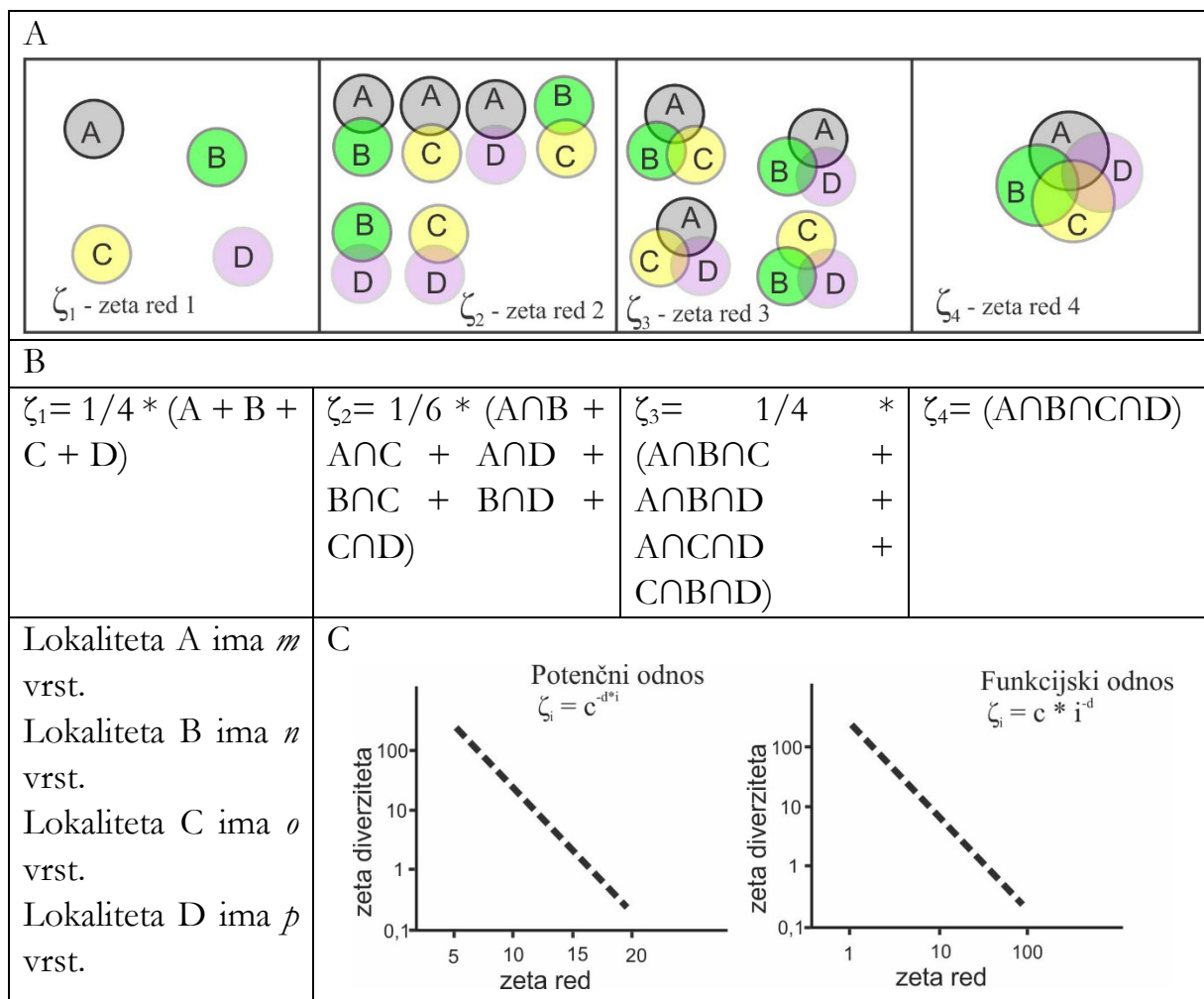
Zeta diverziteta ζ je vidik obravnave vrstnega bogastva, v katerem primerjamo podobnost vrstne sestave več lokalitet, oz. med več vzorci. Metodologijo so razvili Hui in McGeoch (2014) ter McGeoch s sodelavci (2019) in po teh člankih je povzeta vsebina poglavja. Zeta diverziteta ζ je alternativa analizi beta diverzitete β , s katero izvajamo primerjave med pari lokalitet. Beta diverziteta je omejena na primerjave med pari lokalitet, vendar je analizo mnogoterih primerjav težko interpretirati (glej Poglavje 3.1.2). Pristop z analizo zeta diverzitete je ustrezno rešil problem mnogoterih primerjav med lokalitetami

($n > 2$). Zeta diverziteteta ζ je tudi cenilka prostorske (in časovne) variabilnosti vrstnega bogastva pri primerjanju mnogo vzorcev (z vzorcem je mišljeno vrstno bogastvo bodisi na lokaliteti (v vzorcu) bodisi gre za nabor vrst) (Slika 3.7.a). Izračun zeta cenilke ζ temelji na podatkih o prisotnosti vrst (prezenca vrst, PA podatki) (ang. Incidence-based Data). Podatke organiziramo v pravokotni preglednici na način, da so lokalitete v vrsticah in vrste organizmov v stolpcih, vrednosti v celicah so 1 v primeru prisotnosti oz. 0 v primeru odsotnosti vrste na lokaliteti (primer organizacije podatkov je v Preglednici 3.6). Vrednost cenilke zeta ζ_i podaja povprečno število vrst, ki so skupne več lokalitetam i . Število lokalitet, vključenih v analizo (vrednost i), pomeni tudi število nivojev primerjav. Posamezni nivo primerjav se imenuje **zeta red ζ_i** (ang. Zeta Order). Zeta ena ζ_1 je cenilka diverzitetete prvega zeta reda in podaja povprečno število vrst pri obravnavi mnogo lokalitet vključenih v analizo (je primerljiva diverziteti α) (Slika 3.7.b). Zeta dva ζ_2 je cenilka diverzitetete drugega zeta reda in podaja povprečno število skupnih vrst pri primerjanju dveh lokalitet (je primerljiva z β diverziteteto) (Slika 3.7.b). Povprečje izračunamo iz vseh možnih kombinacij med pari lokalitet. Zeta tri ζ_3 je povprečno število skupnih vrst pri primerjanju treh lokalitet (Slika 3.7.b). Povprečje izračunamo iz vseh možnih kombinacij trojic lokalitet. Zeta štiri ζ_4 je povprečno število skupnih vrst pri primerjanju štirih lokalitet, za vse možne kombinacije štirih lokalitet. Postopek nadaljujemo dokler red zeta ni izenačen s številom lokalitet (Slika 3.7.b). Zeta diverziteteto prikažemo grafično s krivuljo oz. z regresijsko premico, ki monotono upada z naraščanjem zeta reda (Slika 3.7.c). Zeta diverziteteta v odnosu do višanja zeta reda vedno upada, zato ta odnos poenostavljeno imenujemo **zeta upadanje diverzitetete** (ang. Zeta Diversity Decline). Ključna informacija za razumevanje in interpretacijo prostorskega vrstnega obrata je naklon premice, ki ga opisuje odnos med zeta diverziteteto in zeta redom, torej upadanje diverzitetete. Odnos med zeta diverziteteto in zeta redom opišemo z regresijsko enačbo, odnos je bodisi potenčni $\zeta_i = c^{-d \cdot i}$ bodisi funkcijski $\zeta_i = c * i^{-d}$, kjer je c presečišče regresijske premice in Y osi, i je število lokalitet (oz. zeta red), d je smerni koeficient regresijske premice (naklon premice). Izbiro med modeli, ki najbolj razložijo odnos opravimo z Akaikovim informacijskim kriterijem AIC (glej Poglavje 15.2). Funkcijski odnos je informacija, da je sprememba v vrstni sestavi predvsem posledica diferenciacije ekoloških niš. V potenčnem modelu upadanja zeta diverzitetete prevladujejo stohastični dogodki. Pri analizah in primerjavah z zeta diverziteteto je pomembna informacija pri katerem zeta redu je zeta diverziteteta dosegla vrednost nič $\zeta_i = 0$. Z drugimi besedami, pri katerem zeta redu nastopijo pogoji, ko si novo dodana lokaliteta ne deli nobene skupne vrste z dosedanjimi lokalitetami. Vrstni obrat je lahko posledica: 1. gradacije ekoloških razmera ali 2. različnih habitatov. V prvem primeru je med lokalitetami vzpostavljen ekološki gradient, ki privede do vrstno specifične sestave biote na konkretni lokaliteti. Npr. spremljanje vrstne sestave plazilcev z naraščanjem nadmorske višine (npr. glej primer v Poglavju 10). V drugem primeru gre za habitatsko heterogenost lokalitet oz. različne habitate, ki povzroči lokalno specifično sestavo biote (Primer 3.4).

Zeta razmerje

Skupne vrste na lokalitetah $n + 1$ so podskupina skupnih vrst na predhodnih lokalitetah n , razmerje med zeta diverzitetama sosednjih redov podamo s cenilko **zeta razmerje** ZR (ang. Zeta Ratio):

$$ZR_n = \frac{\zeta_{n+1}}{\zeta_n}$$



Slika 3.7 Predstavitev računanja vrednosti cenilke zeta diverzitetete ζ_i za posamezni zeta red. A. Z Vennovimi diagrami je predstavljen primer, ki obsega štiri lokalitete A, B, C in D. Vsaka lokaliteta ima lasten nabor vrst, prekrivanje krožnic ponazarja skupne vrste. Pri vsaki primerjavi lokalitet merimo število skupnih vrst. B. Znotraj posameznega zeta reda se izračuna zeta cenilka kot povprečno število skupnih vrst vseh primerjav. C. Krivulja oz. premica upadanja zeta diverzitetete, kot funkcija števila lokalitet i oz. zeta redu; potenčni in funkcijski odnos (razlaga členov enačb je v besedilu).

Zeta razmerje ZR_n predstavlja verjetnost prisotnosti skupnih vrst iz lokalitet n na novi dodatni lokaliteti $n + 1$. Ta informacija o skupnih vrstah, ki zasedajo n število lokalitet, podaja verjetnost o njihovi prisotnosti na novi dodatni lokaliteti. ZR ima nabor vrednosti od 0 do 1, naraščanje vrednosti pomeni večjo verjetnost za skupne vrste na novo vključeni

lokaliteti. Z drugimi besedami, ko vrednost ZR gre proti ena, pomeni, da z dodajanjem novih lokalitet samo potrjujemo že obstoječi nabor vrst. In obratno, ko se vrednost ZR približuje 0, pomeni, da s povečevanjem števila lokalitet ne pričakujemo več skupnih vrst. Z dodajanje novih lokalitet (ali popisov) odkrivamo nove vrste, ki v predhodnih vzorčenjih še niso bile zajete. Odnos med zeta redom in zeta razmerjem podamo podobno kot pri zeta diverziteti, in sicer z regresijsko krivuljo in regresijskim koeficientom. V tem primeru analiz je pomembna informacija o obliki odnosa oz. regresijske krivulje, ki je lahko unimodalna, naraščajoča ali asimptotična, ter pri katerem zeta redu je ZR doseglo maksimalno vrednost. Do maksimalne vrednosti ZR število skupnih vrst narašča, potem pa upada.



PRIMER 3.4

Zeta diverziteti suhih južin v visokogorskih habitatih

Primerjali smo spremembo v vrstni sestavi suhih južin med petimi habitatnimi tipi v visokogorju. Podatki so povzeti po raziskavi, objavljeni v članku Novak in sodelavci (2017). V vsakem posameznem habitatu je bilo izvedenih več vzorčenj (predstavljeno v naslednjem primeru). Tukaj je podan primer analize primerjav v vrstnem obratu suhih južin med petimi habitatnimi tipi.

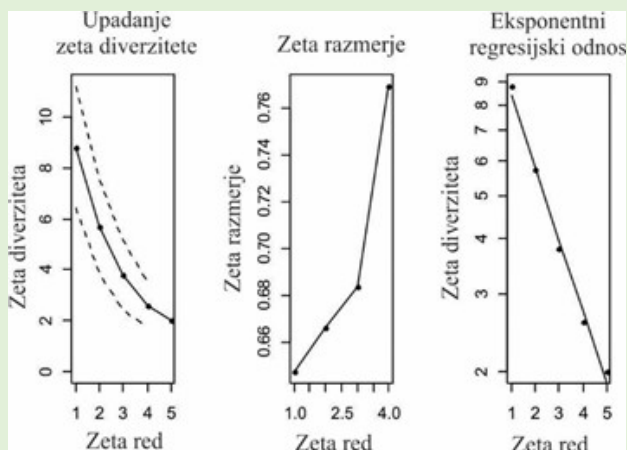
Podatki o prisotnosti vrst so predstavljeni v Preglednici 3.6, ki ima velikost 5 vrstic s habitatnimi tipi in 14 stolpcev z vrstami suhih južin, njihova prisotnost v habitatu je označena z 1.

Zeta diverziteti ima najvišjo vrednost v prvem zeta redu $\zeta_1 = 8,8$ in z vsakim redom upada, to pomeni, da je povprečno število vrst v petih habitatih 8,8. Pri najvišjem (petem redu) ima vrednost $\zeta_5 = 2,0$ in še ni dosegla vrednosti 0 (Preglednica 3.5, Slika 3.8). Slednje pomeni, da sta dve vrsti suhih južin skupni vsem petim habitatnim tipom. V ekološki interpretaciji rezultata lahko sklenemo, da je vrstna sestava suhih južin v različnih habitatnih tipih dokaj podobna oz. da so ekološke razmere v raziskovanih habitatih dokaj podobne. Upadanje diverziteti je potenčno in ga lahko zapišemo v obliki $\zeta_i = 1,0876^{-0,1628 \cdot i}$. Za nadaljnje primerjave z drugimi skupinami organizmov oz. drugimi območji je pomemben smerni koeficient d , ki ima negativno vrednost $-0,1628$.

Zeta razmerje ZR_i ima začetno vrednost 0,65 in z vsakim redom narašča in v zadnjem redu že ima vrednost 0,77 (Preglednica 3.5, Slika 3.8). Opisani vzorec naraščanja interpretiramo v kontekstu, da so v vsakem naslednjem redu primerjav ohranjene skupne vrste ali vsaj nekaj njih. Z drugimi besedami, različni habitatni tipi v visokogorju imajo dokaj enolično sestavo favne suhih južin. Domnevamo, da so zaostrene ekološke razmere v visokih

nadmorskih višinah nad 1.500 m že do te mere prefiltrirale favno suhih južin, da je ohranjenih le nekaj vrst, ki se bolj ali manj pojavljajo v vseh raziskanih habitatnih tipih.

Preglednica 3.5 Upadanje zeta diverzitete. Zeta red pomeni število habitatnih tipov, istočasno vključenih v izračun skupnih vrst suhih južin, ki jih naseljujejo; zeta diverziteta ζ_i je izračun povprečne vrednosti skupnih vrst v posameznem redu in njegova standardna deviacija StDev; povprečje je izračunano iz števila kombinacij primerjav; zeta razmerje ZR podaja kvocient med vrednostjo zeta diverzitete in njegovim predhodnikom ζ_{i+1}/ζ_i .



Slika 3.8 Upadanje zeta diverzitete suhih južin pri primerjanju petih habitatnih tipov.

Preglednica 3.6 Prisotnost suhih južin v petih habitatnih tipih, razlaga habitatnih tipov je v Preglednici 3.1 in vrst v Preglednici 3.2

| HT/SP | HISDEN | PLABUC | MITMOR | LOPPAL | GYAANN | DICGAS | TROROS | TRONEP | NEMTRI | TROPAL | NEMBI | MITALP | MITCHR | TROTIN |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|-------|--------|--------|--------|
| 36.413 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| 36.433 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 |
| 61.231 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 |
| 61.232 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 |
| 31.52 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 |

```
R > package = »zetadiv«
# Suhe južine, podatki = zeta.data
zeta.data
sj<-Zeta.decline.ex(zeta.data, orders = 1:5)
sj
```





PRIMER 3.5

Odkrivnost vrstnega bogastva suhih južin v visokogorskih habitatnih tipih

Novak in sodelavci (2017) poročajo o vzorčenju suhih južin v petih visokogorskih habitatnih tipih. Povzetek z vrstnim bogastvom je predstavljen v Preglednici 3.1. Zanima nas odkrivnost vrst oz. stopnja raziskanosti vrstnega bogastva znotraj posameznega habitatnega tipa. Raziskovalci so v popis posameznega habitatnega tipa vložili različno količina dela od 11 do največ 25 popisov in pri tem odkrili skupno 14 različnih vrst suhih južin od 5 do 11 vrst po habitatu (Preglednici 3.1 in 3.6). Odnos med zeta diverzitetom in zeta redom je v vseh pet primerih potenčni; smerni koeficient d je visok in ima vrednost od $-0,77$ do $-0,35$ (Preglednica 3.7, Slika 3.9). Ta rezultat lahko razumemo, da število istih vrst v popisih znotraj posameznega habitatnega tipa hitro upada. Z drugimi besedami in ob pregledu surovih podatkov vidimo, da posamezni popisi vsebujejo zelo malo vrst. To najbrž pomeni, da so morali vložiti veliko dela za odkritje sploh kakšne vrste. Podrobnejši pregled pokaže, da je v 122 popisih v 41,0 % bila popisana samo ena vrsta, v naslednjih 42,6 % pa samo dve vrsti. Največ je bilo v enem popisu odkritih 5 vrst, takih primerov je bilo pet (Preglednica 3.8). Posledica nizke odkrivnosti živali je, da se večina popisov razlikuje v vrstni sestavi.

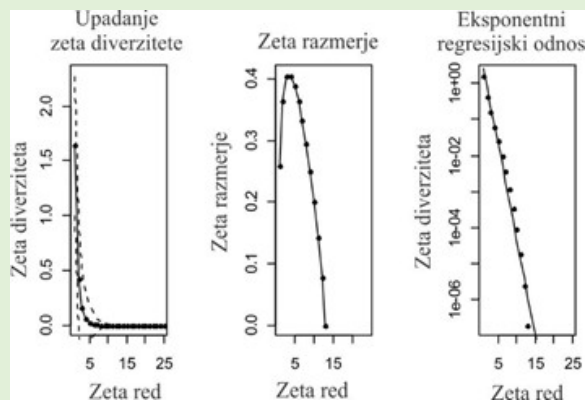
Kljub temu je zeta red, pri katerem je zeta diverziteta dosegla vrednost nič $\zeta = 0$ relativno visok in ima vrednost od 6 do 14. Najvišja vrednost 14 je bila dosežena v habitatnem tipu 36.413, v katerem je bilo opravljenih 25 popisov. To pomeni, da je bila vsaj ena skupna vrsta odkrita oz. prisotna v 14 od 25 popisov oz. v 56 % popisov.

Preglednica 3.7 Zeta diverziteta suhih južin v petih habitatnih tipih HT. Razlaga oznak: i – število vzorčenj, n – število vrst, $\zeta = 0$ – zeta red, kjer zeta diverziteta doseže vrednost 0, ΔAIC – Akaikov informacijski kriterij pri odločitvi za potenčni ali funkcijski odnos med zeta diverzitetom in zeta redom, E/P – informacija, ali je odnos potenčni – E ali funkcijski – P, d – smerni koeficient regresijske premice, $Max\ red$ – zeta red, kjer zeta razmerje ZR doseže maksimum, ZR – maksimalna vrednost zeta razmerja

| HT | i | n | $\zeta=0$ | ΔAIC | E/P | d | $Max\ red$ | ZR |
|--------|-----|-----|-----------|--------------|-------|---------|------------|-------|
| 36.413 | 25 | 10 | 14 | 13,01 | E | -0,5314 | 4 | 0,405 |
| 36.433 | 11 | 8 | 6 | -6,84 | E | -0,6732 | 2 | 0,282 |
| 61.231 | 13 | 5 | 10 | -3,07 | E | -0,3518 | 3 | 0,539 |
| 61.232 | 11 | 10 | 6 | -6,90 | E | -0,6901 | 2 | 0,256 |
| 31.52 | 22 | 11 | 9 | 0,28 | E | -0,7708 | 3 | 0,225 |

Preglednica 3.8 Frekvenca zastopanosti števila vrst na posamezni popis

| Št. vrst na popis | Frekvenca | % |
|-------------------|-----------|-------|
| 1 | 50 | 41,0 |
| 2 | 52 | 42,6 |
| 3 | 15 | 12,3 |
| 5 | 5 | 4,1 |
| Skupaj | 122 | 100,0 |



Slika 3.9 Zeta diverzitet, zeta razmerje in eksponentni regresijski odnos za habitatni tip 36.413.

```
R > package = »zetadiv«
#HT 36.413, 25 vzorčenj
ht63413
zd.63413<-Zeta.decline.ex(ht63413, orders = 1:25)
zd.63413
#HT 36.433, 11 vzorčenj
ht63433
zd.63433<-Zeta.decline.ex(ht63433, orders = 1:11)
zd.63433
#HT 36.231, 13 vzorčenj
ht63231
zd.63231<-Zeta.decline.ex(ht63231, orders = 1:13)
zd.63231
#HT 61.232, 11 vzorčenj
ht61232
zd.61232<-Zeta.decline.ex(ht61232, orders = 1:11)
zd.61232
#HT 11.52, 22 vzorčenj
```



Zeta diverzitet v kontekstu redkih oz. pogostih vrst

Vrstno bogastvo mnogoterih lokalitet lahko razdelimo na pet vidikov redkosti oz. pogostosti vrst, in sicer v tri vidike znotraj alfa diverzitete (zeta, eta in theta) in v dva vidika znotraj beta diverzitete (jota in kapa) (Slika 3.10). Definicija posameznih vidikov je naslednja:

Komponente alfa diverzitete:

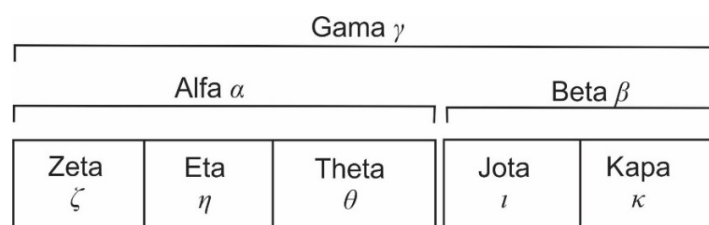
- Zeta ζ – sestavljajo jo vrste skupne vsem naborom vrst. Komponenta zeta je enakovredna zeta diverziteti.

- Eta η – sestavljajo jo vrste, ki niso edinstvene samo za en nabor vrst niti niso skupne vsem naborom vrst.
- Theta θ – sestavljajo vrste, ki so edinstvene samo za en nabor vrst.

Komponenti beta diverzitete:

- Jota ι – pričakovane vrste, vendar v naboru vrst manjkajo. Komponenta jota je enakovredna temni diverziteti (glej Poglavje 3.5).
- Kappa κ – vrste, prisotne v naboru vrst, vendar niso pričakovane v konkretnem vzorčenju. Komponenta jota je enakovredna prostorskemu vrstnemu obratu (glej Poglavje 3.1.2).

Povzeto po Riva in Mammola (2021).



Slika 3.10 Pet vidikov redkih oz. pogostih vrst v kontekstu alfa, beta in gama diverzitete. Zeta – skupne vrste na več lokalitetah (pogoste vrste). Eta – nikjer pogoste in nikjer redke vrste. Theta – prisotne samo na eni lokaliteti (redke vrste). Jota – manjkajoče vrste (temna diverziteteta). Kapa – nepričakovane vrste na lokaliteti (prostorski vrstni obrat). Prirejeno po Riva in Mammola 2021.

3.2 Ocena vrstnega bogastva

Sestavni del biodiverzitetnih študij je ugotavljanje stopnje raziskanosti vrstnega bogastva, torej preverjanje, ali smo vložili dovolj raziskovalnega dela za odkrivanje vrst, oz. kdaj smo vložili dovolj dela, da lahko s popisovanjem končamo. Enak pristop lahko uporabimo pri primerjanju popisov z različno količino vložene delo. V ta namen izdelamo **akumulacijsko krivuljo vrst SAC** (ang. Species Accumulation Curve) (glej Poglavje 16.1), iz katere razberemo:

1. Ali še lahko pričakujemo dodatne vrste?
2. Ali smo v iskanje vrst vložili dovolj dela, da lahko zaključimo s popisovanjem?
3. SAC uporabimo za primerjavo vrstnega bogastva med območji.

3.2.1 Akumulacija vrst

Akumulacijsko krivuljo vrst lahko izvajamo iz dveh izhodišč: 1. glede na število vzorčenj – na vzorcih temelječa analiza ali 2. glede na skupno število vzorčenih osebkov – na osebkih temelječa analiza. Pri oceni vrstnega bogastva lahko torej izhajamo iz **števila vzorčenih enot** (ang. Sample-based Assessment) ali iz **števila vzorčenih osebkov** (ang. Individual-based Assessment). Ob naključnem vzorčenju ob enakem vložku raziskovalnega dela je razmerje med številom vzorčenj in številom osebkov linearno in ekvivalentno. To pomeni, da se oceni bogastva vrst, pridobljeni z enim ali drugim pristopom, bistveno ne razlikujeta.

Za oceno vrstnega bogastva obstajata dve skupini cenilk: ena skupina temelji na abundanci, druga na prezenci. Za izračun ocene bogastva vrst na podlagi abundance uporabimo cenilki Chao1 ali ACE, za podatke o incidenci oz. pojavljanju (ang. Incidence Data), tj. za PA podatke, pa cenilke Chao2, ICE, Jackknife1, Jackknife2 ali bootstrap.

Poglavje o akumulaciji vrst je narejeno po virih Colwell 2013, Gotelli in Chao 2013 in Oksanen 2022.

Cenilko *Chao1* uporabimo pri podatkih o abundanci, pri čemer imajo največjo utež redke vrste. Ideja izhaja iz podmene, da število redkih vrst daje največ informacij o manjkajočih vrstah. Izračun temelji na **enojcih** (ang. Singleton) – vrstah s samo enim osebkom in **dvojcih** (ang. Doubleton) – vrstah z dvema osebkoma. Ocena števila vrst se približuje dejanskemu številu, ko je vsaka vrsta zastopana z vsaj dvema osebkoma. Na tej točki popis lahko zaključimo. Cenilko *Chao1* priporočajo pri popisih, kjer prevladujejo redke vrste, torej z nizko abundanco. Formula za izračun cenilke Chao1 je:

$$S_{\text{Chao1}} = S_{\text{obs}} + \frac{a^2}{2b},$$

kjer so S_{Chao1} – ocena števila vrst, S_{obs} – skupno število vrst, a – število enojcev, b – število dvojcev.

S cenilko *ACE* (ang. Abundance-based Coverage Estimator) ocenjujemo vrstno bogastvo na osnovi abundance. S to metodo razdelimo opazovane frekvence vrst v dve skupini: na **pogoste** in na **redke vrste**. Pogoste vrste so tiste z 10 in več osebkov v vzorcu, redke pa tiste z manj kot 10 osebkov v vzorcu. Po tem kriteriju v izračunu upoštevamo le število pogostih in število redkih vrst. Abundance pogostih vrst niso vključene v izračun, upoštevamo samo abundance redkih vrst, saj na njih temelji ocena o manjkajočih vrstah. Izračun ocene števila vrst s pomočjo *ACE*:

$$S_{ace} = S_{abund} + \frac{S_{rare}}{C_{ace}} + \frac{F_1}{C_{ace}} \gamma_{ace}^2$$

$$C_{ace} = 1 - \frac{F_1}{N_{rare}}$$

$$\gamma_{ace}^2 = \max \left[\frac{S_{rare}}{C_{ace}} \frac{\sum_{i=1}^{10} i(i-1)F_1}{(N_{rare})(N_{rare}-1)} - 1 \right].$$

kjer so S_{ace} – ocena števila vrst, S_{abund} – število pogostih vrst (abundanca ≥ 10), S_{rare} – število redkih vrst (abundanca < 10), γ_{ace}^2 – koeficient variabilnosti za redke vrste, C_{ace} – indeks, F_1 – število enojcev, N_{rare} – celokupno število osebkov redkih vrst, i – število osebkov posamezne redke vrste.

Cenilko *Chao2* uporabimo na PA podatkih; v izračunu so bolj pomembne redke vrste. Cenilko priporočajo pri popisih, v katerih prevladujejo vrste z majhnimi areali. Izračun temelji na razmerju med številom vrst, najdenih le v eni vzorčni enoti – **edinstvene** (ang. Unique), in številom vrst, najdenih samo v dveh vzorčenih enotah – **dvojne** (ang. Dupes). Ocena števila vrst doseže vrh, ko je vsaka vrsta zastopana vsaj v dveh vzorčenih enotah. Takrat lahko popis zaključimo. Formula za izračun cenilke *Chao2* je:

$$S_{Chao2} = S_{obs} + \frac{L^2}{2M},$$

kjer so S_{Chao2} – ocena števila vrst, S_{obs} – skupno število vrst, L – število edinstvenih vrst, M – število dvojnih vrst.

S cenilko *ICE* (ang. Incidence-based Coverage Estimator) ocenimo vrstno bogastvo na osnovi prezenca. Opazovane frekvence vrst z vseh vzorčenih enot razdelimo v dve skupini: na skupino s pogostimi vrstami (prisotne v 10 in več vzorčenjih) in na skupino z redkimi vrstami (prisotne v manj kot 10 vzorčenjih). Formula za izračun cenilke *ICE* je:

$$S_{ice} = S_{freq} + \frac{S_{infr}}{C_{ice}} + \frac{Q_1}{C_{ice}} \gamma_{ice}^2,$$

$$C_{ice} = 1 - \frac{Q_1}{N_{infr}},$$

$$\gamma_{ice}^2 = \max \left[\frac{S_{infr}}{C_{ice}} \frac{m_{infr}}{(m_{infr}-1)} \frac{\sum_{j=1}^{10} j(j-1)Q_1}{(N_{infr})^2} - 1 \right],$$

kjer so S_{ice} – ocena števila vrst, S_{freq} – število pogostih vrst (prisotnih vsaj v 10 vzorčnih enotah), Q_1 – frekvenca edinstvenih (unique) vrst, S_{infr} – število redkih vrst (prisotnih v < 10 vzorčnih enotah), γ_{ice}^2 – koeficient variabilnosti za redke vrste, m_{infr} – število vzorčnih enot z vsaj eno redko vrsto, $m_{infr}-1$ – (število vzorčnih enot z vsaj eno redko vrsto) – 1, j –

posamezna redka vrsta, C_{ie} – indeks vzorčne pokritosti, $N_{inf r}$ – število vrst s pogostostjo (z incidenco) < 10 .

RAZLAGA

Vzorčna pokritost

V vsakdanji ekološki in biodiverzitetni praksi, v kateri izvajamo popise organizmov, se dogajajo primeri, ko nekatere vrste ali osebke spregledamo in jih ne vključimo v štetje. Nekaterih redkih vrst pa tudi naključno ne zajamemo v popis. Indikator za navedeno napako merjenja je delež edinstvenih vrst, tistih ki jih popišemo samo v enem popisu (uniques); oz. enojcev, tistih samo z enim osebkom (singleton). Na redkih vrstah temelječi ocena vrstnega bogastva imenujemo **vzorčna pokritost** (ang. Sample Coverage) in jo vključimo v izračun ocene vrstnega bogastva in vrstne pestrosti.

Jackknife1 je cenilka vrstnega bogastva na osnovi prisotnosti vrst. Utež v izračunu ocene imajo edinstvene vrste, prisotne le v eni vzorčeni enoti. Formula za izračun cenilke *Jackknife1* je:

$$S_{Jack1} = S_{obs} + L \frac{n-1}{n},$$

kjer so S_{Jack1} – ocena števila vrst, S_{obs} – skupno število vrst, L – število edinstvenih vrst, n – število vzorčenih enot.

Jackknife2 je cenilka vrstnega bogastva na osnovi prisotnosti vrst. Izračun temelji na razmerju med edinstvenimi vrstami, ki so prisotne le v eni vzorčeni enoti, in dvojnimi vrstami, ki so prisotne v dveh vzorčenih enotah. Formula za izračun cenilke *Jackknife2* je:

$$S_{Jack2} = S_{obs} + \left(L \frac{2n-3}{n} \right) - \left(M \frac{(n-2)^2}{n(n-1)} \right),$$

kjer so S_{Jack2} – ocena števila vrst, S_{obs} – skupno število vrst, L – število edinstvenih vrst, M – število dvojnih vrst, n – število vzorčenih enot.

Bootstrap je cenilka vrstnega bogastva na osnovi prisotnosti vrst. Izračun temelji na deležu edinstvenih vrst, ki so prisotne le v eni vzorčeni enoti (vrsta k). Formula za izračun cenilke *bootstrap* je:

$$S_{boot} = S_{obs} + \sum_{k=1}^{S_{obs}} (1 - p_k)^m,$$

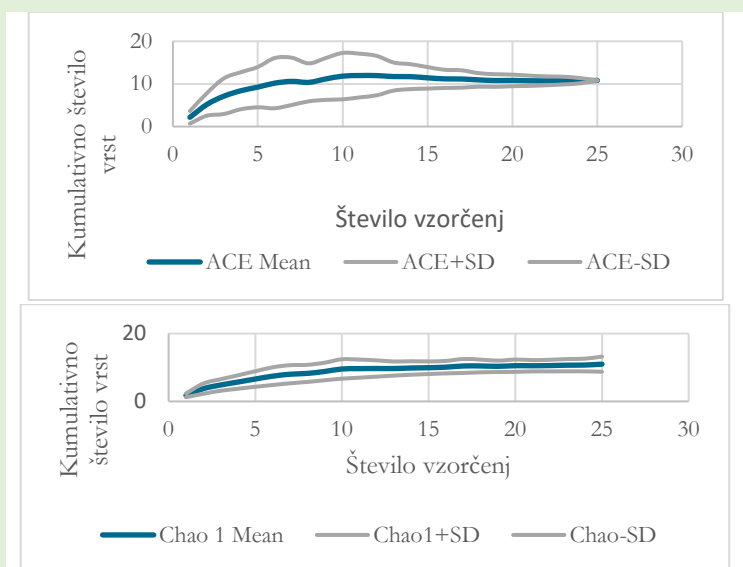
kjer so S_{boot} – ocena števila vrst, S_{obs} – skupno število vrst, p_k – delež vzorca, ki vsebujejo vrsto k , m – število vzorčnih enot.



PRIMER 3.6

Suhe južine v visokogorskih habitatih – ocena vrst – abundanca

V raziskavi suhih južin v Kamniško-Savinjskih Alpah KSA je bilo v HT 36.413 izvedenih 25 popisov, popisanih je bilo 60 osebkov iz 10 različnih vrst. Akumulacijsko krivuljo vrst smo izdelali glede na število vzorčenih enot, na podatkih z abundancami (Slika 3.11 in 3.12). Obe cenilki sta dosegli asimptotično vrednost pri okrog 10 vzorčenjih.



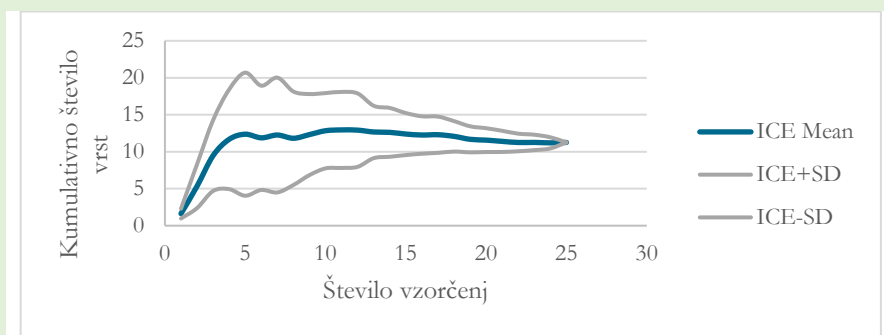
Slika 3.11 Krivulja ACE cenilke vrstnega bogastva.

Slika 3.12 Krivulja Chao1 cenilke vrstnega bogastva.

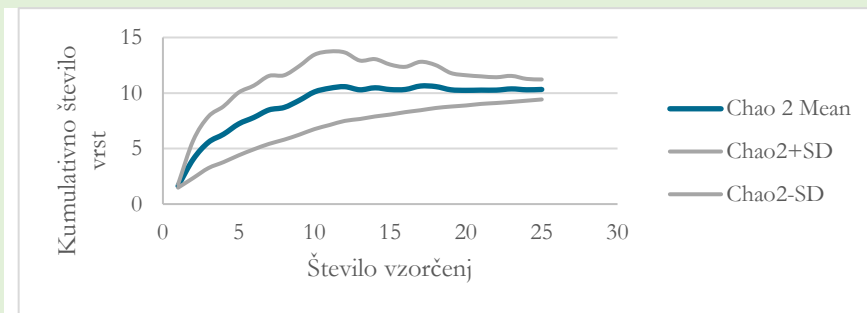


PRIMER 3.7

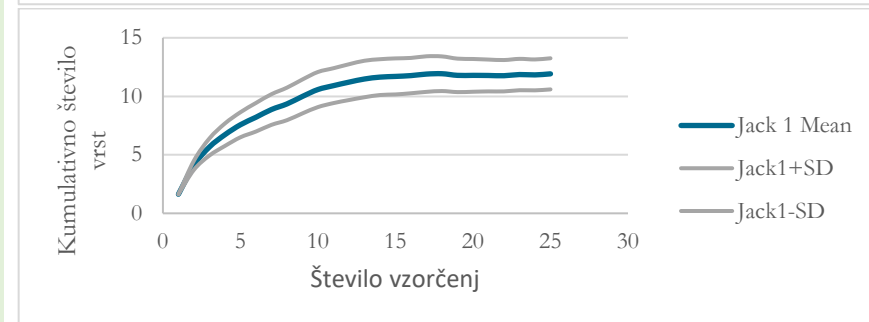
Suhe južine v visokogorskih habitatih – ocena vrst – prisotnost



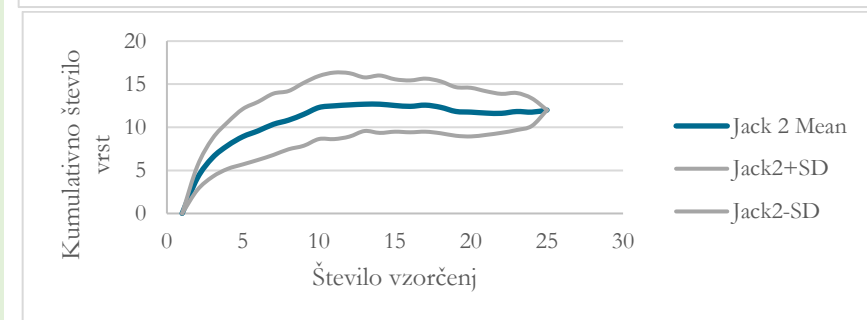
Slika 3.13 Krivulja ICE cenilke vrstnega bogastva.



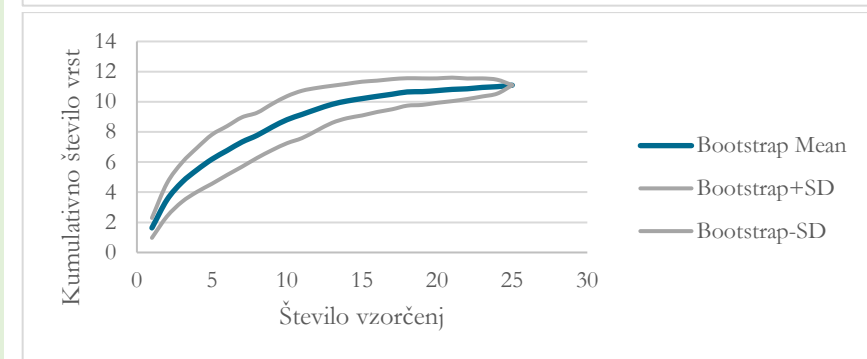
Slika 3.14 Krivulja Chao2 cenilke vrstnega bogastva.



Slika 3.15 Krivulja Jackknife1 cenilke vrstnega bogastva.



Slika 3.16 Krivulja Jackknife2 cenilke vrstnega bogastva.



Slika 3.17 Krivulja Bootstrap cenilke vrstnega bogastva.

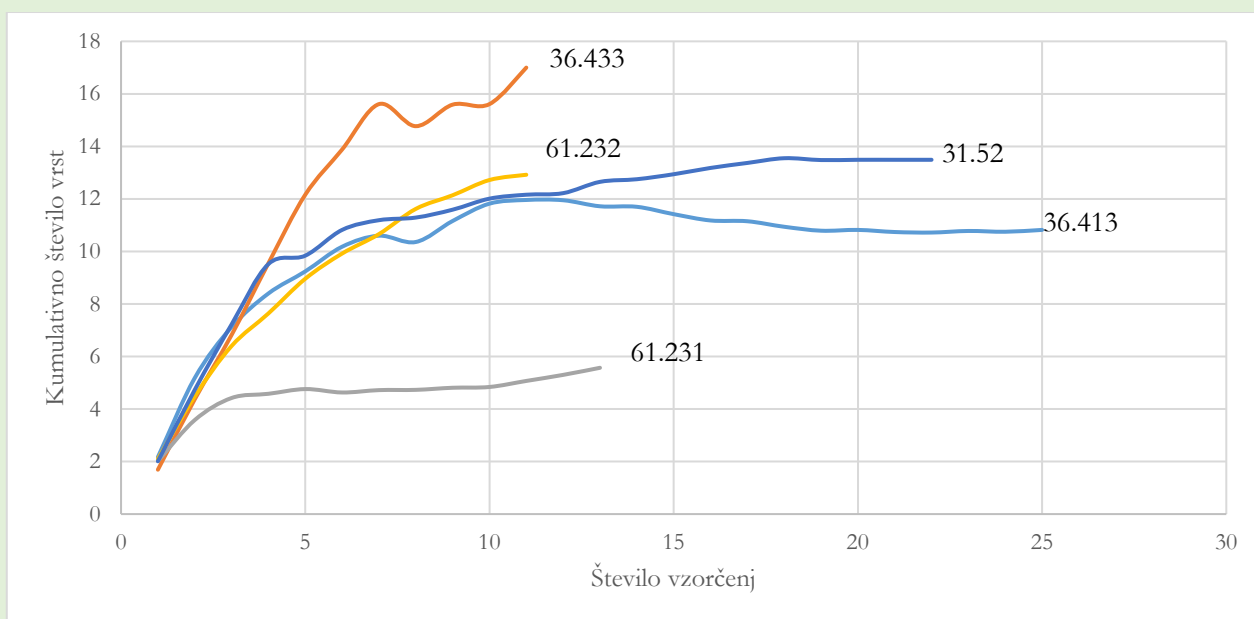


PRIMER 3.8

Suhe južine v visokogorskih habitatih – SAC

V raziskavi suhih južin v KSA smo popise izvajali v petih različnih habitatnih tipih HT. Raziskovalno delo (št. vzorčenj), vloženo v popis suhih južin je bilo od 11 do 25 popisov v posameznem HT. Preverjanje raziskanosti vrstnega bogastva smo izvedli po metodi ACE. Raziskanost vrstnega bogastva v HT 31.52 in 36.413 je dosegla asimptoto pri okrog 15 popisih. Nadaljnje vzorčenje v teh dveh HT ni bilo potrebno in s stališča racionalnosti

časa in sredstev nepotrebno. Raziskanost HT 61.232 in 61.231 je dosegla prevoj, vendar še ne vemo ali sta krivulji dosegli asimptoto, potrebno bi bilo še nekaj vzorčenj (približno 5). Raziskanost HT 36.433 ni zadostna, potrebno bi bilo več kot 5 vzorčenj (Slika 3.18).



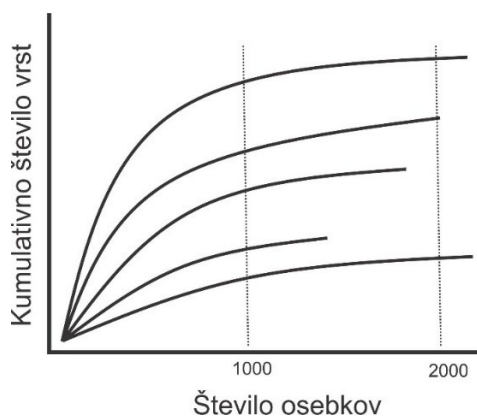
Slika 3.18 Akumulacijske krivulje cenilka vrstnega bogastva ACE v petih habitatnih tipih (za razlago HT glej Preglednico 3.1), primer nezadostnega (36.433) in prekomernega vzorčenja (36.413).



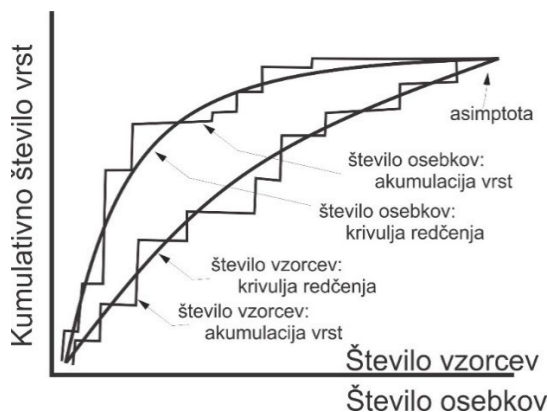
3.3 Redčenje

Redčenje (ang. Rarefaction) je metoda za primerjanje vrstnega bogastva dveh območij, na katerih smo vložili različno količino dela. Primerjavo večjega vzorca (območje z večjim številom vzorčenih enot ali z večjim številom opazovanih osebkov) z manjšim vzorcem izvedemo na način, da večjemu vzorcu zmanjšamo število vzorčenih enot ali število osebkov na enako število, kot jih ima manjši vzorec. Ocena vrstnega bogastva obeh primerjanih vzorcev (zmanjšanega in osnovnega) lahko izvedemo po metodi *Chao1* za podatke, temelječe na številu osebkov, ali po metodi *Chao2* za podatke, temelječe na številu vzorcev. Nato primerjamo oceno številčnosti vrst med obema vzorcema (Slika 3.19).

Ocena vrstnega bogastva na osnovi števila vzorčenih enot ali iz števila vzorčenih osebkov se deloma razlikujeta; krivulja na osnovi števila vzorčenih osebkov je v začetnem delu vzorčenja bolj strma (Slika 3.20).



Slika 3.19 Regresijske krivulje odnosov med številom vzorčenih osebkov in številom vrst za različne habitate. Krivulje predstavljajo posamezne habitate, ki so bili vzorčeni z različno količino dela oz. so v njih vzorčili različno število osebkov. Metoda redčenja omogoča primerjavo vrstnega bogastva med habitati na osnovi izbranega števila vzorčenih osebkov – npr. pri 1000 ali 2000 osebkih (črtkana črta).



Slika 3.20 Krivulji akumulacije vrst in krivulji redčenja; pristop z metodo števila vzorcev in pristop z metodo števila osebkov. Asimptota je točka, v kateri krivulja doseže maksimum, se ustali in ne narašča več.



PRIMER 3.9

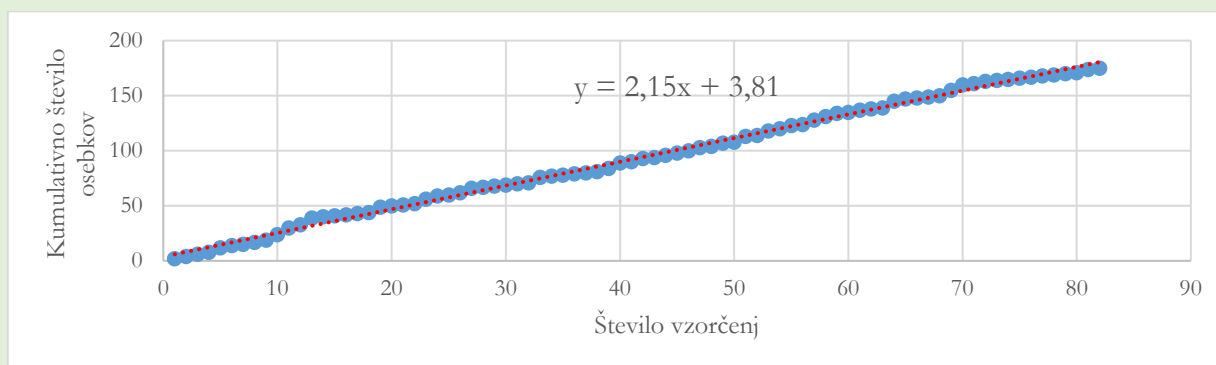
Suhe južine v visokogorskih habitatih – abundanca

Preglednico z abundanco vrst v posameznem habitatnem tipu (HT) orientiramo, da so v vrsticah vrste suhih južin in v stolpcih HT v (Preglednica 3.9). Vrednost v posamezni celici pomeni, da smo v tem HT prešteli toliko osebkov določene vrste, vrednost 0 pa, da vrste v tem HT nismo našli, oz. v numeričnem žargonu, da je vrsta zastopana z 0 osebki. Z barvami so označene posamezne kategorije vrstnega bogastva (vrste potrjene v vseh HT, enojci potrjene v samo enem HT, dvojci potrjeni v dveh HT).

Kumulativno število prešteti osebkov je v zelo tesni povezavi s številom vzorčenj (Slika 3.21). Regresijska enačba ima vrednost $y = 2,15 x + 3,81$ in pomeni, da z vsakim vzorčenjem naraste število osebkov v povprečju za 2,15 osebkov.

Preglednica 3.9 Abundanca suhih južin (za razlago glej Preglednico 3.2) v petih habitatnih tipih (HT, za razlago glej Preglednico 3.1) na območju Kamniško-Savinjskih Alp. Vrste prisotne v vseh habitatih so označene z modro, enojci z rdečo in dvojci z zeleno barvo

| Koda / HT | 36.413 | 36.433 | 61.231 | 61.232 | 31.52 |
|-----------|--------|--------|--------|--------|-------|
| HISDEN | 0 | 1 | 0 | 4 | 1 |
| TROROS | 0 | 1 | 4 | 1 | 0 |
| TRONEP | 6 | 0 | 2 | 0 | 3 |
| TROTIN | 1 | 0 | 0 | 2 | 2 |
| TROFAL | 1 | 0 | 0 | 0 | 0 |
| LOPPAL | 3 | 0 | 0 | 3 | 1 |
| PLABUC | 19 | 31 | 0 | 0 | 10 |
| MITMOR | 17 | 7 | 15 | 2 | 7 |
| MITALP | 3 | 4 | 6 | 3 | 0 |
| GYAANN | 0 | 0 | 0 | 0 | 3 |
| NEMTRI | 4 | 1 | 1 | 1 | 7 |
| NEMBBI | 3 | 1 | 0 | 2 | 3 |
| DICGAS | 1 | 1 | 0 | 1 | 1 |
| MITCHR | 0 | 0 | 0 | 1 | 1 |



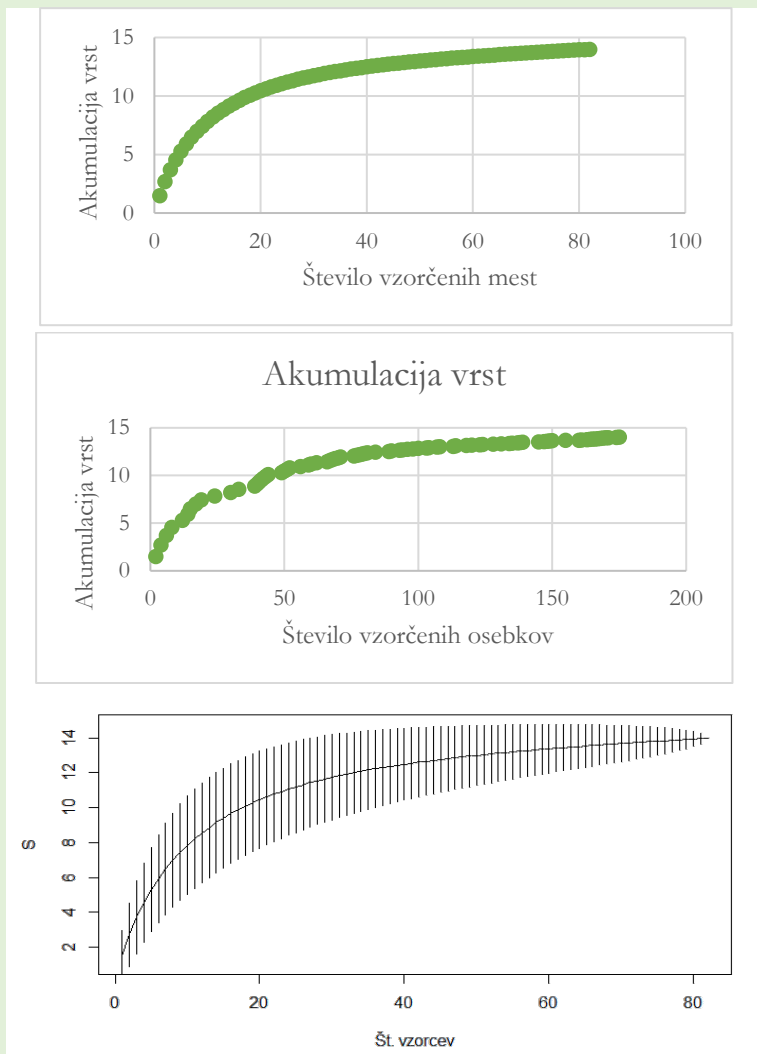
Slika 3.21 Razmerje med številom vzorčenj in kumulativnim številom osebkov (modre pike), ter regresijska premica (rdeča črtkana črta) in regresijska enačbe.



PRIMER 3.10

Suhe južine v visokogorskih habitatih – akumulacija vrst

V začetni fazi vzorčenja akumulirano število vrst strmo narašča, z nadaljevanjem vzorčenja sledi prevoj in vse počasnejše naraščanje vrst, dokler krivulja ne doseže asimptotične vrednosti. Vzorec akumulacije vrst ima identično obliko, če ga izvajamo s stališča števila vzorcev (Slika 3.22 in 3.24) ali s stališča števila vzorčenih osebkov (Slika 3.23).



Slika 3.22 Graf akumulacije števila vrst suhih južin glede na število vzorčenih mest.

Slika 3.23 Graf akumulacije števila vrst suhih južin glede na kumulativno število osebkov.

Slika 3.24 Krivulja kopičenja vrst suhih južin na število vzorčenih mest, prečne črte na krivulji predstavljajo standardno deviacijo.

```
R >
# Akumulacijska krivulja vrst
data.sac <- specaccum(data)
data.sac
# Izvoz podatkov
write.table(data.sac, file = "data_sac.txt", row.names = TRUE)
# SAC plot, na vseh popisih
plot(specaccum(data), xlab= "Št. vzorcev", ylab="S")
```



PRIMER 3.11

Suhe južine v visokogorskih habitatih – redčenje

Oceno števila vrst za posamezni habitatni tip HT z metodo redčenja na 18 vzorčenih osebkov. Primerjamo vrstno bogastvo med HT, če bi na vsakem HT vzorčili 18 osebkov (Preglednica 3.10).

Preglednica 3.10 Dejansko število popisanih osebkov in število vrst za pet habitatnih tipov HT, ter z metodo redčenja ocenjeno števila vrst pri 18 vzorčenih osebkih

| Habitatni tip | 36.413 | 36.433 | 61.231 | 61.232 | 31.52 |
|---------------|------------------------|--------|--------|--------|-------|
| Št. osebkov | 60 | 18 | 28 | 28 | 40 |
| Št. vrst | 10 | 8 | 5 | 10 | 11 |
| Ocena | Redčenje na 18 osebkov | | | | |
| Št. osebkov | 18 | 18 | 18 | 18 | 18 |
| Št. vrst | 6,9 | 7,7 | 4,5 | 8,5 | 8,3 |

```
R >
# Redčenje – vrstno bogastvo
rarefy(data, sample=18)
```

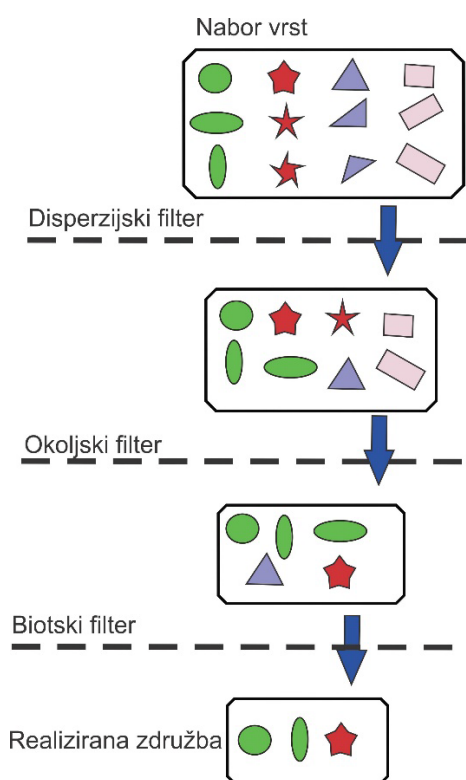


3.4 Temna diverziteta

Za razlago koncepta **temne diverzitete** (ang. Dark Diversity) uvedemo pojme: nabor vrst SP (ang. Species Pool), tarčna združba FC (ang. Focal Community) in filtriranje vrst (ang. Species Filtering). **Nabor vrst** definiramo s tremi lastnostmi: 1. prostorsko, z opredelitvijo regije; 2. ekološko, z opredelitvijo habitatnega tipa ali skupine sorodnih habitatnih tipov; 3. filogenetsko, z opredelitvijo taksonomske skupine organizmov. To pomeni, da nabor vrst sestavljajo vrste znotraj regije, ki naseljujejo podobno skupino habitatov, npr. vrste malih sesalcev v iglastem gozdu na Smrekovcu, ali pa suhe južine na traviščih nad gozdno mejo na Grintovcu. Nabor vrst je lahko enak številu vrst v regiji, kadar celotna regija pripada samo enemu habitatnemu tipu ali skupini podobnih habitatov. V tem primeru je nabor vrst sinonim za vrstno bogastvo regije oz. diverziteto gama (primerjaj s Poglavjem 3.1). Nabor vrst predstavlja potencial vrst, ki jih lahko opazujemo na posamezni mikrolokaciji oz. na konkretnem vzorčenem mestu znotraj regije. V primeru, da regija vključuje več izrazito različnih habitatnih tipov, npr. travniško združbo in ribnik, potem vsak posamezni habitatni tip oblikuje lasten nabor vrst, npr. nabor žuželčnih vrst travnika in nabor žuželčnih vrst ribnika. **Tarčna združba** je nabor vrst na konkretni lokaliteti in je sinonim za točkovno oz. alfa diverziteto.

Opazovano vrstno bogastvo v tarčni združbi (to je njena alfa diverziteta) je manjše od nabora vrst regije. Razlika v številu vrst je posledica več mehanizmov, ki preprečujejo, da bi vse vrste iz nabora vrst poselile konkretno mikrolokacijo. Mehanizem, ki vzdržuje razliko med alfa diverziteto in gama diverziteto, se imenuje **filtriranje vrst**. V analizi biodiverzitetnih vzorcev govorimo o treh tipih filtrov: 1. filter disperzijske mobilnosti; 2. abiotski filter okolja; 3. biotski filter medvrstnih interakcijski (Cadott in Tucker 2017).

Filter disperzijske mobilnosti najprej izloči vrste, ki zaradi omejitev v mobilnosti ne morejo dostopiti do mikrolokacije. Naslednja stopnja izločevanja je filtriranje vrst v **gradientih abiotskih dejavnikov okolja**, filtri okolja izločijo vrste, pri katerih vsaj eden ali več abiotskih dejavnikov presega njihovo strpno območje. Tretji nivo izločanja vrst predstavljajo **biotski odnosi med vrstami**. Biotski filtri izločijo vrste predvsem zaradi kompeticije in tudi predacije in parazitizma (Slika 3.25). V realni združbi vsi trije procesi delujejo diskretno in po sekvencah, kar pomeni, da delujejo brez izjem in v zaporedju.



Slika 3.25 Trije tipi filtriranja vrst s katerimi razložimo razliko med začetnim naborom vrst (gama diverziteta) in vrstnim bogastvom na konkretni lokaliteti (alfa diverziteta), oz. sestavi vrst v tarčni združbi. Simboli predstavljajo posamezne vrste iz nabora vrst. Disperzijski filter izloči vrste, ki se ne uspejo razširiti na mikrolokacijo. Abiotski filter okolja izloči vrste, za katere abiotski dejavniki niso vzdržni. Biotski filter izloči vrste, kot posledico medvrstnih odnosov npr. kompeticije (prirejeno po Cadotte in Tucker 2017).

Posledica filtriranja vrst so razlike v alfa diverzitetah med posameznimi vzorčnimi mesti znotraj regije, oz. filtriranje vrst preprečuje, da bi se alfa diverzitete izenačile z gama diverzitetjo. S terminom **temna diverziteta** definiramo delež iz nabora vrst, ki jih nismo potrdili v tarčni združbi. To pomeni, da so to lokalno odsotne vrste, oz. je njihova zastopanost na lokalnem nivoju preredka, da bi jih zaznali. Z drugimi besedami, vrste temne diverzitete so tiste, ki so prisotne na večini drugih vzorčenih mestih v isti regiji, niso pa v tarčni združbi, imajo pa potencial in možnost naseliti tarčno mesto.

Temno diverziteto za posamezno vzorčeno mesto DD_p izračunamo kot razliko med naborom vrst SP in vrstnim bogastvom tarčne združbe FC :

$$DD_p = SP - FC .$$

Metodologija izračuna relativne vrednosti temne diverzitete, predvsem za namene primerjav med prostorsko časovnimi vzorci, še ni stabilna. V uporabi sta dva pristopa: izračun *indeksa popolnosti združbe* in izračun *Bealovega indeksa*.

Indeks popolnosti združbe

Indeks popolnosti združbe CCI (ang. Community Completeness Index) (Pärtel in sod. 2013) je koncept, pri katerem ocenimo obseg temne diverzitete za posamezno vzorčeno mesto in predstavlja logaritem razmerja med bogastvom tarčne združbe FC oz. opazovanim bogastvom vrst (ang. Observed Richnes) in temno diverziteteto DD_p :

$$CCI = \ln \frac{FC}{DD_p}.$$

Indeks CCI računamo za vzorčena mesta, na katerih so prisotne vrste iz temne diverzitete. Če je alfa diverziteteta enaka gama diverziteti indeksa ne moremo izračunati oz. ni smiseln. Indeks CCI ima pozitivno vrednost, ko je število vrst temne diverzitete manjše od števila prisotnih vrst ($FC > DD$), in ima negativno vrednost, kadar število vrst temne diverzitete presega število opazovanih vrst ($FC < DD$). Vrednost indeksa CCI nič ($CCI = 0$) pomeni, da je število manjkajočih vrst enako številu opazovanih vrst.

Z indeksom CCI primerjamo zastopanost temne diverzitete bodisi v geografskem gradientu med vzorčenimi lokalitetami znotraj regije bodisi v časovni vrsti dolgoročnih raziskav stabilnosti združbe oz. odzivu združbe na motnje v habitatu, ali pa z njim merimo uspešnost ekološke obnove habitata.

Bealov indeks

Bealov indeks (ang. Beals Index) daje verjetnost, da je določena vrsta prisotna na izbrani lokaliteti. Izračun indeksa temelji na **konceptu sovrst** (ang. Co-occured Species). Sovrsti sta dve vrsti, torej par vrst, ki sta pogosto skupaj prisotni na posameznih lokalitetah. Glede na Bealov pristop lahko eno vrsto iz para sovrst uporabimo kot indikator za napoved prisotnosti druge vrste na lokaliteti s podobnim okoljem. Na podlagi števila znanih skupnih opazovanj dveh vrst, temelji napoved verjetnosti opazovanja vrste j na novi lokaliteti i , če je na njej prisotna vrsta k . Vrednost Bealovega indeksa za prisotnost vrste j na lokaliteti i je lahko od 0 do 1, z višanjem vrednosti indeksa se povečuje verjetnost za prisotnost vrste j . Vrednost indeksa je tudi merilo ustreznosti habitata za vrsto j . Vrednosti bližje 1 pomenijo višjo verjetnost prisotnosti vrste in ustrežnejši habitat za vrsto j .



PRIMER 3.12

Temna diverziteta – indeks popolnosti združbe

Nabor vrst suhih južin v petih habitatih visokogorskih travnikov na območju Kamniško Savinjskih alp obsega 14 vrst, njihova prisotnost po habitatih je predstavljen v Preglednici 3.2. Tarčna diverziteta (alfa diverziteta) po habitatnih tipih obsega od 5 do 10 vrst (Preglednica 3.8), temna diverziteta oz. odsotne vrste pa od 3 do 9 vrst (Preglednica 3.8). Najvišjo vrednost indeksa popolnosti združbe $CCI = 1,3$ dosega habitat 31.52, najnižjo vrednost $CCI = -0.6$ pa habitat 61.231 (Preglednica 3.11).

Preglednica 3.11. Vrstno bogastvo tarčne združbe FC in temna diverziteta DD za 14 vrst, ki sestavljajo nabor vrst suhih južin na lokalitetah v petih travniških habitatnih tipih HT nad gozdno mejo. Indeks popolnosti združb CCI

| HT | FC | DD | CCI |
|--------|----|----|--------|
| 36.413 | 10 | 4 | 0,916 |
| 36.433 | 8 | 6 | 0,288 |
| 61.231 | 5 | 9 | -0,588 |
| 61.232 | 10 | 4 | 0,916 |
| 31.52 | 11 | 3 | 1,299 |



Podatkovna matrika X za izračun Bealovega indeksa ima dimenzijo r vrstic z vzorčenimi mesti (lokalitete) in p stolpcev z vrstami, vrednosti v poljih so PA podatki (npr. Preglednica 3.2). To pomeni, da posamezno polje vsebuje vrednost ena ali nič, oz. je vrsta k na lokaliteti i prisotna ali odsotna, kar označimo z x_{ik} . Prvi korak v izračunu Bealovega indeksa je izdelava simetrične matrike parov sovrst, kjer za vsak par vrst j in k izračunamo število skupnih pojavljanj m_{jk} . Bealov indeks b_{ij} daje verjetnost prisotnosti vrste j na lokaliteti i , izračunamo ga s formulo:

$$b_{ij} = \frac{1}{s_i} \sum_{k=1}^p \frac{m_{jk} x_{ik}}{n_k},$$

kjer je s_i število vrst na lokaliteti i in n_k število sovrst vrste k (de Cáceres in Legendre 2008, Brown is sod. 2019).



PRIMER 3.13

Temna diverziteta – Bealov indeks

Matriko habitati x vrste (Preglednica 3.2) smo vključili v izračun verjetnosti za prisotnost posamezne vrste v posameznem habitatnem tipu. Napoved je narejena na konceptu sovrst, to pomeni, če sta dve vrsti večkrat prisotni skupaj, bosta verjetno prisotni skupaj tudi v naslednjem habitatu. Najvišjo maksimalno verjetnost prisotnosti imata vrsti MITMOR in NEMTRI $b_{ij} = 1$, verjetnost prisotnosti na izbranih lokalitetah je 100 %. Najnižjo povprečno vrednost Bealovega indeksa ima vrsta TROFAL $b_{ij} = 0,22$ (Preglednica 3.12), njena povprečna verjetnost prisotnosti na lokaliteti je 22 %. Absolutno najnižjo verjetnost prisotnosti $b_{ij} = 0,15$ ima vrsta GYAANN v habitatu 61.231. Izračun Bealovega indeksa izvedemo z ukazom *beal* v paketu *vegan*.

| Sp / HT | 36.413 | 36.433 | 61.231 | 61.232 | 31.52 | povprečje |
|---------|--------|--------|--------|--------|-------|-----------|
| HISDEN | 0,553 | 0,692 | 0,540 | 0,720 | 0,730 | 0,647 |
| PLABUC | 0,720 | 0,650 | 0,540 | 0,603 | 0,715 | 0,646 |
| MITMOR | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |
| LOPPAL | 0,753 | 0,608 | 0,540 | 0,720 | 0,791 | 0,683 |
| GYAANN | 0,223 | 0,196 | 0,147 | 0,240 | 0,370 | 0,235 |
| DICGAS | 0,902 | 0,877 | 0,737 | 0,902 | 0,933 | 0,870 |
| TROROS | 0,428 | 0,619 | 0,657 | 0,578 | 0,427 | 0,542 |
| TRONEP | 0,670 | 0,504 | 0,607 | 0,520 | 0,639 | 0,588 |
| NEMTRI | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |
| TROFAL | 0,348 | 0,185 | 0,197 | 0,182 | 0,203 | 0,223 |
| NEMBBI | 0,902 | 0,877 | 0,737 | 0,902 | 0,933 | 0,870 |
| MITALP | 0,777 | 0,804 | 0,853 | 0,760 | 0,630 | 0,765 |

Preglednica 3.12 Vrednosti Bealovega indeksa, ki na podlagi koncepta sovrst, daje verjetnosti prisotnosti posamezne vrste v posameznem habitatu.

```
R > package = »vegan«
d<-beals(dd, species=NA,
include=TRUE)
d
```



3.5 Prikrita diverziteta

Prikrito diverziteto (ang. Cryptic Diversity) tvori skupina sorodnih organizmov, ki so fenotipsko zelo podobni, morfološke razlike med njimi niso zaznavne, ali pa so zelo slabo izražene. Za opis tega fenomena je v uporabi več sinonimov ali smiselno podobnih terminov: kompleks vrst (ang. Species Complex), vrste dvojčice (ang. Sibling Species), agregat vrst (ang. Species Agregat), vrsta v širšem pomenu (ang. Sensu Lato Species).

Prikrito diverzitet

o običajno oblikuje monofiletska skupina v zgodnji fazi speciacije. Lahko gre tudi za vrste, ki so ločene že dolgo časovno obdobje, vendar pri njih ni prišlo do morfološke diferenciacije, ali pa gre za konvergentni razvoj. Prikrita diverziteto je prisotna pri vseh skupinah organizmov. Prepoznamo jo lahko s poglobljenim proučevanjem podrobnih morfoloških lastnosti, ali s preverjanjem reproduktivne izolacije, ali pa z

molekularnimi metodami analize DNK. Prisotnost prikrite diverzitete privede do podcenjene vrstne pestrosti.

V vretenčarski favni Slovenije je prisotnih več primerov kompleksov vrst, npr. med sesalci travniška voluharica *Microtus agrestis/lavernedii*, med plazilci kuščar zelenec *Lacerta viridis/trilineata/bilineata*, med dvoživkami zelena žaba *Pelophylax k.l. esculenta*.

Največje presenečenje so odkrile raziskave močerila *Proteus anguinus*, kjer so prepoznali devet prikritih vrst (Recknagel in sod. 2023).

Zaključki

- Vrstno bogastvo je skupno število vrst v vzorcu, združbi ali območju.
- S povečanjem raziskovalnega dela zmanjšamo pristransko podcenjeno število vrst. Raziskanost ocenimo z akumulacijsko krivuljo vrst, ko krivulja doseže asimptoto, je raziskanost števila vrst ustrezna.
- Ocena vrstnega bogastva s pomočjo akumulacijske krivulje vrst izvajamo s stališča števila vzorčenj ali z vidika števila prešteti osebkov. Ocene za obe skupini podatkov izhajata iz števila vrst enojcev in dvojcev oz. deleža redkih vrst v združbi.
- Primerjave vrstnega bogastva med združbami, ki so raziskane z različno količino raziskovalnega dela, izvajamo z redčenjem. To je z ocena števila vrst različnih združb, ki jih standardiziramo na enako količino raziskovalnega dela. Standardizacija je lahko izvedena glede na število osebkov ali glede na število vzorčenj.
- Z enako intenzivnim in konstantnim vzorčenjem kumulativno število prešteti osebkov linearno narašča skupaj s številom vzorčenj.
- Združbo raziskujemo z vidika lokalne – alfa in regionalne – gama biodiverzitete. Znotraj ene regije je prisotnih več lokalnih alfa diverzitet, ki skupaj sestavljajo gama diverzitetu.
- Biodiverziteta v prostoru in času ni konstantna, vrstna sestava združbe se spreminja vzdolž gradienta okolja in med habitati. To povzroči spremembe v številu vrst, spremembe v vrstni sestavi ter izmenjave vrst.
- Znotraj beta diverzitete razlikujemo dve kategorij vrst: ugnezdene vrste so prisotne v vseh lokalnih favnah; vrste prostorskega obrata se izmenjujejo vzdolž gradienta.
- Primerjave vrstne sestave med posameznimi združbami izvajamo s cenilkami parnih primerjav, ob hkratni primerjavi treh ali več lokalnih diverzitet jih vrednotimo s cenilkami mnogoterih primerjav.
- Zeta diverzitetu je analiza vrstnega obrata med mnogo lokalitetami.
- Terensko vzorčenje z namenom evidentiranja vrstnega bogastva izbrane skupine organizmov načrtujemo v skladu z optimalno količino raziskovalnega dela, da

ujamemo ravnotežje med ustrezno raziskanostjo ter časovno in materialno naložbo v izvedbo vzorčenja.

- Pomembno informacijo o združbi dobimo tudi iz pričakovanih vendar manjkajočih vrst v združbi, to imenujemo temna diverziteteta. Temno diverziteteto izrazimo kot razliko med pričakovanim in opazovanim vrstnim bogastvom, njeno relativno vrednost pa z indeksom popolnosti združbe in z Bealovim indeksom verjetnosti prisotnosti vrste na lokaliteti.
- Koncept temne diverzitetete temelji na lastnostih in procesih v združbi: nabor vrst, tarčna združba in filtriranje vrst.
- Kompleks vrst je skupina sorodnih organizmov, ki so fenotipsko zelo podobni, morfološke razlike med njimi niso izražene. Kompleks vrst je praviloma monofiletska skupina organizmov bodisi v zgodnji fazi speciacije bodisi že dolgo časovno obdobje ločene skupine, pri katerih ni prišlo do morfološke diferenciacije.

Literatura

- Baselga A. 2010. Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, 19: 134–143.
- Beta diversity. 2019. Wikipedia. Pridobljeno dne 22. 9. 2019, iz https://en.wikipedia.org/wiki/Beta_diversity.
- Brown J.J., Mennicken S., Massante J.C., Dijoux S., Telea A., Benedek M.A., Götzenberger L., Májeková M., Lepš J., Šmilauer P., Hrček J., de Bello F. 2019. A novel method to predict dark diversity using unconstrained ordination analysis. *Journal of Vegetation Science*, 00: 1–10. <https://doi.org/10.1111/jvs.12757>.
- Cadotte M.W., Tucker C.M. 2017. Should Environmental Filtering be Abandoned? *Trends in Ecology & Evolution*, 32(6): 429–437.
- Colwell R.K. 2013. EstimateS: Statistical estimation of species richness and shared species from samples. Version 9. User's Guide and application published at: <http://purl.oclc.org/estimates>.
- de Cáceres M., Legendre P. 2008. Beals smoothing revisited. *Oecologia*, 156(3): 657–669. <https://doi.org/10.1007/s00442-008-1017-y>.
- Gotelli N.J., Chao A. 2013. Measuring and Estimating Species Richness, Species Diversity, and Biotic Similarity from Sampling Data. V: Levin S.A. (ed.) *Encyclopedia of Biodiversity*. Second edition, Volume 5, pp. 195–211. MA: Academic Press. Waltham.
- Hui C., McGeoch A.M. 2014. Zeta Diversity as a Concept and Metric That Unifies Incidence-Based Biodiversity Patterns. *The American Naturalist*, 184(5): 684–694.
- Jurasinsky G., Retzer V., Beierkuhnlein C. 2009. Inventory, differentiation, and proportional diversity: a consistent terminology for quantifying species diversity. *Oecologia*, 159:15–26.
- Koleff P., Gaston K.J., Lennon J.J. 2003. Measuring beta diversity for presence–absence data. *Journal of Animal Ecology*, 72: 367–382.
- Legendre P., Legendre L. 2012. *Numerical Ecology*. Elsevier. Amsterdam.
- McGeoch M.A., Latombe G., Andrew N.R., Nakagawa S., Nipperess D.A., Roige M., Marzinelli E.M., Campbell A.H., Verges A., Thomas T., Steinberg P.D., Selwood K.E., Henriksen M.V., Hui C. 2019. Measuring continuous compositional change using decline and decay in zeta diversity. *Ecology* 100(11): e02832.0.1002/ecy.2832.
- Novak T., Janžekovič F., Rak M., Ivajnsič D., Kozel P., Slana Novak L., Škornik S. 2017. Can highland habitat type distributions replace mapping of harvestman species? *Journal of Insect Conservation*, 21: 331–343.
- Oksanen J. 2022. Vegan: ecological diversity. Processed with vegan 2.6-4 in R version 4.2.1 (2022-06-23) on October 11, 2022.
- Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., Wagner H. 2019. *vegan: Community Ecology Package*. R package version 2.5-6. <https://CRAN.R-project.org/package=vegan>.
- Pärtel M., Szava-Kovats R., Zobel M. 2013. Community Completeness: Linking Local and Dark Diversity within the Species Pool Concept. *Folia Geobotanica*, 48: 307–317. DOI 10.1007/s12224-013-9169-x.
- Recknagel H., Zakšek V., Delić T., Gorički Š., Trontelj P. 2023. Multiple transitions between realms shape relict lineages of *Proteus* cave salamanders. *Molecular Ecology*, 00, 1–16. <https://doi.org/10.1111/mec.168681365294x>.
- Riva F., Mammola S. 2021. Rarity facets of biodiversity: Integrating Zeta diversity and Dark diversity to understand the nature of commonness and rarity. *Ecology and Evolution*, 11: 13912–13919.
- Tome D. 2006. *Ekologija : organizmi v prostoru in času*. Tehniška založba Slovenije. Ljubljana.

4. poglavje



VRSTNA PESTROST

Povzetek

4. poglavje: Vrstna pestrost

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

Vrstna pestrost je razmerje med številom vrst (vrstno bogastvo) in številom osebkov posamezne vrste (vrstna abundanca). Cenilke vrstne pestrosti sestavljata komponenti: vrstno bogastvo (število vrst) in poravnano vrst (razmerja med vrstami glede na relativno število osebkov posamezne vrste). Primerjave vrstne pestrosti izvajamo glede na: območja, časovna obdobja, časovno zaporedje ali vzdolž gradientov (fizični gradient: temperatura, svetloba, globina; kemični gradient: slanost, pH; geografski gradient: geografska širina in dolžina, nadmorska višina; biotski gradient: kompetitorji; časovni gradient: časovna vrsta). Osnovni cenilki vrstne pestrosti sta Margalefov in Menhinickov indeks, v katerih je upoštevano samo razmerje med številom vrst in skupnim številom osebkov. Simpsonov indeks D je ocena verjetnosti, da dva naključna osebkoma pripadata isti vrsti. Z naraščanjem njegove vrednosti opisujemo nižanje pestrosti. Nasprotna vrednost Simpsonovega indeksa se imenuje Gini-Simpsonov indeks, z njegovim naraščanjem opisujemo večanje pestrosti. Tretja oblika je recipročni Simpsonov indeks, kjer je pestrost podana v enotah števila vrst. Shannonov indeks H' je ocena verjetnosti, da naključni osebek pripada določeni vrsti. Vrednost indeksa narašča s številom vrst in z enakomerno zastopano številom osebkov posamezne vrste. Indeks H' ima vrednost 0, če je v vzorcu samo ena vrsta in ima maksimalno vrednost, kadar imajo vse vrste enako število osebkov.

Interpretacije in primerjave pestrosti z diverzitetnimi indeksi so lahko težko razložljive, zato so nadgrajeni v Hillova diverzitetna števila H_q , kjer diverzitetno izrazimo v enotah števila vrst. Druga oblika diverzitetnih indeksov je njihova standardizacija, ki jo podajamo z izračunom vrstne poravnane vrednosti E_a , s katerim je pestrost izražena v deležu od maksimalne vrednosti 1. Za uravnotežene primerjave vrstne pestrosti med združbami moramo zagotoviti primerljiv vložek raziskovalnega dela. V primeru različne količine raziskovalnega dela surove podatke standardiziramo na enak raziskovalni vložek, na enako število osebkov in na enako vzorčno pokritost ali ocenjeni delež spregledanih osebkov. Abundanca vrst ni enakomerno razporejena, večina vrst je zastopana z nizkim številom osebkov, z visokim številom osebkov je zastopanih le nekaj vrst. V analizi vrstne abundance vrednotimo relativne odnose med vrstami. Porazdelitev vrstnih abundanc v združbi lahko prikažemo na tri različne načine s Fisherjevo logaritemsko porazdelitvijo, s Prestonova logaritemsko normalno porazdelitvijo in z Whittakerjevo porazdelitvijo rangiranih abundanc.

Analitska orodja za vrednotenje biodiverzitete so razvita v okolju R, prikazana je uporaba paketov »vegan« in »iNEXT«.

Vrstna pestrost (ang. Species Diversity) je razmerje med vrstnim bogastvom in abundanco vrst. Vrstno bogastvo podamo s številom vrst, abundanco pa s številom osebkov posamezne vrste.

V izračun cenilke vrstne pestrosti vključimo dve komponenti: **vrstno bogastvo** in **poravnanost vrst** (razmerja med vrstami glede na relativno število osebkov posamezne vrste, tj. relativna abundanca vrst).

Terminološko

Biodiverzitetu združbe ali vzorca vrednotimo z dvema osnovnima merama, to sta *vrstno bogastvo* in *vrstna pestrost*. Uporaba biodiverzitetnih terminov še ni ustaljena, še posebej terminov *vrstno bogastvo* in *vrstna pestrost*, zato je precej neenotnosti v strokovni in znanstveni literaturi. Termina se uporabljata tudi v različnih pomenih (*Sensu Lato* in *Sensu Stricto*). Dodatno zmedo v slovenščini predstavlja sočasna raba slovenskih in sloveniziranih angleških terminov.

Termin *vrstno bogastvo* (ang. Species Richness) v ožjem pomenu besede pomeni število vrst v vzorcu. V uporabi za opredelitev vrstnega bogastva je tudi termin *vrstna pestrost*, ki je v tem kontekstu sinonim oz. širša raba termina.

Termin *vrstna pestrost* (ang. Species Diversity) pomeni razmerje med vrstnim bogastvom in relativno abundanco vrst v vzorcu. V slovenščini je v uporabi tudi slovenizirani termin *vrstna diverziteteta*, ki je v tem kontekstu sinonim za vrstno pestrost.

Primerjave biodiverzitet med vzorci se lahko nanašajo na primerjanje med različnimi območji, med različnimi časovnimi obdobji, v časovnem zaporedju ali vzdolž gradientov (npr. fizični gradient: temperatura, svetloba, tlak, elektroprevodnost; kemični gradient: slanost, pH, trdota vode; geografski gradient: geografska širina, geografska dolžina, nadmorska višina; biotski gradient: kompetitorji; časovni gradient: časovna vrsta).

V tem poglavju obravnavamo naslednje vsebine:

- Cenilke vrstne pestrosti.
 - Simpsonovi indeksi diverzitet D .
 - Shannonov indeks H' .
- Hillova diverzitetna števila H_a .
- Vrstna poravnanost E_a .
- Primerjave vrstne pestrosti med združbami.
- Relativna vrstna abundanca.

Vsebino poglavja o vrstni pestrosti smo črpali iz virov: Clarke in sod. 2014, Gotelli in Graves 1996, Hsieh in sod. 2016 in 2019, Legendre in Legendre 2012, Oksanen 2022, Oksanen in sod. 2019, Peter 1995, Tome 2006, Upton 2020 in Zelený 2020.

4.1 Cenilke vrstne pestrosti

Osnovne cenilke vrstne pestrosti, ki temeljijo na razmerju med številom vrst S in številom vseh osebkov v vzorcu N , so Margalefov indeks $R1$, Menhinickov indeks $R2$ in Odumov indeks $R3$:

$$R1 = \frac{S-1}{\ln(N)}; \quad R2 = \frac{S}{\sqrt{N}}; \quad R3 = \frac{S}{\log(N)}.$$

Od kompleksnejših indeksov predstavljamo Simpsonov in Shannonov indeks diverzitete.

Simpsonovi indeksi diverzitete

Simpsonovi indeksi diverzitete so trije med seboj sorodni indeksi: Simpsonov indeks D , Simpsonov indeks diverzitete $1-D$ in Simpsonov recipročni indeks $1/D$.

Simpsonov indeks D je izračun verjetnosti, da dva osebka, ki ju naključno odvezamemo iz vzorca, pripadata isti vrsti. Izračunamo ga po formuli:

$$D = \sum_{i=1}^S p_i^2,$$

kjer so: D – Simpsonov indeks, p_i – relativna abundanca vrste i ($p_i = \frac{n_i}{N}$), n_i – število osebkov vrste i , S – število vrst, N – število osebkov vseh vrst v vzorcu.

Simpsonov indeks je ocena verjetnosti z naborom vrednosti od 0 do 1. Če je izračun D visok in se približuje vrednosti 1, se večja verjetnost, da sta osebka iste vrste. To pomeni, da je pestrost vzorca nizka, in obratno oz. višja, kot je vrednost D , nižja je pestrost. Tak zapis intuitivno ni logičen, zato naredimo preračun in vrednost D preobrnemo. Lahko jo odštejemo od 1 in dobimo Simpsonov indeks diverzitete ($1-D$), ali pa vrednost 1 delimo z D in dobimo Simpsonov recipročni indeks ($1/D$).

Simpsonov indeks diverzitete 1-D (ali Gini-Simpsonov indeks)

Indeks izračunamo tako, da od 1 odštejemo D :

$$1 - D = 1 - \sum_{i=1}^S p_i^2,$$

kjer so $1-D$ – Simpsonov indeks diverzitete, S – število vrst, p_i – relativna abundanca vrste i ($p_i = \frac{n_i}{N}$), n_i – število osebkov vrste i , N – število osebkov vseh vrst v vzorcu.

Simpsonov indeks diverzitete $1-D$ ima nabor vrednosti od 0 do 1 in je ocena verjetnosti, da dva osebka, ki ju naključno odvzamemo iz vzorca, pripadata isti vrsti. Če je verjetnost velika, sta osebka različne vrste, potem je diverziteteta vzorca visoka. Torej, višja kot je vrednost $1-D$, višja je diverziteteta.

Simpsonov recipročni indeks $1/D$

Indeks izračunamo tako, da 1 delimo z D :

$$\frac{1}{D} = \frac{1}{\sum_{i=1}^S p_i^2},$$

kjer so $1/D$ – Simpsonov recipročni indeks, S – število vrst, p_i – relativna abundanca vrste i ($p_i = \frac{n_i}{N}$), n_i – število osebkov vrste i , N – število osebkov vseh vrst v vzorcu.

Vrednost tega indeksa se začne z 1 kot najnižjo možno vrednostjo in predstavlja vzorec z eno samo vrsto. Višja kot je vrednost, višja je pestrost. Najvišja možna vrednost indeksa je določena s številom vrst v vzorcu. Če je npr. v vzorcu pet vrst, je največja možna vrednost 5.

Shannonov indeks H'

Shannonov indeks H' je najpogosteje uporabljen indeks diverzitete v ekologiji in z njim ocenimo verjetnost, da bo osebek, ki ga naključno izberemo izmed S vrst in N osebkov, pripadal definirani vrsti. Ta povprečna stopnja nedoločenosti (ang. Uncertainty) narašča z naraščanjem števila vrst in z izenačevanjem številčnosti osebkov posameznih vrst. Izpostavljamo dve lastnosti H' :

- 1) $H' = 0$, ko je v vzorcu samo ena vrsta;
- 2) $H' =$ maksimalen, ko so vse vrste enako številčne (so v vzorcu zastopane z enakim številom osebkov).

Enačba za Shannonov indeks:

$$H' = - \sum_{i=1}^S p_i \ln p_i ,$$

kjer so H' – Shannonov indeks, S – število vrst, p_i – relativna abundanca vrste i ($p_i = \frac{n_i}{N}$), n_i – število osebkov vrste i , N – število vseh osebkov v vzorcu.

V strokovni literatu je za Shannonov indeks v uporabi več sinonimov, npr: Shannonov diverzitetni indeks (ang. Shannon Diversity Index), Shannon-Wienerjev indeks (ang. Shannon-Wiener Index), Shannon-Weaverjev indeks (ang. Shannon-Weaver Index) in Shannonova entropija (ang. Shannon Entropy).

Indeksi diverzitete združujejo **vrstno bogastvo** (število vrst), **vrstno pestrost** (število vrst v kombinaciji s številom osebkov teh vrst) in **pogostnost** (sin. incidenca, t.j. prisotnost izbrane vrste v zaporedju vzorcev), zato jih nekateri imenujejo tudi **indekse heterogenosti** (ang. Heterogeneity Indices). Šibkost biodiverzitetnih indeksov je v njihovi interpretaciji. V nekaterih primerih so lahko vrednosti indeksa diverzitete enake, čeprav izhajajo iz različnih kombinacij števila vrst in številčnosti. Zato je le na osnovi vrednosti indeksa (biodiverziteta *per se*) nemogoče ugotoviti, kolikšen je njihov absoluten pomen.



PROJEKT 4.1

Akcija »Zimsko štetje vodnih ptic«

Zimsko štetje vodnih ptic (IWC – International Waterbirds Census) je svetovna akcija zbiranja podatkov o prezimovanju vodnih ptic. V akciji IWC na sistematičen način zbirajo podatke o vodnih pticah in s tem zagotavljajo dolgoročni monitoring, s katerim spremljajo naravne procese. Obenem je namen akcije popularizacija opazovanja ptic, delovanja v korist narave in druženje zagovornikov narave. Slovenski koordinatorski center akcije je Društvo za opazovanje in proučevanje ptic Slovenije DOPPS (Slika 4.1). Rezultate vsakoletnega štetja objavijo v znanstveni ornitološki reviji *Acrocephalus*, iz teh člankov smo povzeli podatke za to analizo (Božič 2010, 2012, 2013, 2014, 2015, 2016, 2017 in 2018).

Glavni cilji IWC so: monitoring številčnosti populacij vodnih ptic; opis sprememb v številu in razširjenosti vrst; evidentiranje mokrišč mednarodnega pomena; zagotavljanje informacij z namenom varovanje in upravljanja populacij in habitatov vodnih ptic ter izvajanje mednarodne in nacionalne naravovarstvene zakonodaje.

Predstavitev poslanstva DOPPS, vzeto iz spletne strani: <https://www.ptice.si/>

Društvo za opazovanje in proučevanje ptic Slovenije (DOPPS) je neprofitna nevladna organizacija, ustanovljena leta 1979, ki deluje v javnem interesu ohranjanja narave, varstva okolja in na področju raziskovalne dejavnosti. Pri številnih dejavnostih združujemo več kot 1000 članov, prostovoljcev in ljubiteljev narave po vsej Sloveniji, in zaposluje 30 strokovnjakov z različnih področij. DOPPS dela za varstvo ptic in njihovih življenjskih okolij. S tem prispeva k ohranjanju narave in dobrobiti celotne družbe. Njegovo poslanstvo se uresničuje z naravovarstvenim, znanstvenoraziskovalnim, izobraževalnim in publicističnim delom ter s sodelovanjem z drugimi nevladnimi in vladnimi organizacijami. DOPPS je polnopravni partner svetovne organizacije za varstvo narave BirdLife International.



Slika 4.1 V Mariboru so se zimskega štetja vodnih ptic udeležili študentje Biologije in ekologije z naravovarstvom iz Fakultete za naravoslovje in matematiko Univerze v Maribor (levo).

Foto. Eva Horvat.

Samica žvižgavke *Anas penelope*, redna prezimovalka na celinskih vodah (desno).

Foto. Alen Ploj.



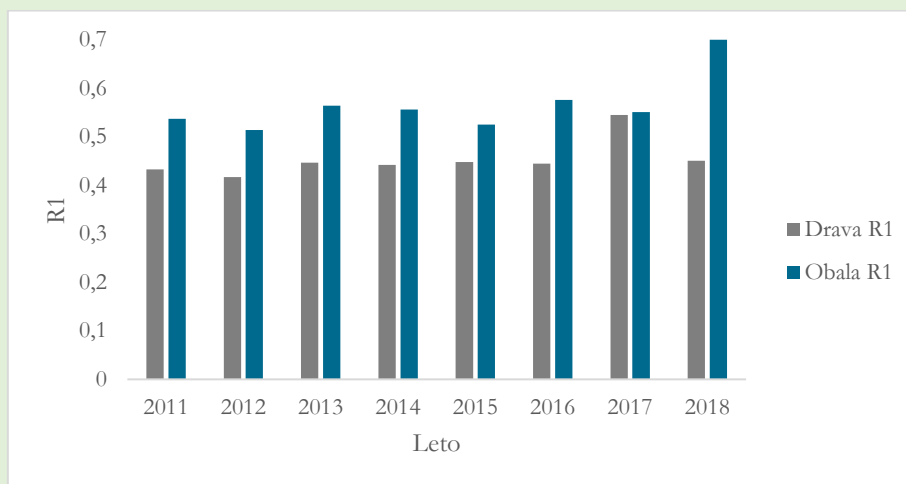
| območje | Obala | | | | | | | |
|---------------------------|-------|------|------|------|------|------|------|------|
| Vrsta/leto | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
| <i>Anas penelope</i> | 153 | 29 | 21 | 77 | 123 | 81 | 60 | 46 |
| <i>Anas strepera</i> | 44 | 18 | 7 | 2 | 4 | 16 | 30 | 23 |
| <i>Anas crecca</i> | 261 | 833 | 290 | 451 | 538 | 119 | 239 | 270 |
| <i>Anas platyrhynchos</i> | 1157 | 1440 | 807 | 684 | 1154 | 738 | 1037 | 664 |
| <i>Anas acuta</i> | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| <i>Anas clypeata</i> | 97 | 76 | 75 | 111 | 211 | 78 | 49 | 145 |
| Osebkki SKUPAJ | 1712 | 2396 | 1200 | 1325 | 2030 | 1032 | 1415 | 1149 |
| Vrste SKUPAJ | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 6 |



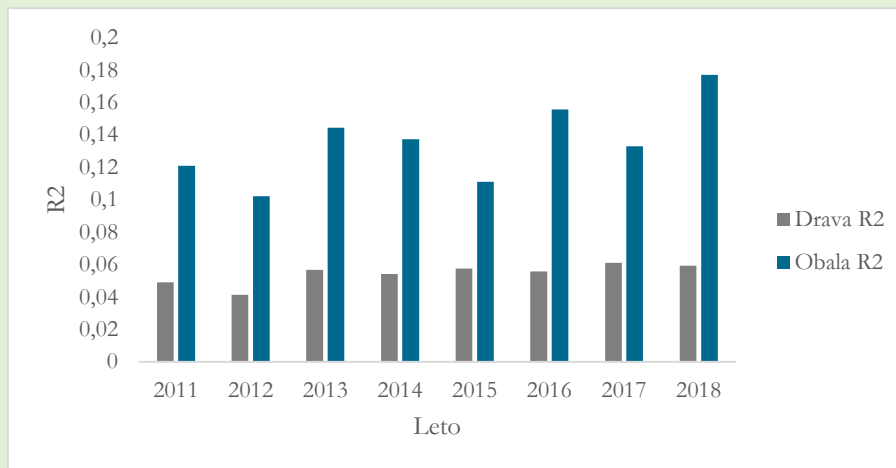
PRIMER 4.2

Zimsko štetje vodnih ptic – R1 in R2

Grafična predstavitev najpreprostejših indeksov pestrosti R1 in R2 (Sliki 4.4 in 4.5), ki temeljita na razmerju med številom vrst in številom osebkov. Število vrst na obeh lokalitetah in po letih je skoraj izenačeno, medtem je na lokaliteti Drava bistveno večje število osebkov. Vidimo, da večanje števila osebkov ob konstantnem številu vrst, znižuje vrednosti obeh indeksov, zato so vrednosti R1 in R2 višje na Obali. Indeksa nista več v vsakdanji rabi, saj je njuna uporaba omejena na posebne primere, interpretacija v kompleksnejših okoliščinah je problematična.



Slika 4.4 R1 – Margalefov indeks pestrosti prezimujočih rac rodu *Anas* na dveh lokalitetah po letih.



Slika 4.5 R2 – Menhinickov indeks pestrosti prezimujočih rac rodu *Anas* na dveh lokalitetah po letih.

```

R >
apply(data > 0, 1, sum) # Izpis števila vrst po letih
apply(data, 1, sum) # Izpis števila osebkov po letih
# R2
n <- apply(data > 0, 1, sum)
N <- apply(data, 1, sum)
R2 <- n / sqrt(N)
R2
# R1
n <- apply(data > 0, 1, sum)
N <- apply(data, 1, sum)
R1 <- (n - 1) / log(N)
R1

```

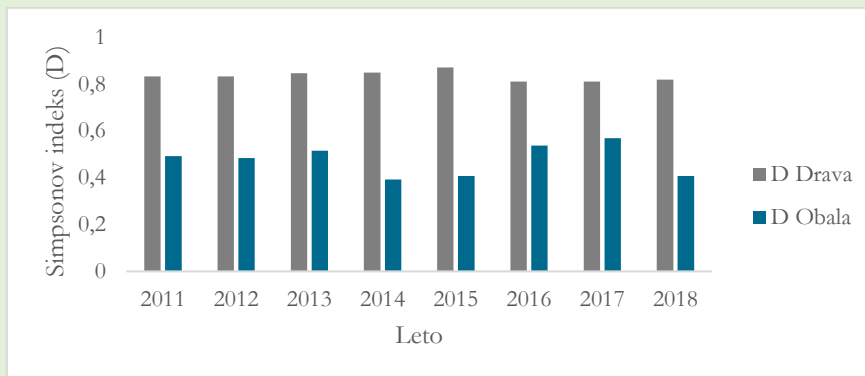


PRIMER 4.3

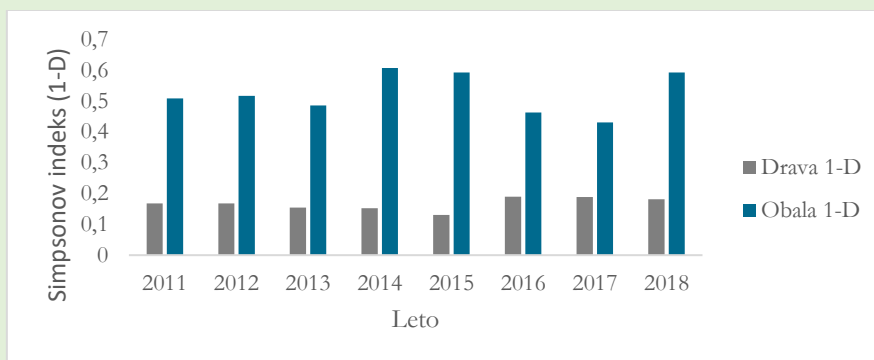
Zimsko štetje vodnih ptic – *D*

Simpsonov indeks je izračun verjetnosti, da dva osebkata naključno odvzeta iz vzorca, pripadata isti vrsti. Višja vrednost indeksa pomeni nižjo pestrost in obratno nižja vrednost indeksa pomeni višjo vrednost. Z drugimi besedami, bolj kot so vrste številčno enakomerno zastopane, nižja je vrednost indeksa, in obratno, če v vzorcu prevladuje ena vrsta, se vrednost indeksa večja. Absolutno število vrst in/ali osebkov ne vpliva izrazito na velikost indeksa. Opisani primer je predstavljen na Sliki 4.6, kjer je indeks *D* visok na Dravi diverzitetata pa nižja, prevladuje samo ena vrsta, ki večja vrednost indeksa, saj se povečuje verjetnost, da dva naključna osebkata pripadata isti vrsti, to je raci mlakarici. Izpeljanka, v kateri z enostavnim računskim postopkom $(1-D)$ obrnemo vrednost indeksa, da postane bolj logičen, se imenuje Gini-Simpsonov indeks (Slika 4.7). Z njegovim naraščanjem se

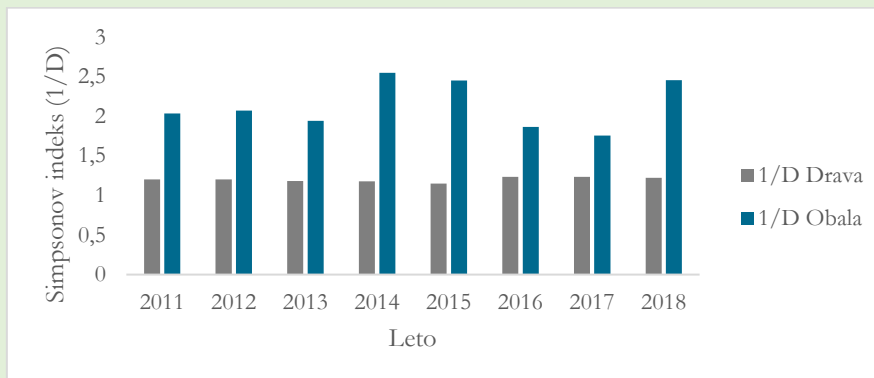
veča diverziteteta. Tretja oblika je recipročni Simpsonov indeks ($1/D$), v tem primeru je diverziteteta podana v številu vrst (Slika 4.8).



Slika 4.6 Simpsonov indeks D pestrosti prezimujočih rac rodu *Anas* na dveh lokalitetah po letih.



Slika 4.7 Simpsonov indeks $1-D$ pestrosti prezimujočih rac rodu *Anas* na dveh lokalitetah po letih.



Slika 4.8. Simpsonov indeks ($1/D$) pestrosti prezimujočih rac rodu *Anas* na dveh lokalitetah, po letih.

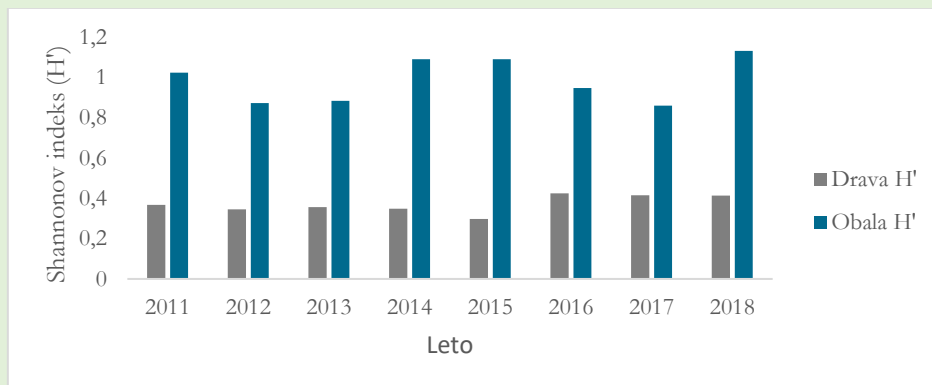




PRIMER 4.4

Zimsko štetje vodnih ptic – H'

Shannonov indeks je izračun verjetnosti, da bo naključno izbrani osebek pripadal definirani vrsti. Verjetnost doseže maksimum, ko so vse vrste zastopane z istim številom osebkov. V primeru rac je Shannonov indeks bistveno višji na Obali, saj so tam vrste veliko bolj enakomerno zastopane. Na lokaliteti Drava velika številčnost rase mlakarice zmanjšuje vrednost indeksa (Slika 4.9).



Slika 4.9 Shannonov indeks pestrosti prezimujočih rac rodu *Anas* na dveh lokalitetah po letih.

```
R > package = »vegan«
# H'
diversity(data, index="shannon")
# 1-D
diversity(data, index="simpson")
# Poravnost Pilou evenness (J)
S <- apply(data>0,1,sum)
diversity(data, index="simpson")/log(S)
```



4.2 Hillova diverzitetna števila in prava diverzitetna

Zaradi nedorečenosti v interpretaciji in primerjanju vrednosti diverzitetnih indeksov je prišlo do razvoja postopkov za njihovih standardizacijo. To pomeni, da je pestrost, podana v enotah števila vrst, govorimo o **efektivnem številu vrst** (ang. Effective Number of Species). V uporabi sta tudi termina **Hillova diverzitetna števila in prava diverzitetna** (ang. True Diversity). Z drugimi besedami: efektivno število vrst je mera za primerjanje pestrosti med vzorci. Na primer, če ima sosednja združba enako diverzitetno kot referenčna

združba, potem morata imeti obe združbi enako vrednost *prave diverzitete* oz. enako *efektivno število vrst*. Pravo diverzitetno izračunavamo s skupino indeksov, ki jih imenujejo **družina diverzitetnih števil** in izhajajo iz enačbe:

$$H_q = \frac{1}{1-q} \log \sum_{i=1}^S p_i^q,$$

kjer so H_q – Renyijeva entropija, q – red enačbe, S – število vrst, p_i – relativna abundanca posamezne vrste i ($p_i = \frac{n_i}{N}$), n_i – število osebkov vrste i , N – število vseh osebkov v vzorcu.

Izpeljave iz teh enačb ničelnega, prvega in drugega reda ($q = 0$, $q = 1$, $q = 2$) sovpadajo s tremi indeksi Hillovih diverzitetnih števil – N_d :

$$\begin{aligned} N_0 &= S \\ N_1 &= e^{H'} \\ N_2 &= 1/D \end{aligned}$$

kjer so: S – število vrst, H' – Shannon indeks, $1/D$ – Simpsonov recipročni indeks, e – matematična konstanta = 2,718. Diverzitetni indeksi in iz njega izvedena Hillova diverzitetna števila oz. prava diverzitetna so predstavljeni v Preglednici 4.2.

Preglednica 4.2 Diverzitetni indeksi, izvedena Hillova diverzitetna števila (sin. efektivno število vrst v vzorcu oz. prava diverzitetna) in red enačbe

| Diverzitetni indeks | Hilovo diverzitetno število | Red enačbe |
|-----------------------|--|------------|
| Vrstno bogastvo S | Vrstno bogastvo S | $q = 0$ |
| Shannonov indeks H' | Shannonova diverzitetna $e^{H'}$ | $q = 1$ |
| Simpsonov indeks D | Simpsonova recipročna diverzitetna $1/D$ | $q = 2$ |

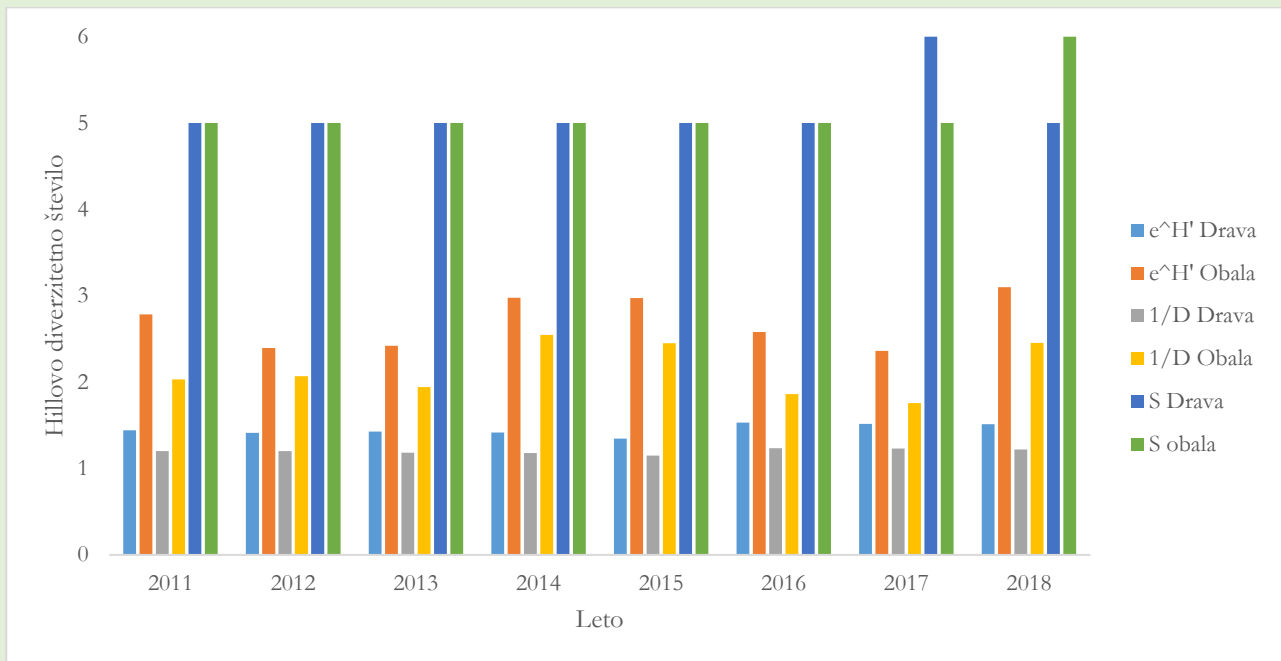


PRIMER 4.5

Zimsko štetje vodnih ptic – Hillova števila

Hill je uvedel diverzitetna števila, s katerimi pestrost vzorca izrazimo s številom vrst. S tem pristopom je želel odpraviti zmedo v interpretaciji diverzitetnih indeksov. Diverzitetna števila izhajajo iz diverzitetnih indeksov in predstavljajo enačbe 0, 1 in 2 reda. Višji, kod je red enačbe, manjši je vpliv redkih vrst na izračun. Redke vrste najbolj vplivajo na 0 red – S , kjer je vpliv redkih in številčnih vrst izenačen.

Primeri izračunov za vse tri oblike prave diverzitete za Dravo in Obalo vzdolž časovne vrste so predstavljeni na sliki 4.10



Slika 4.10 Hillova diverzitetna števila za prezimujoče rase rodu *Anas* na dveh lokalitetah po letih. $e^{H'}$ – Shannonova diverzitetna, $1/D$ – Simpsonova recipročna diverzitetna, S – vrstno bogstvo.



4.3 Vrstna poravnost

Vrstna poravnost (ang. Species Evenness, sin. Equitability) vsebuje dve komponenti variabilnosti, to sta število vrst in poravnost njihovih abundanc. Govorimo o **indeksih poravnosti** (ang. Evenness Indices) E_i . Njihova zaloga vrednosti je od 0 do 1. V primeru, ko imajo vse vrste v vzorcu enako abundanco, tj. enako število osebkov, ima indeks poravnosti največjo vrednost 1, in obratno, če je v vzorcu samo ena vrsta, je vrednost indeksa 0. Vrednost indeksa se približuje ničli, kakor abundance osebkov posameznih vrst divergirajo od enakosti in se približuje vrednosti 1, ko se abundance osebkov posameznih vrst izenačujejo. Poznamo več indeksov poravnosti:

$$E_1 = \frac{H'}{\ln(S)} = \frac{\ln(N_1)}{\ln(N_0)},$$

$$E_2 = \frac{e^{H'}}{S} = \frac{N_1}{N_0},$$

$$E_3 = \frac{e^{H'} - 1}{S - 1} = \frac{N_1 - 1}{N_0 - 1},$$

$$E_4 = \frac{\frac{1}{D}}{e^{H'}} = \frac{N_2}{N_1},$$

$$E_5 = \frac{\frac{1}{D} - 1}{e^{H'} - 1} = \frac{N_2 - 1}{N_1 - 1}.$$

kjer so: E_{1-5} – indeksi poravnosti (E_1 – Pieloujev indeks poravnosti, E_4 – Hillovo razmerje, E_5 – modificirano Hillovo razmerje), H' – Shannonov indeks, D – Simpsonov indeks, S – število vrst, N_1 , N_2 , in N_3 – Hillova diverzitetna števila, e – matematična konstanta = 2,718.

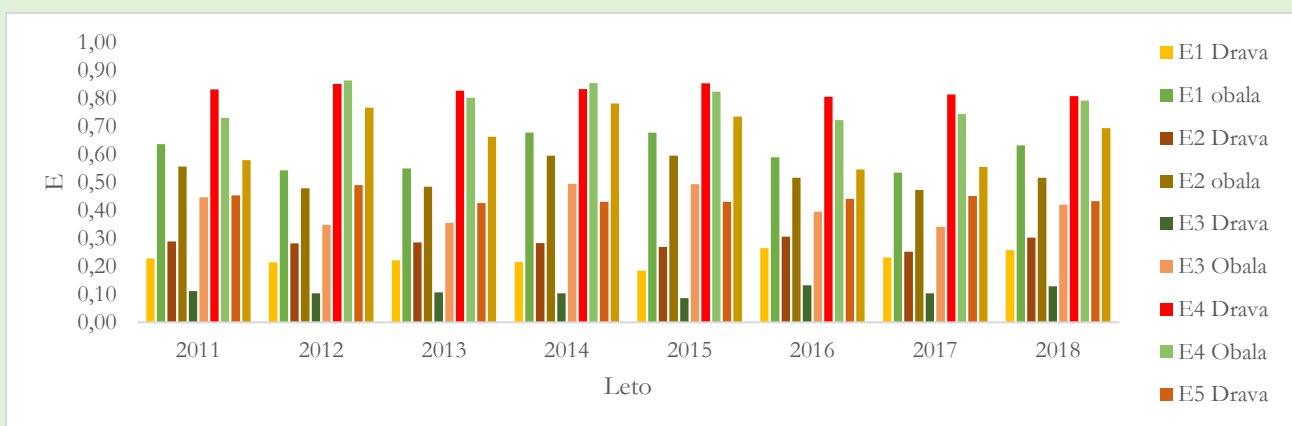


PRIMER 4.6

Zimsko štetje vodnih ptic – vrstna poravnost

Številni avtorji makroekoloških razprav interpretirajo vrstno pestrost kot funkcijo fizikalnih, geografskih, bioloških in časovnih spremenljivk. Vrstno pestrost razlagajo, kot da sestoji iz dveh komponent: iz vrstnega bogastva (število vrst) in poravnosti vrst (relativna vrstna abundanca). S krivuljo relativne abundance (predstavljena v Poglavju 3 Vrstno bogastvo) razvrstimo vrste glede na njihovo številčnost, s tem prikažemo razmerje med številom pogostih in številom redkih vrst. Praviloma je pogostih vrst zelo malo, število vrst z nizko abundanco pa je mnogo. Iz oblike porazdelitvene krivulje sklepamo na medvrstne odnose in kompeticijske pritiske. V analizi vrstne poravnosti združimo obe komponenti, tako vrstno bogastvo, kot njihovo poravnost in večja, kot so razhajanja v poravnosti vrst, nižji je indeks. Indeksi poravnosti imajo dobro lastnost. Njihove vrednosti so v razponu od 0 do 1. Vrednost 1 pomeni popolno poravnost vrst oz. zastopanost vseh vrst z enakim številom, vrednost 0 pa prisotnost samo ene vrste.

Poravnost vrst vzdolž osemletnega spremljanja prezimovanja rac na Dravi in Obali izkazuje stabilno nižjo poravnost vrst na Dravi, kar razložimo s prevlado rase mlakarice. Medletna variabilnost ni izrazita (Slika 4.11).



Slika 4.11 Indeksi poravnosti za prezimujoče rase rodu *Anas* na dveh lokalitetah po letih.



4.4 Primerjave vrstne pestrosti med združbami

V tem poglavju predstavljamo pristop k primerjavam biodiverzitete med vzorci s pomočjo Hillovih diverzitetnih števil oz. s primerjanjem efektivnega števila vrst (sin. prava diverziteteta). Analizo izvajamo s pakotom »iNEXT« Chao in sod. 2014 in Hsieh in sod. 2022.

V prvem koraku primerjamo akumulacijske krivulje pestrosti na surovih podatkih, v nadaljevanju vrstno pestrost redčimo oz. standardiziramo: 1. glede na število osebkov (ang. Sample-size) posamezne vrste (abundance) ali na njihovo incidenco (število prisotnosti v vseh vzorčenjih); 2. na vzorčno pokritost (ang. Coverage-based) in iz njih izračunamo vrednosti prave diverzitetete oz. Hillova števila (za $q = 0$, $q = 1$, $q = 2$).

Na izračun efektivnega števila vrst posameznih vzorcev lahko vplivajo: 1. **velikost vzorčene enote**, vzorčene enote se razlikujejo v velikosti (dolžini transekt, površini, volumnu ...) oz. v količini vložene delo (človek/ura/vzorčena enota); 2. **število prešteti osebki** in 3. **vzorčna pokritost SC** (ang. Sample Coverage), ocena deleža spregledanih vrst in/ali osebki, dejansko delež enojcev. Za odstranitev vpliva posameznega dejavnika izvedemo standardizacijo podatkov, s čimer podatke priredimo na skupno osnovo.

Standardizacijo zaradi različno velikih vzorčenih enot izvedemo tako, da za vsako vzorčeno enoto preračunamo število osebki posameznih vrst na izbrano referenčno vrednost velikosti, po navadi na vrednost najmanjše vzorčene enote (najkrajši transekt, najmanjša površina, najmanjši volumen ali podobno).

Standardizacijo zaradi različnega števila osebki izvedemo tako, da vsaki vzorčeni enoti z redčenjem zmanjšamo število osebki na izbrano referenčno vrednost, po navadi na vrednost vzorca z najmanjšim številom osebki.

Standardizacijo zaradi različne vzorčne pokritosti izvedemo na način, da vsaki vzorčeni enoti izračunamo vrstno pestrost glede na izbrano referenčno vrednost vzorčne pokritosti, po navadi na vrednost vzorca z najnižjo vrednostjo pokritosti.



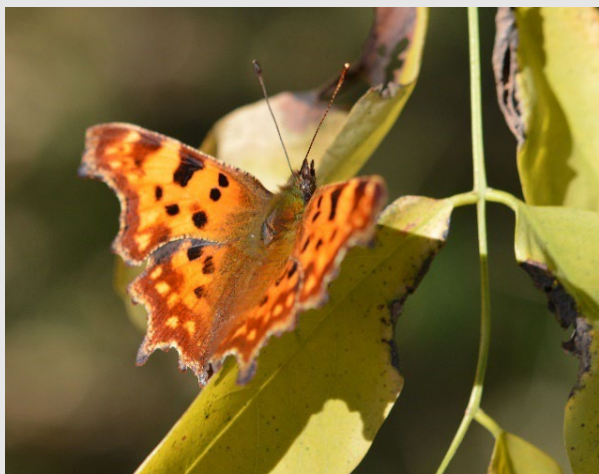
PROJEKT 4.2

Fenologija metuljev

Metulji (Slika 4.12) so pogosto objekt biodiverzitetnih raziskav, v tej predstavitvi so povzeti podatki iz članka Predraga Jakšića in sodelavcev (2017). Najpogostejša metoda za popis metuljev je *linijski transekt* oz. Pollardov transekt – standardizirani protokol za popis metuljev. Metoda je znana pod imenom Pollardova hoja (ang. Pollard Walks), po raziskovalcu metuljev Ernieju Pollardu in jo pogosto uporabijo pri monitoringu metuljev.

Popisovalec na terenu določi pot znane dolžine (npr. 100 m) in širine 5 m, ki jo počasi prehodi (Slika 4.13), praviloma večkrat v sezoni. Vsakič popiše vse metulje in njihovo številčnost pred seboj in na razdalji približno 2,5 m na levo in na desno od začrtane poti. Druga oseba lahko sledi popisovalcu in pomaga pri štetju, determinaciji in/ali fotografiranju metuljev.

Število vrst metuljev in število njihovih osebkov se tekom leta spreminjata. Na diverzitetu metuljev zelo vplivajo geografska širina, nadmorska višina in sezonska variabilnost klime. Metulji so vezani na prisotnost svojih krmnih rastlin, gosenice predvsem na liste, odrasli metulji pa na cvetove.



Slika 4.12 Metulj vrste beli C *Polygonia c-album*.
Foto. Predrag Jakšić.



Slika 4.13 Predrag Jakšić pri vzorčenju žuželk na travniku.
Foto. Franc Janžekovič.





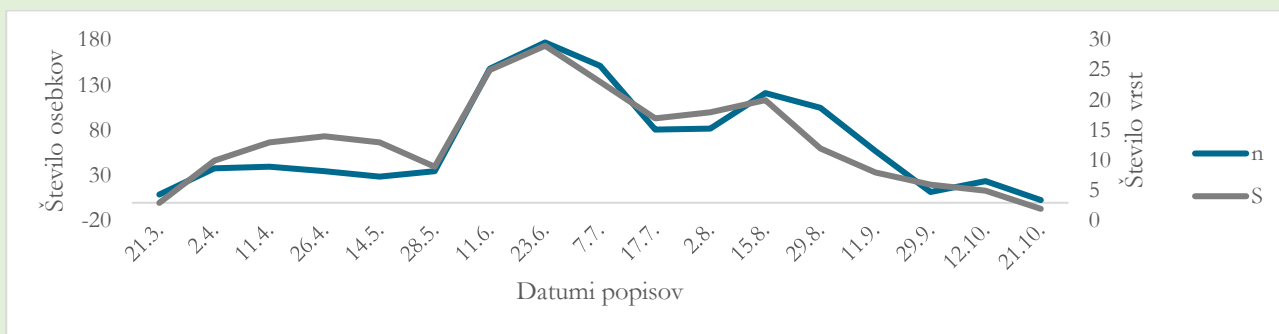
PRIMER 4.7

Fenologija metuljev – podatki

Jakšič in sod. 2017 so predstavili rezultate enoletnega monitoringa metuljev na Avali (Srbija). Popisi so bili izvedeni v treh habitatih: na travniku v zaraščanju – A, v naravnem gozdu – B in na njivi v zaraščanju – C. Monitoring metuljev je obsegal 17 popisov in je bil izveden po metodi Pollardovega transekta. Popisovalec je na terenu počasi prehodil v naprej definirano pot in popisoval vse metulje pred seboj in na razdalji približno 2,5 m na levo in na desno od začrtane poti. Površina popisne ploskve se izračuna kot dolžina poti krat širina popisa (dolžine poti: A = 641 m, B = 752 m, C = 439 m).

Podatke s terenskih popisov so uredili v Excelu. Posamezna vrstica vnosa je vsebovala naslednje elemente: lokaliteta, datum, vrsta metulja, abundanca metulja. Z drugimi besedami, za izbrani habitat in na en datum je za vsako vrsto metulja opravljen zapis v novo vrstico.

Skupno so ugotovili 50 različnih vrst metuljev in prešteli 1.147 osebkov. Podatki o številu vrst in številu osebkov v časovni vrsti so predstavljeni na grafu (Slika 4.14). Aktivnost metuljev je trajala od sredine marca do druge polovice oktobra. Na grafu so izrisani trije vrhovi tako v številu vrst kot v številu osebkov, prvi je v sredini aprila in ga tvorijo lanske živali, ki so bodisi prezimile kot odrasli osebki bodisi so prezimile kot bube in se izlegle v zgodnji pomladi. Drugi, najvišji vrh je dosežen v drugi polovici junija, ko se izleže prva generacija letošnjih metuljev. Druga generacija letošnjih metuljev tvori tretji vrh v sredini avgusta.



Slika 4.14 Število osebkov in število vrst v časovni vrsti; S = število vrst, n = število osebkov.

Za primerjanje vrstne pestrosti metuljev med tremi habitatimi (A, B in C) moramo upoštevati tri vire variabilnosti, ki vplivajo na izračun pestrosti: 1. različno velike površine raziskanih habitatov (različen vložek raziskovalnega dela v urah/habitat) – standardizacija na velikost območja; 2. različno število popisanih osebkov – standardizacija na število osebkov; 3. različen uspeh pri odkrivnosti metuljev – standardizacija na vzorčno pokritost.

Standardizacijo na velikost območja izvedemo tako, da najmanjši ploskvi – C določimo referenčno vrednost ($KV = 1$) in preostalima ploskvama preračunamo koeficient vpliva sorazmerno glede na velikost oz. dolžino popisne ploskve. Sledi preračun abundanc glede na koeficient vpliva ($\text{Standardizirana_abundanca} = \text{Surova_abundanca} * KV$), (Preglednica 4.3).

Preglednica 4.3 Koeficient vpliva (KV) za tri različno velike habitate

| Habitat | Dolžina (m) | Izračun KV | KV |
|--------------------------|-------------|------------|--------|
| A – travnik v zaraščanju | 641 | 439/641 | 0,6848 |
| B – naravni gozd | 752 | 439/752 | 0,5837 |
| C – njiva v zaraščanju | 439 | 439/439 | 1,0000 |



PRIMER 4.8

Fenologija metuljev – lastnosti popisov

Za izračun Hillovih števil ter standardizacijo na število osebkov, standardizacijo na vzorčno pokritost in nadaljnjo statistično vrednotenje v paketu »iNEXT« programskega okolja R morajo biti podatki pripravljene v dveh ločenih matrikah, posebej za abundanco (število osebkov posamezne vrste) in posebej za prezenco (prisotnost posamezne vrste).

Podatki o abundanci (`data.abund`) vsebujejo skupno število osebkov posamezne vrste v posameznem habitatu. Podatki o prezenci (`data.incid`) vsebujejo število prisotnosti posamezne vrste v posameznem habitatu, naslovna vrstica v tej tabeli je število popisov (`plo.no = 17`).

Najprej pregledamo lastnosti vzorcev s funkcijo *DataInfo*, datoteki analiziramo vsako posebej `data.abund` (Preglednica 4.4) in `data.incid` (Preglednica 4.5). Rezultat so naslednje vrednosti:

- za abundanco: *site* – oznaka habitata (A, B, C); *n* – število osebkov; *S.obs* – število vrst; *SC* – vzorčna pokritost (ocena pokritosti, od 0 do 1); vrednosti v stolpcih *f1*, *f2*, *f3*, ..., *f10* – število vrst, ki so na lokaliteti zastopane s samo 1, 2, 3, ..., 10 osebki (enojci, dvojci, trojci itd.).
- za prezenco: *site* – oznaka habitata (A, B, C); *T* – število popisov; *U* – skupno število prisotnosti vseh vrst na vseh popisih; *S.obs* – število vrst; *SC* – vzorčna pokritost (ocena pokritosti, od 0 do 1); vrednosti v stolpcih *Q1*, *Q2*, *Q3*, ... *Q10* – število vrst, ki so na lokaliteti opazovane samo 1, 2, 3, ... 10 krat (edinstvene, dvojne itd.).

Preglednica 4.4 Lastnosti abundanca osebkov za tri habitate A – travnik v zaraščanju, B – naravni gozd in C – njiva v zaraščanju, razlaga oznak je v besedilu

| Habitat | n | S.obs | SC | f1 | f2 | f3 | f4 | f5 | f6 | f7 | f8 | f9 | f10 |
|---------|-----|-------|--------|----|----|----|----|----|----|----|----|----|-----|
| A | 576 | 44 | 0.9844 | 9 | 5 | 2 | 0 | 1 | 4 | 3 | 0 | 0 | 3 |
| B | 213 | 30 | 0.9625 | 8 | 0 | 4 | 5 | 0 | 1 | 2 | 1 | 2 | 1 |
| C | 358 | 37 | 0.9777 | 8 | 2 | 4 | 2 | 4 | 1 | 2 | 3 | 0 | 0 |

Preglednica 4.5 Lastnosti prezenca vrst za tri habitate A – travnik v zaraščanju, B – naravni gozd in C – njiva v zaraščanju, razlaga oznak je v besedilu

| Habitat | T | U | S.obs | SC | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 |
|---------|----|-----|-------|--------|----|----|----|----|----|----|----|----|----|-----|
| A | 17 | 159 | 44 | 0.9227 | 13 | 6 | 7 | 5 | 5 | 2 | 0 | 2 | 1 | 2 |
| B | 17 | 94 | 30 | 0.8787 | 12 | 5 | 2 | 1 | 4 | 2 | 3 | 0 | 1 | 0 |
| C | 17 | 104 | 37 | 0.8892 | 12 | 4 | 11 | 5 | 2 | 0 | 3 | 0 | 0 | 0 |

```
R > package = »iNEXT«
# Lastnosti vzorcev
DataInfo (data.abund, datatype = "abundance")
DataInfo (data.incid, datatype = "incidence")
```



PRIMER 4.9

Fenologija metuljev – Hillova števila

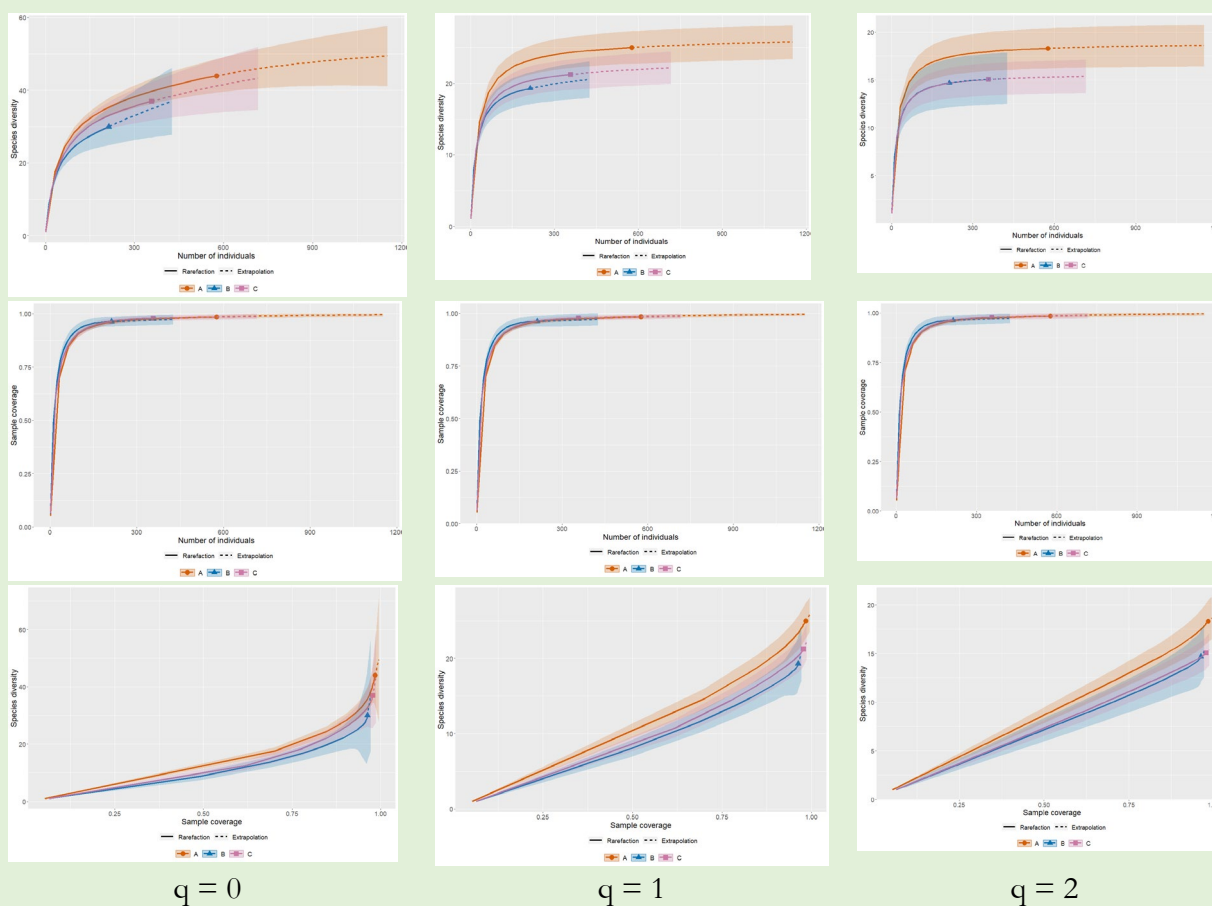
Matriko podatkov z abundancami metuljev za tri habitate smo vključili v izračun Hillovih števil, analizo smo izvedli z ukazom iNEXT. Rezultati za vsa tri Hillova števila ($q = 0$, $q = 1$, $q = 2$) so prikazani grafično na Sliki 4.15 in tabelarično v Preglednici 4.6.

Podatke o abundancah metuljev smo pretvorili v podatke o prezenci, t.j. v PA matriko. Analizo izračuna Hillovih števil smo ponovili na podatkih o prezenci (Preglednica 4.6).

Primerjave med pestrostjo metuljev po habitatih izvajamo znotraj istega Hillovega števila.

Preglednica 4.6 Vrednosti Hillovih diverzitetnih števil: $q = 0$ vrstno bogastvo S , $q = 1$ Shannonova diverzitetna $e^{H'}$ i $q = 2$ Simpsonova recipročna diverzitetna $1/D$, ki podajajo pestrost v enotah števila vrst. Izračun je narejen posebej na podatkih o abundancah in posebej na podatki o prezenci. Za združbe metuljev v treh habitatih A – travnik v zaraščanju, B – naravni gozd in C – njiva v zaraščanju

| | Na abundanci | | | Na prezenci | | |
|---|--------------|-------|-------|-------------|-------|-------|
| | q = 0 | q = 1 | q = 2 | q = 0 | q = 1 | q = 2 |
| A | 44 | 25 | 18 | 44 | 34 | 28 |
| B | 30 | 19 | 15 | 30 | 23 | 19 |
| C | 37 | 21 | 15 | 37 | 31 | 27 |



Slika 4.15 Grafični prikaz odnosov med vrstno pestrostjo (ang. Species Diversity), številom osebkov (ang. Number of Individuals) in vzorčno pokritostjo (ang. Sample Coverage) v Hillovih številih: $q = 0$ vrstno bogastvo S , $q = 1$ Shannonova diverziteteta $e^{H'}$, $q = 2$ Simpsonova recipročna diverziteteta $1/D$, za biodiverzitetu metuljev v treh habitatih A – travnik v zaraščanju, B – naravni gozd in C – njiva v zaraščanju. Zgornja vrstica: Projekcija vrstne pestrosti na število osebkov. Srednja vrstica: Projekcija vzorčne pokritosti na število osebkov. Spodnja vrstica: Projekcija vrstne pestrosti na vzorčno pokritost. Polna krivulja predstavlja dejanske podatke, črtkana črta predstavlja ekstrapolacijo. Barvno polje predstavlja variabilnost podatkov.

```

R > package = »iNEXT«
#Izračun Hillovih števil - prezenca
inc<-iNEXT(me.incid, q=1, datatype = "incidence_freq")
inc #Prikaz izračuna
ggiNEXT(inc, type = 3) #Plot

```





PRIMER 4.10

Fenologija metuljev – primerjave med vzorci

Primerjave vrstne pestrosti med vzorci z različno količino vložene delo ali različno vzorčno pokritostjo izvajamo po metodi redčenja (ang. Rarefaction), ki je opisana v Poglavjih 3.6 in 16.1. Primerjave lahko izvajamo na podatkih o abundancah, kot na podatkih o prezencah. V nadaljevanju sta predstavljeni oba postopka in njuni rezultati.

Izračun Hillovih števil na abundancah in primerjave med pestrostjo metuljev v treh habitatih smo dosegli: 1. s standardizacijo na enako število popisanih osebkov, v tem primeru na 426 osebkov; 2. s standardizacijo na enako vzorčno pokritost $SC = 0,9718$. Standardizirane vrednosti Hillovih števil so podana v Preglednici 4.7.

Izračun Hillovih števil na prezencah in primerjave med pestrostjo metuljev v treh habitatih smo dosegli: 1. s standardizacijo na enako število izvedenih popisov, v tem primeru na 34 popisov; 2. s standardizacijo na enako vzorčno pokritost $SC = 0,9447$. Standardizirane vrednosti Hillovih števil so podana v Preglednici 4.7 posebej za abundance in posebej za prezenco.

Znotraj istega Hillovega števila lahko sedaj izvajamo primerjave med vrstno pestrostjo metuljev iz različnih habitatov. Pri abundancah v vseh šestih primerih izkazuje najvišjo vrstno pestrost združba metuljev v habitatu A – travnik v zaraščanju. Medtem sta združbi v habitatih B in C precej izenačena predvsem v Hillovih številih $q = 1$ in $q = 2$, kjer med vzorcema ni razlik.

Preglednica 4.7 Vrednosti Hillovih števil za tri združbe metuljev v habitatih A – travnik v zaraščanju, B – naravani gozd in C – njiva v zaraščanju. A. Standardizacija abundanc na isto število osebkov in na isto vzorčno pokritost (SC). B. Standardizacija prezenc na isto število vzorčenj in isto vzorčno pokritost

| A | Št. osebkov $n = 426$ | | | Vzorčna pokritost $SC = 0,9718$ | | |
|---|-----------------------|---------|---------|---------------------------------|---------|---------|
| | $q = 0$ | $q = 1$ | $q = 2$ | $q = 0$ | $q = 1$ | $q = 2$ |
| A | 41 | 25 | 18 | 38 | 24 | 18 |
| B | 37 | 21 | 15 | 37 | 21 | 15 |
| C | 38 | 21 | 15 | 35 | 21 | 15 |

| B | Št. vzorčenj $n = 34$ | | | Vzorčna pokritost $SC = 0,9447$ | | |
|---|-----------------------|---------|---------|---------------------------------|---------|---------|
| | $q = 0$ | $q = 1$ | $q = 2$ | $q = 0$ | $q = 1$ | $q = 2$ |
| A | 52 | 37 | 30 | 48 | 35 | 29 |
| B | 38 | 26 | 21 | 37 | 26 | 20 |
| C | 45 | 35 | 30 | 45 | 36 | 30 |

```

R > package = »iNEXT«
#Standardizacija na število osebkov - abundanca
out<-estimateD(me.abund, q=2,datatype = "abundance", base = "size")
out
#Standardizacija na pokritost - abundanca
out1<-estimateD(me.abund, q=2,datatype = "abundance", base = "coverage")
out1
#Standardizacija na pokritost-incidence
out2 <- estimateD (me.incid, datatype = 'incidence_freq', base = 'coverage')
out2
#Standardizacija na število vzorčenj-incidence
out3 <- estimateD (me.incid, datatype = 'incidence_freq', base = 'size')
out3

```



4.5 Relativna vrstna abundanca

Vrste, ki sestavljajo naravno združbo, niso zastopane z enakim številom osebkov. Praviloma je malo vrst z visokim številom osebkov. Prevladujejo vrste z nizkim številom osebkov, večina vrst je redkih. Število osebkov izbrana vrste v združbi je **abundanca vrste**, npr. vrstna abundanca metulja navadni modrin je 102 osebkov v združbi metuljev v izbranem habitatu. V analizi vrstne abundance nas poleg absolutne vrednosti abundance zanima tudi njena relativna vrednost. Z **relativno vrstno abundanco** (ang. Relative Species Abundance) podamo informacijo o tem, kako pogoste ali redke so vrste relativno glede na druge vrste v združbi ali v vzorčeni enoti. Relativno abundanco vrst po navadi podamo za skupino organizmov iz istega (enega) trofičnega nivoja. To pa zato, ker vrste iz istega trofičnega nivoja dejansko ali potencialno konkurirajo za iste dobrine in takšna primerjava med vrstami izkazuje ekološki oz. evolucijski uspeh vrst v združbi. Z relativno abundanco vrst lahko opišemo npr. združbo ptic v gozdu ali metuljev na travniku ipd.

Poglavje o vrstni abundanci je narejeno na virih: Abundance 2021, Gotelli in Graves 1996, McGill in sod. 2007, Oksanen 2022, Oksanen in sod. 2019 in Upton 2020.

Porazdelitev vrstnih abundanc v združbi lahko prikažemo na tri različne načine s Fisherjevo logaritemsko porazdelitvijo, s Prestonova logaritemsko normalno porazdelitvijo in z Whittakerjevo porazdelitvijo rangiranih abundanc.

Z modelom **Fisherjeve logaritemske porazdelitve** abundanc (ang. Fisher's Log-series) pričakovano število vrst v združbi \hat{S} (ali v vzorcu) ter ob znanem številu osebkov n izračunamo z enačbo:

$$\hat{S}_n = \frac{\alpha x^n}{n},$$

kjer je α diverzitetni paramater, x je parameter definiran z α in skupnim številom osebkov N v vzorcu:

$$x = \frac{N}{N-\alpha}.$$

Rezultat Fisherjevega pristopa je vrednost α , ki jo interpretiramo kot cenilko diverzitete. Grafična predstavitev rezultatov je s stolpičnim diagramom s frekvencami vrstnih abundanc (Slika 4.19).

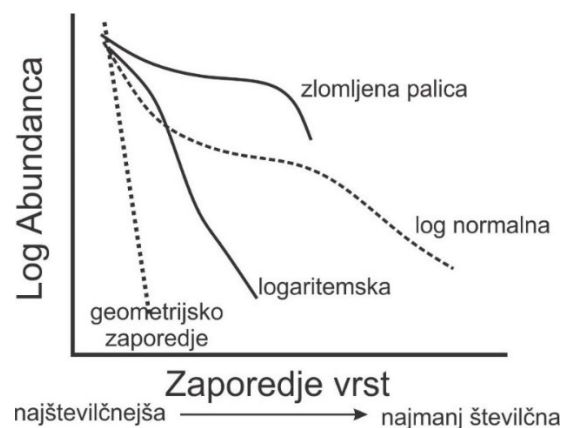
Prestonova logaritemsko normalna porazdelitev (ang. Preston's Log-normal) temelji na porazdelitvi frekvenc posameznih vrstnih abundanc. Za razliko od Fisherjevega pristopa v katerem so abundance razporejene po surovih vrednostih, so pri Prestonoven pristopu abundance razporejene v razrede. Preston je te razrede imenoval oktave. Velikost razredov oz. oktav narašča podobno logaritemski porazdelitvi. Velikost naslednjega razreda se poveča s podvojitvijo prehodnega razreda. Razredi imajo vrednosti: 0, 1, 2-3, 4-7, 8-15, 16-31, 32-63, 64-127, 128-255, 256-511, 512-1.033, 1.034-2.067... števila osebkov. Grafični rezultat praviloma izkazuje najvišje frekvence v začetnih razredih kjer so razporejene vrste z nizko abundanco (Slika 4.20), v nekaterih primerih je rezultat porazdelitve podoben Gaussovi zvonasti porazdelitvi.

Whittakerjeva porazdelitev rangiranih abundanc (ang. Ranked Abundance Distribution) je porazdelitev logaritmiranih vrednosti abundanc urejenih v padajočem zaporedju. Za krivuljo, ki opisuje porazdelitev vrednosti uporabljamo več imen: krivulja porazdelitve rangiranih abundanc, krivulja vrstnih abundanc, krivulja diverzitetne dominace ali Whittakerjev graf. Najpogostejši modeli krivulj so: model zlomljene palice (ang. Brokenstick) v katerem gre za naključno porazdelitev; logaritemski model (ang. Preemption), logaritemsko normalni model (ang. Log-normal) in dva modela v skladu z geometrijskim zaporedjem to sta modela Zipf in Zipf-Mandelbrot (Slika 4.16). Funkcijski zapisi modelov za oceno največjega verjetja so:

| | |
|--|-----------------------------|
| $a_r = \frac{N}{S} \sum_{k=r}^S \frac{1}{k}$ | model zlomljene palice |
| $\hat{a}_r = N\alpha(1 - \alpha)^{r-1}$ | logaritemski model |
| $\hat{a}_r = \exp[\log(\mu) + \log(\sigma) \varphi]$ | logaritemsko normalni model |
| $\hat{a}_r = N\hat{p}_1 r^\gamma$ | Zipf model |
| $\hat{a}_r = Nc(r + \beta)^\gamma$ | Zipf-Mandelbrotov model |

Pomen členov je naslednji: \hat{a}_r je pričakovana abundanca za vrste v rangi r , S je število vrst, N je število osebkov, φ je standardna normalna funkcija, \hat{p}_1 je ocena deleža najbolj abundantne vrste in $\alpha, \mu, \sigma, \gamma, \beta$ in c so ocene parametrov za posamezni model.

Model logaritemske porazdelitve predvideva, da v združbi (ali v vzorcu) prevladujejo redke vrste z nizkimi abundancami. Logaritemska porazdelitev abundanc je lahko tudi posledica nezadostnega vzorčenja. V logaritmsko normalnem modelu prevladujejo vrste z abundancami nekje v sredini med obema skrajnostma. To je lahko tudi posledica preobsežnega vzorčenja. V modelu geometrijskega zaporedja abundance vrst sukcesivno upadajo, v povprečju je vsaka naslednja abundanca za polovico manjša od predhodne. Takšni primeri so prisotni v revnih oz. ekstremnih habitatih. Model zlomljene palice oz. naključna porazdelitev je prisotna v združbah pri katerih je dostopnost do virov neenakomerna oz. je naključno porazdeljena. Slednje pomeni, da so lahko med abundancami sosedov naključno velike razlike.



Slika 4.16 Štirje osnovni modeli krivulj rangiranih abundanc: model zlomljene palice, logaritmsko normalni model, model logaritemske porazdelitve in model geometrijskega zaporedja (prirejeno po Gotelli in Graves 1996 in po McGill in sod. 2007).

Indeks dominance

Relativno vrstno abundanco za izbrano taksonomsko ali ekološko skupino lahko podamo v obliki **indeksa dominance** D , ki ga izračunamo s formulo:

$$D = \frac{n_i}{N} 100,$$

kjer so: D – indeks relativne vrstne dominance, n_i – število osebkov vrste i , N – število vseh osebkov v združbi.

Relativno vrstno abundanco po navadi prikažemo v razredih. V Preglednici 4.8 je predstavljena slovenska terminologija, povzeta po Tarmanu (1992), in angleška, povzeta po Peter (1995) in Abundance (2021).

Preglednica 4.8 Stopnje relativne vrstne abundance po Tarmanu (1992) in Peter (1995) (kratici ACFOR in DAFOR sta začetnici črk stopenj abundance)

| Delež populacije, ki ga zavzema vrsta | Tarman 1992 | A.C.F.O.R. | D.A.F.O.R. | Stopnja relativne abundance |
|---------------------------------------|----------------------|------------|------------|-----------------------------|
| 10 % in več | Evdominantna vrsta | Abundant | Dominant | Prevladujoča |
| 5 – 10 % | Dominantna vrsta | Common | Abundant | Številna |
| 2 – 5 % | Subdominantna vrsta | Frequent | Frequent | Pogosta |
| 1 – 2 % | Recendentna vrsta | Occasional | Occasional | Občasna |
| 1 % ali manj | Subrecendentna vrsta | Rare | Rare | Redka |



PRIMER 4.11

Relativna vrstna abundanca treh združb metuljev

V analizo porazdelitve vrstnih abundanc smo vključili abundance metuljev popisane v treh habitatih: na travnik v zaraščanju – A, v naravnem gozdu – B in na njivi v zaraščanju – C. Podatke o abundancah metuljev smo povzeli po Jašič in sod. 2017. Popisali so 50 vrst metuljev, njihove abundance za posamezni habitat so od 0 (nič) osebkov do 102 osebkov pri vrsti navadni modrin *Polyommatus icarus*. Podatke smo organizirali v matriko dimenzije 3 vrstice x 50 stolpcev, trije habitat in 50 vrst, vrednosti v poljih so bile abundance vrst. V primerih, kjer posamezne vrste niso našli smo jo zapisali z vrednostjo 0.

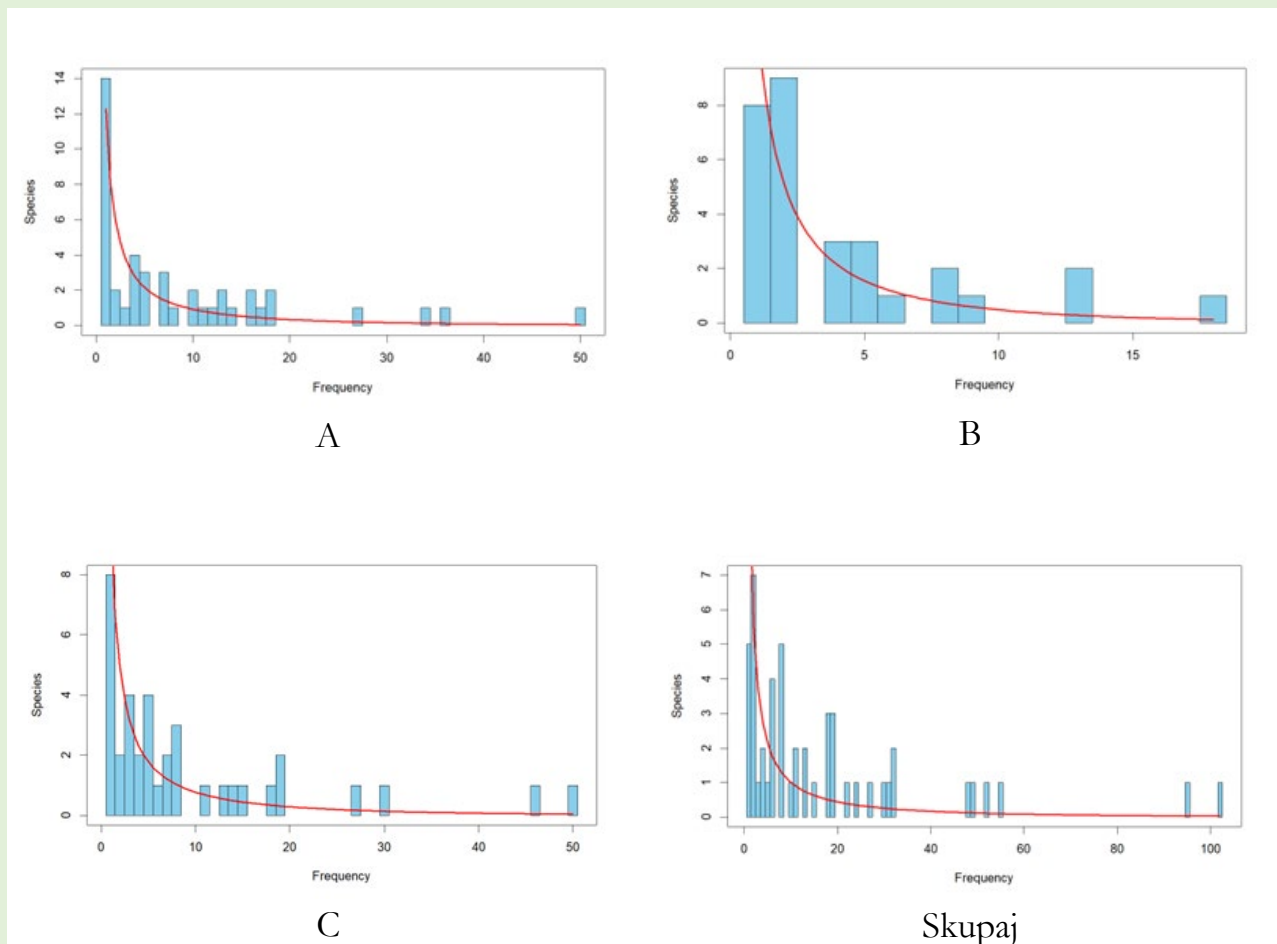
Izvedli smo analize za vse tri pristope Fisherjev, Prestonov in Whittakerjev. Abundance smo obravnavali ločeno po habitatih in potem še združeno za celoten nabor vrst.

Raziskovalna vprašanje je bilo, kakšno obliko porazdelitve vrstne abundance izkazujejo združbe metuljev v posameznem habitatu, ter ali se porazdelitve vrstnih abundanc metuljev iz različnih habitatov razlikujejo med seboj.

Fisherjev pristop

Za posamezni habitatni tip je vzorec porazdelitve vrst glede na njihovo abundanco predstavljen na Sliki 4.19. S stolpci so predstavljene dejanske vrednosti, s krivuljo je predstavljena teoretična logaritemska porazdelitev frekvence vrst glede na njihovo abundanco. Bolj kot se dejanska porazdelitev približuje teoretični porazdelitvi, višja je vrednost Fisherjeve α (alfa). Najvišjo vrednost alfa $\alpha = 12,67$ je dosegla združba metuljev v habitatu A, kjer na grafični predstavitvi vidimo najvišjo skladnost med dejansko in teoretično porazdelitvijo (Preglednica 4.9 in Slika 4.17). Najnižjo vrednost alfa $\alpha = 10,37$ izkazuje združba v habitatu C.

V združeni obravnavi abundanc metuljev iz vseh habitatov Slika 4.17 vidimo, da prevladujejo vrste z nizko abundanco 37 vrst ima abundanco manjšo od 20 osebkov in samo pri 10 vrstah abundanca presega 30 osebkov. Fisherjeva alfa $\alpha = 11,48$ na združenem vzorcu je dosegla srednjo vrednost.



Slika 4.17 Fisherjev pristop prikaza porazdelitve vrstne abundance metuljev v treh habitatih A – travnik v zaraščanju, B – naravni gozd, C – njiva v zaraščanju in skupno. Na X osi so abundance vrst in na Y osi je število vrst. Velikost stolpca podaja število vrst pri izbrani abundanci. Z rdečo krivuljo je vrisana logaritemska porazdelitev.

Preglednica 4.9 Vrednosti Fisherjeve alfa α , ki je cenilka diverzitete in število vrst metuljev za posamezni habitat ter skupna vrednost združeno za vse tri habitate

| Habitat | Fisherjeva alfa α | Št. vrst |
|--------------------------|--------------------------|----------|
| A – travnik v zaraščanju | 12,67 | 44 |
| B – naravni gozd | 12,34 | 30 |
| C – njiva v zaraščanju | 10,36 | 37 |
| Skupaj | 11,48 | 50 |

```

R > package = »vegan«
##Fisherjev pristop; data = metulji 3 x 50
fisA<-fisherfit(metulji[1,]) #A-travnik v zaraščanju
fisA
plot(fisA)
fisB<-fisherfit(metulji[2,]) #B-naravni gozd
fisB
plot(fisB)
fisC<-fisherfit(metulji[3,]) #C-njiva v zaraščanju
fisC
plot(fisC)
fito<-fisherfit(colSums[metulji]) #Skupaj
fito
plot(fito)

```

Prestonov pristop

Prestonov pristop analize vrstne abundance temelji na oktavah oz. oblikovanju razredov, ki so podobni logaritemsko normalni porazdelitvi. Rezultat analize so oktave z opazovanimi in pričakovanimi frekvencami metuljev (Slika 4.18). Za vse tri habitate so opazovane in predvidene frekvence vrstne abundance metuljev podane v preglednicah 4.10, 4.11 in 4.12 in za združene abundance (Preglednica 4.13). Za vsako posamezno združbo metuljev smo s Hi-kvadrat testom preverjali razliko med obema frekvencama. Pri nobeni združbi metuljev razlike med opazovanimi in pričakovanimi frekvencami niso statistično značilne, vrednost p je v vseh primerih večja od 0,5 ($p > 0,5$) (Preglednica 4.14). To pomeni, da se v vseh treh primerih opazovane frekvence skladajo z logaritemsko normalno porazdelitvijo.

Preglednica 4.10 Opazovane in pričakovane frekvence abundanc metuljev po oktavah za habitat A – travnik v zaraščanju

| Oktave | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|-------------|------|------|------|------|------|------|------|
| Opazovane | 7,0 | 8,0 | 4,0 | 8,5 | 8,5 | 5,0 | 3,0 |
| Pričakovane | 6,35 | 7,32 | 7,69 | 7,37 | 6,44 | 5,13 | 3,72 |

Preglednica 4.11 Opazovane in pričakovane frekvence abundanc metuljev po oktavah za habitat B – naravni gozd

| Oktave | 0 | 1 | 2 | 3 | 4 | 5 |
|-------------|------|------|------|------|------|------|
| Opazovane | 4,0 | 8,5 | 6,0 | 6,5 | 4,0 | 1,0 |
| Pričakovane | 4,40 | 6,99 | 7,80 | 6,12 | 3,38 | 1,31 |

Preglednica 4.12 Opazovane in pričakovane frekvence abundanc metuljev po oktavah za habitat C – njiva v zaraščanju

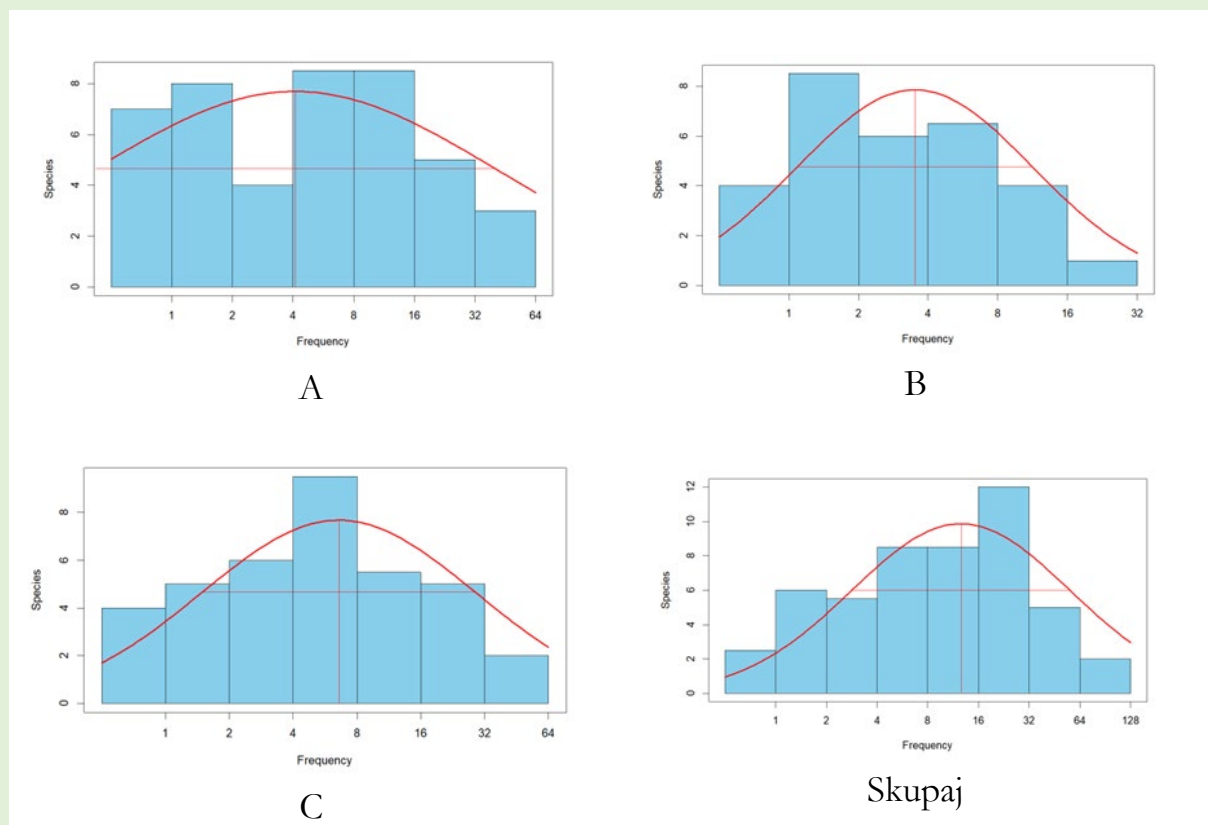
| Oktave | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|-------------|------|------|------|------|------|------|------|
| Opazovane | 4,0 | 5,0 | 6,0 | 9,5 | 5,5 | 5,0 | 2,0 |
| Pričakovane | 3,43 | 5,57 | 7,26 | 7,61 | 6,41 | 4,34 | 2,36 |

Preglednica 4.13 Opazovane in pričakovane frekvence abundanc metuljev po oktavah za vse tri habitate skupaj

| Oktave | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|-------------|------|------|------|------|------|------|------|------|
| Opazovane | 2,5 | 6,0 | 5,5 | 8,5 | 8,5 | 12,0 | 5,0 | 2,0 |
| Pričakovane | 2,35 | 4,62 | 7,35 | 9,42 | 9,74 | 8,11 | 5,45 | 2,95 |

Preglednica 4.14 Rezultati Hi-kvadrat testa preverjanja razlike med frekvencami abundanc metuljev po oktavah med opazovanimi in pričakovanimi frekvencami

| Habitat | Hi-kvadrat | d.f. | p |
|--------------------------|------------|------|--------|
| A – travnik v zaraščanju | 35,0 | 30 | 0,2426 |
| B – naravni gozd | 24,0 | 20 | 0,2424 |
| C – njiva v zaraščanju | 35,0 | 30 | 0,2426 |
| Skupaj | 48,0 | 42 | 0,2426 |



Slika 4.18 Stolpčni diagrami s frekvencami vrstne abundance za združbe metuljev v treh habitatih A – travnik v zaraščanju, B – naravni gozd, C – njiva v zaraščanju in Skupno. Z rdečo krivuljo je predstavljena teoretična logaritemsko normalna porazdelitev frekvenc – pričakovane frekvence.

```

R > package = »vegan«
#Prestonov pristop
pre.octA<-prestonfit(metulji[1,]) #A-travnik v zaraščanju
pre.octA #Frekvence
plot(pre.octA) #Graf
chisq.test(pre.octA$fitted, pre.octA$freq) # Hi kvadrat test
pre.octB<-prestonfit(metulji[2,]) #B-naravni gozd
pre.octB #Frekvence
plot(pre.octB) #Graf
chisq.test(pre.octB$fitted, pre.octB$freq) # Hi kvadrat test
pre.octC<-prestonfit(metulji[3,]) #C-njiva v zaraščanju
pre.octC #Frekvence
plot(pre.octC) #Graf
chisq.test(pre.octC$fitted, pre.octC$freq) # Hi kvadrat test
pre.octSku<-prestonfit(colSums(metulji)) #Skupaj
pre.octSku #Frekvence
plot(pre.octSku) #Graf
chisq.test(pre.octSku$fitted, pre.octSku$freq) #Hi kvadrat test

```

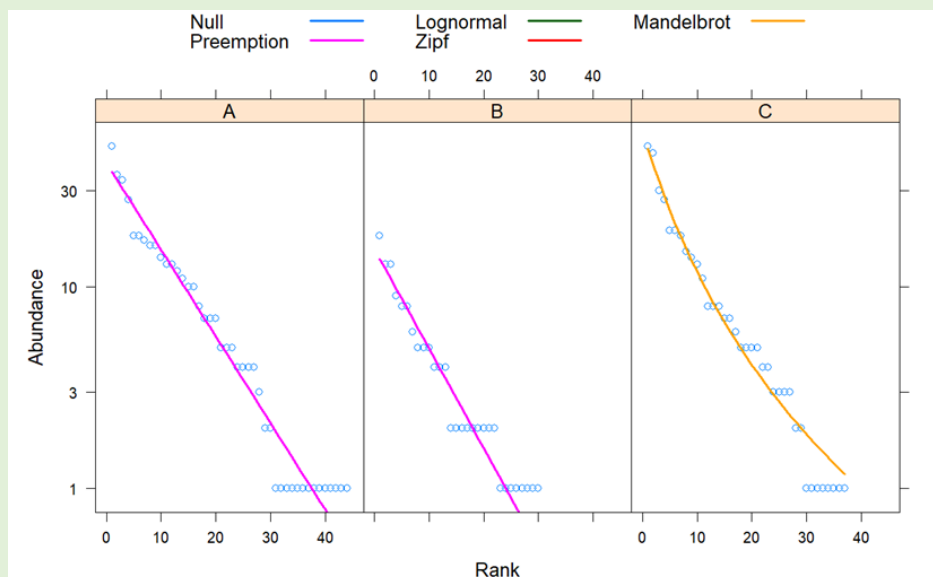
Whittakerjev pristop

Whittakerjev pristop temelji na porazdelitvi rangiranih vrstnih abundanc. Z drugimi besedami, vrste razporedimo v zaporedje s padajočimi abundancami in jih nanašamo na X os, na Y os nanašamo logaritmirane vrednosti abundanc (Slika 4.16).

Grafi rangiranih abundanc za vse tri združbe metuljev iz habitata A, B in C so predstavljeni na Sliki 4.19. Kateri od petih najpogostejših modelov se najbolj sklada s podatki smo preverjali z informacijsko teoretičnim kriterijem AIC (Preglednica 4.15). Združbi metuljev v habitatih A in B se najbolj prilega logaritemski model, porazdelitvi abundanc združbe C se najbolj prilega model geometrijskega zaporedja Zipf-Mandelbrot (Slika 4.19).

Preglednica 4.15 Vrednosti Akaikovega informacijskega kriterija AIC najboljšega prileganja modela k podatkom za posamezno združbo metuljev v habitatih A – travnik v zaraščanju, B – naravni gozd in C – njiva v zaraščanju. Najnižja vrednost AIC je merilo najboljšega prileganja modela k dejanskim podatkom. Najnižja vrednost je v krepki pisavi

| Habitat / Model | A AIC | B AIC | C AIC |
|-----------------------------------|---------------|--------------|---------------|
| Naključni (Null) | 175,09 | 96,01 | 151,83 |
| Logaritemski (Preemption) | 165,48 | 95,74 | 145,05 |
| Logaritemski normalni (Lognormal) | 172,64 | 97,06 | 144,60 |
| Geometrijski (Zipf) | 213,499 | 101,78 | 172,72 |
| Geometrijski (Zipf-Mandelbrot) | 168,16 | 96,89 | 142,61 |



Slika 4.19 Whittakerjev graf rangiranih vrstnih abundanc za tri združbe metuljev v habitatih A – travnik v zaraščanju, B – naravni gozd in C – njiva v zaraščanju. Barvna krivulja opisuje najboljše prilegajoči se model porazdelitve. Za združbi v habitatih A in B ga najbolje opisuje logaritemski model (Preemption), združbo v habitatu C najbolje opisuje geometrijski model Zipf-Mandelbrot (Mandelbrot). Na grafu smo pozorni še na višino abundanco (višina krivulje ob Y osi) in število vrst (konec krivulje ob X osi).

```
R > package = »vegan«
#Whittakerjev pristop
rA<-radfit(metulji[1,]) #A-travniki v zaraščanju
rA
rB<-radfit(metulji[2,]) #B-naravni gozd
rB
rC<-radfit(metulji[3,]) #C-njiva v zaraščanju
rC
rS <- radfit(metulji) #Skupaj
plot(rS)
```



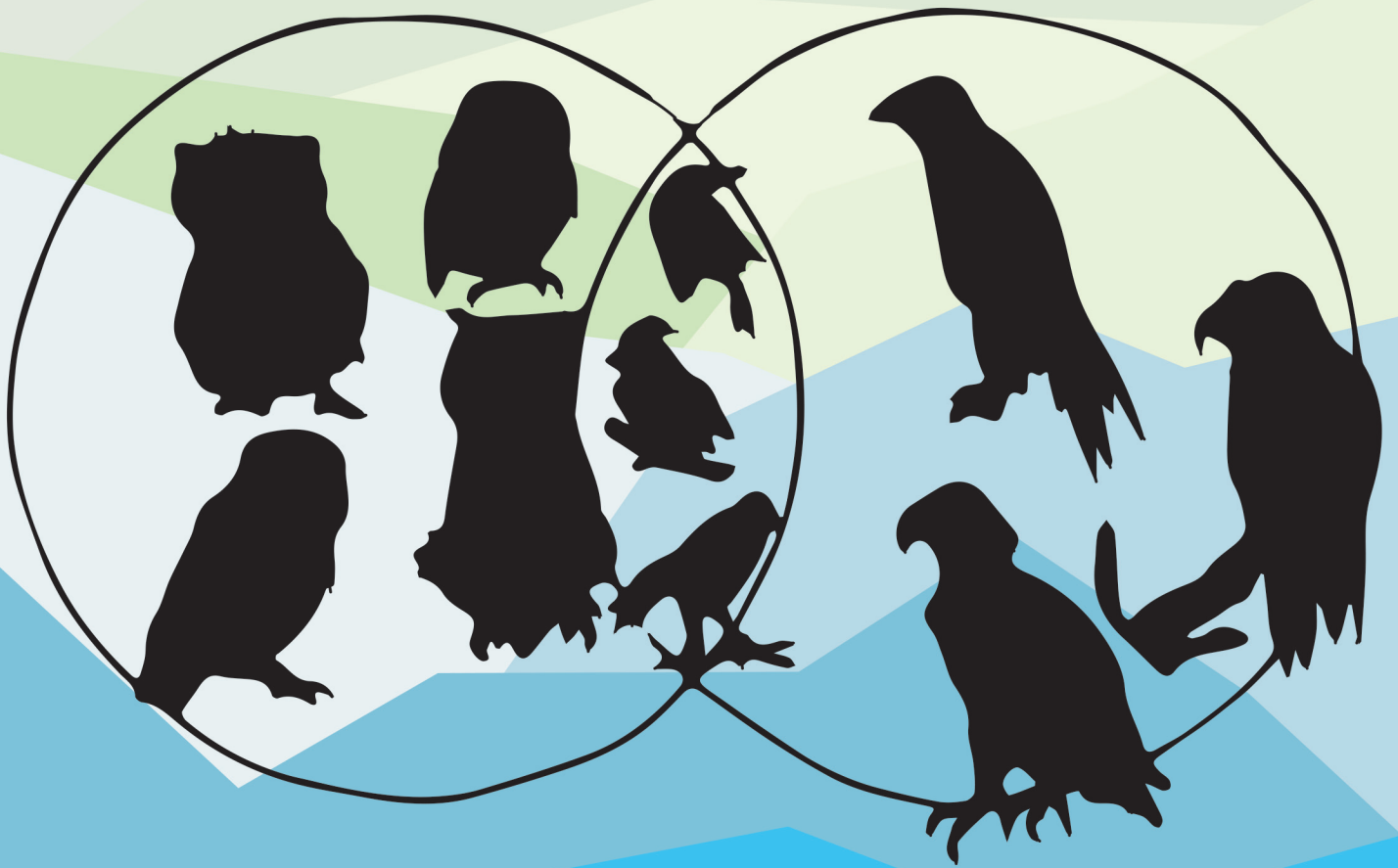
Zaključki

- Vrstna pestrost je razmerje med vrstnim bogastvom in abundanco vrst.
- V izračun cenilk vrstne pestrosti sta vgrajeni komponenti: vrstno bogastvo (število vrst) in poravnano vrst, to je razmerja med vrstami glede na relativno število osebkov posamezne vrste.
- Primerjave med vrstnimi pestrostmi izvajamo glede na: območja, časovna obdobja, časovno zaporedje ali vzdolž gradientov (fizični, kemični, geografski gradient, biotski ali časovni gradient).
- Osnovni cenilki vrstne pestrosti sta Margalefov in Menhinickov indeks, v katerih je upoštevano samo razmerje med številom vrst in skupnim številom osebkov.
- Simpsonov indeks D je ocena verjetnosti, da dva naključno odvzeta osebkata pripadata isti vrsti. Z naraščanjem njegove vrednosti se pestrost niža, obratna vrednost Simpsonovega indeksa $1-D$ ali Gini-Simpsonov indeks, z njegovim naraščanjem opisujemo večanje pestrosti. Tretja oblika je recipročni Simpsonov indeks $1/D$, kjer je pestrost podana v številu vrst.
- Shannonov indeks H' je ocena verjetnosti, da naključno odvzeti osebek pripada določeni vrsti. Vrednost indeksa narašča s številom vrst in z enakomerno zastopanostjo v številu osebkov posamezne vrste. H' ima vrednost 0, ko je v vzorcu samo ena vrsta in ima maksimalno vrednost, kadar imajo vse vrste enako število osebkov.
- Hillova diverzitetna števila H_a so nadgradnja diverzitetnih indeksov, kjer diverzitetno izrazimo v enotah števila vrst.
- Posebna oblika izpeljave diverzitetnih števil je izračun vrstne poravnano E_a . E_a ima nabor vrednosti od 0 do 1, vrednost 1 pomeni, da so vse vrste zastopane z enakim številom osebkov.
- Pri primerjanju vrstne pestrosti med združbami moramo zagotoviti primerljivi vložek raziskovalnega dela. To dosežemo s standardizacijo surovih podatkov: 1. na enak vložek dela, 2. na enako število osebkov in 3. na enako vzorčno pokritost.
- V naravnih združbah je malo vrst z visokim številom osebkov, večina vrst je redkih. Relativno abundanco vrst podamo za skupino organizmov iz istega trofičnega nivoja, takšna primerjava izkazuje ekološki oz. evlucijski uspeh vrst. Porazdelitev vrstnih abundanc v združbi prikažemo s Fisherjevo logaritemsko porazdelitvijo, s Prestonova logaritemsko normalno porazdelitvijo in z Whittakerjevo porazdelitvijo rangiranih abundanc.
- Analitska orodja za vrednotenje biodiverzitet so razvita v okolju R, prikazali smo uporabo paketov »vegan« in »iNEXT«.

Literatura

- Abundance. 2021. Wikipedia. Pridobljeno dne 18. 2. 2021, iz [https://en.wikipedia.org/wiki/Abundance_\(ecology\)](https://en.wikipedia.org/wiki/Abundance_(ecology)).
- Božič L. 2010. Rezultati januarskega štetja vodnih ptic leta 2011 v Sloveniji. *Acrocephalus*, 31: 148/149, 378–381.
- Božič L. 2012. Rezultati januarskega štetja vodnih ptic leta 2012 v Sloveniji. *Acrocephalus*, 32: 152, 207–210.
- Božič L. 2013. Rezultati januarskega štetja vodnih ptic leta 2013 v Sloveniji. *Acrocephalus*, 34: 156/157, 93–103.
- Božič L. 2014. Rezultati januarskega štetja vodnih ptic leta 2014 v Sloveniji. *Acrocephalus*, 35: 160/161, 73–83.
- Božič L. 2015. Rezultati januarskega štetja vodnih ptic leta 2015 v Sloveniji. *Acrocephalus*, 36: 164/165, 57–67.
- Božič L. 2016. Rezultati januarskega štetja vodnih ptic leta 2016 v Sloveniji. *Acrocephalus*, 38: 170/171, 209–219.
- Božič L. 2017. Rezultati januarskega štetja vodnih ptic leta 2017 v Sloveniji. *Acrocephalus*, 38: 174/175, 203–215.
- Božič L. 2018. Rezultati januarskega štetja vodnih ptic leta 2018 v Sloveniji. *Acrocephalus*, 39: 178/179, 85–126.
- Chao A., Gotelli J.N., Hsieh T.C., Sander E.L., Ma K.H., Colwell K.R., Ellison M.A. 2014. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs*, 84: 45–67.
- Clark K.R., Gorley R.N., Somerfield P.J., Warwick R.M. 2014. *Change in Marine Communities: An Approach to Statistical Analysis and Interpretation*. 3rd edition. PRIMER-E. Plymouth.
- Gotelli J.N., Graves R.G. 1996. *Null models in ecology*. Smithsonian Institution Press. Washington and London.
- Hsieh T.C., Ma K.H., Chao A. 2022. iNEXT: iNterpolation and EXTRapolation for species diversity. R package version 3.0.0 URL: <http://chao.stat.nthu.edu.tw/wordpress/software-download/>.
- Jakšič P., Klenovšek T., Janžekovič F. 2017. Monitoring butterfly biodiversity on prime butterfly area Avala Mt. (Serbia) by the transect method (Pollard walks) in the year 2017. *University Thought, Publication in Natural Sciences*, 7(2), 28–35.
- Legendre P., Legendre L. 2012. *Numerical Ecology*. Elsevier. Amsterdam.
- McGill B., Etienne R., Gray J.S., Alonso D., Anderson M., Habtamu Benecha K., Dornelas M., Enquist B., Green J., He F., Hurlbert H.A., Magurran A., Marquet P., Maurer B.A., Ostling A., Soykan C., Uglund K.I., White E. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, 10: 995–1015.
- Oksanen J. 2022. Vegan: ecological diversity. Processed with vegan 2.6-4 in R version 4.2.1 (2022-06-23) on October 11, 2022.
- Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., Wagner H. 2019. *vegan: Community Ecology Package*. R package version 2.5-6. <https://CRAN.R-project.org/package=vegan>.
- Peter M. 1995. *Methods of Environmental Assessment*. University College London Press. London.
- Tarman K. 1992. *Osnove ekologije in ekologija živali*. DZS. Ljubljana.
- Tome D. 2006. *Ekologija : organizmi v prostoru in času*. Tehniška založba Slovenije. Ljubljana.
- Upton G. 2020. *Measuring Abundance. Methods for the Estimation of Population Size and Species Richness*. Pelagic Publishing. Exeter.
- Zelený D. 2020. Analysis of community ecology data in R. Pridobljeno dne 21. 8. 2019, iz <https://www.davidzeleny.net/anadat-r/doku.php/en:start>.

5. poglavje



FUNKCIONALNA PESTROST

Povzetek

5. poglavje: Funkcionalna pestrost

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

Funkcionalna pestrost je pestrost morfoloških, fizioloških, vedenjskih in/ali fenoloških lastnosti na nivoju osebkov, populacije, vrste ali življenjske združbe – objekt. Posamezna lastnost je funkcionalni znak, ki je merljiva značilnost osebkov, s katero se bodisi odziva na biotske in abiotske dražljaje bodisi vpliva na ekosistemske funkcije ali oboje.

Funkcionalno pestrost prepoznavamo po treh dimenzijah: funkcionalno bogastvo, funkcionalna raznolikost in funkcionalna poravnost. Funkcionalno bogastvo je merilo celokupnega obsega funkcionalne pestrosti. Prikažemo jo z velikostjo poligona znakovnega prostora in z istoimensko cenilko. Funkcionalna raznolikost je merilo porazdelitve vrst in njihovih abundanc vzdolž osi funkcionalnih znakov. Izrazimo jo kot povprečno razdaljo med objektom in centroidom celotne združbe. Večja razpršenost objektov pomeni večjo funkcionalno raznolikost, navadno jo razlagamo kot posledico specializacije, in s tem razhajanje med vrstami, s čimer se zmanjšuje kompeticija med njimi. Za prikaz funkcionalne pestrosti poznamo več cenilk: funkcionalno razhajanje, funkcionalna razpršenost, Raova kvadratna entropija in funkcionalna specializiranost. Funkcionalna poravnost je komponenta, s katero vrednotimo funkcionalno pestrost, če bi bile vse vrste zastopane v enakem številu.

Cenilke vrstne pestrosti računamo bodisi na prezenci bodisi na abundanci vrst. Na večjem, makroekološkem merilu lahko prikažemo odnos funkcionalne pestrosti vzdolž gradienta, npr. vzdolž geografske širine, nadmorske višine ali klimatskega gradienta in podobno.

Lastnosti funkcionalnih znakov so lahko kvantitativnega ali kvalitativnega tipa. Kvalitativni tip podatkov pretvorimo v kvantitativnega s pomočjo analize glavnih koordinat.

Analizo samo enega funkcionalnega znaka izvedemo na surovih vrednostih z metodami opisne statistike, vpliv abundance pa s cenilko tehtano povprečje združbe.

Multivariatno obravnavo funkcionalne pestrosti izvajamo bodisi na razdaljah bodisi na večdimenzionalnem hipervolumnu. Cenilke, temelječe na razdaljah, podajajo razporeditev vrst okrog povprečja. Večja variabilnost pomeni večjo funkcionalno pestrost združbe. Cenilke, ki temeljijo na analizi poligona, funkcionalno pestrost združbe vrednotijo na podlagi velikosti n -dimenzionalnega poligona in razporeditev vrst v poligonu. Pomemben dodaten dejavnik za oba koncepta je abundanca vrst, ki lahko zelo vpliva na vrednost cenilke.

Funkcionalna pestrost (ang. Functional Diversity) je pestrost morfoloških, fizioloških, vedenjskih in/ali fenoloških lastnosti na nivoju osebk, populacije, vrste ali življenjske združbe. V nadaljevanju je v rabi termin objekt. Lastnosti so spremenljivke, s katerimi opišemo ekološke, morfološke, biokemijske, molekularne, vedenjske in/ali fenološke značilnosti objekta, s katerimi je v interakciji z živim in neživim okoljem. Lastnosti dajejo sposobnost osebk, da se odziva na dražljaje oz. da deluje na okolje in druge organizme, vplivajo na njegovo preživetje, rast, razvoj in razmnoževanje. Posamezna lastnost organizma se imenuje **funkcionalni znak** (ang. Functional Trait), ki je merljiva značilnost osebk. Lastnost lahko prepoznamo na kateremkoli organizacijskem nivoju organizma, s katero se bodisi odziva na biotske in abiotske dražljaje bodisi vpliva na ekosistemske funkcije ali oboje. Funkcionalni znaki obsegajo vse lastnosti, s katerimi je osebek v interakciji z življenjsko združbo, katere član je, oz. z živo in neživo komponento ekosistema, v katerem živi. Za spremljanje funkcionalnih znakov in njihovo vrednotenje znotraj skupine in primerjave med skupinami definiramo spremenljivke iz petih temeljnih področij: morfologije, razmnoževanja, fiziologije, etologije in fenologije. Za nekatere skupine organizmov obstaja usklajen nabor znakov. Največji napredek je na področju opisovanja rastlin, kjer je ustaljena raba termina »funkcionalne poteze«. Znotraj živalskih skupin obstajajo predlogi funkcionalnih znakov, vendar le-ti še niso standardizirani, ne harmonizirani, niti ni poenotenega merjenja znakov (de Bello in sod. 2021, Kearney in sod. 2021, Morel in sod. 2021).

Uvedba funkcionalnih znakov v raziskave biodiverzitete prispeva k boljšemu razumevanju nekaterih determinističnih procesov v združbah npr. okoljsko filtriranje vrst (glej Poglavje 3.5) in k razumevanju medvrstnih interakcij, ki dejansko delujejo na lastnostih osebk in ne na njegovi taksonomski identiteti. Za razumevanje sestave združbe in procesov v združbi ter odzivanje združbe na spremembe okolja je bolj informativen nabor funkcionalnih znakov v združbi in njihova kombinacija kot sama taksonomska sestava združbe (de Bello in sod. 2021, Gorczyński in sod. 2021).

Terminologija osnovnih pojmov v metriki funkcionalnih znakov je predstavljena v Preglednici 5.1. **Matrika znakov T** (ang. Trait Matrix) je pravokotna matrika, kjer so objekti n (ang. Observations) v vrsticah in znaki z (ang. Traits) v stolpcih (Preglednica 5.2). Objekt raziskave je lahko osebek, populacija, vrsta oz. takson ali združba organizmov. Znaki so lastnosti objekta, ki jih izmerimo na posameznem osebk. V posamezni celici je navedena vrednost x_i funkcionalnega znaka i , ki je numerična ali kategorična meritev znaka. V nadaljnjih analizah vrednosti bodisi analiziramo z metodami opisne statistike bodisi transformiramo v matriko razdalj, bodisi transformiramo npr. v glavne koordinate v primeru kvalitativnih znakov ali v glavne komponente v primeru kvantitativnih znakov. **Matrika objektov O** (ang. Observation Matrix) je pravokotna matrika, kjer je številčnost objektov a v vrstici in objekti n v stolpcih (Preglednica 5.3). Matrika objektov ima samo eno vrstico, v kateri je podana številčnost objektov. **Matrika združb C** (ang. Community

Matrix) je pravokotna matrika, kjer so skupine g (ang. Groups) v vrsticah in objekti s v stolpcih (Preglednica 5.4). Skupine so oblikovane glede na raziskovalno vprašanje in so lahko kladi, združbe, ekosistemi ipd. Objekti so člani skupin, npr. osebek, populacija, vrsta oz. takson. **Znakovni prostor** (ang. Trait Space) je grafična predstavitev matrike znakov (Slika 5.1), predstavljena je v obliki poligona, funkcionalnega dendrograma ipd. **Funkcionalna metrika** (ang. Functional Metric) vključuje različne vidike statistične analize variabilnosti funkcionalnih znakov, lastnosti in variabilnost znakovnega prostora ter izračun cenilk funkcionalne pestrosti (npr. funkcionalno bogastvo, funkcionalna raznolikost, funkcionalna poravnost).

Podatke z meritvami organiziramo v dveh, v primeru analiz med skupinami pa v treh preglednicah. Znaki so v preglednici, imenovani matrika znakov (Preglednica 5.2), številčnost organizmov je v preglednici matrika objektov (Preglednica 5.3) in skupine so v preglednici matrika združb (Preglednica 5.4).

Preglednica 5.1 Slovar osnovnih pojmov v metriki funkcionalne pestrosti

| Pojem | Metrika | Definicija | Primeri uporabe |
|-----------------------|--------------------------------------|--|--|
| Matrika znakov, T | Matrika ($n \times z$) | Tabela z vrednostmi, izmerjene lastnosti znakov | Funkcionalni prostor, ekoprostor, matrika razdalj ipd. |
| Objekti | Vrstice v tabeli T (n) | Objekti raziskave | Takson, osebek, populacija, vrsta |
| Znaki | Stolpci v tabeli T (z) | Merljive lastnosti objekta funkcionalni znaki | Znak, funkcionalni znak, življenjska lastnost, vrednosti glavnih koordinat |
| Številčnost | Vektor (a) | Številčnost objekta | Število osebkov, relativna abundanca, gostota, biomasa, prezenca |
| Matrika objektov, O | Matrika ($a \times n$) | Tabela s številčnostjo objektov | Število osebkov, relativna abundanca, gostota, prezenca |
| Matrika združb, C | Matrika ($g \times n$) | Tabela s številčnostjo objektov glede na skupino | Število osebkov, relativna abundanca, gostota, biomasa, prezenca |
| Skupine | Vrstice v tabeli C (g) | Skupina objektov oblikovana glede na namen raziskave | Združbe, ekosistemi, vrste, kladi |
| Znakovni prostor | Grafična predstavitev matrike znakov | Kakršnakoli transformacija in/ali grafična vizualizacija (navadno 2D) znakovnega prostora | Hipervolumen, konveksni poligon, funkcionalni dendrogram |
| Funkcionalna metrika | Statistično vrednotenje | Statistično vrednotenje lastnosti funkcionalnih znakov in njihove variabilnosti v znakovnem prostoru | Funkcionalna pestrost, funkcionalno bogastvo, funkcionalna poravnost |

Preglednica 5.2 Matrika znakov T , z vrstami s v vrsticah in znaki z v stolpcih (z_1 – intervalni znak, z_2 – nominalni znak, z_3 – binarni znak)

| $s \times z$ | z_1 | z_2 | z_3 | z_4 |
|--------------|-------|-------|-------|-------|
| Sp_1 | 9,3 | D | 1 | p_1 |
| Sp_2 | 8,6 | C | 0 | p_2 |
| ... | ... | ... | ... | ... |
| Sp_n | 9,1 | A | 1 | p_n |

Preglednica 5.3 Matrika objektov O , s številčnostjo a v vrstici in objekti n_i v stolpcih

| $n \times a$ | n_1 | n_2 | ... | n_n |
|--------------|-------|-------|-----|-------|
| a | 5 | 40 | ... | a_n |

Preglednica 5.4 Matrika združb C , s skupinami Skp_i v vrsticah in objekti n_i v stolpcih

| Skupine m | n_1 | n_2 | n_3 |
|-------------|-------|-------|-------|
| Skp_1 | 12 | 0 | 5 |
| Skp_2 | 7 | 9 | 28 |
| Skp_3 | 34 | 5 | 42 |

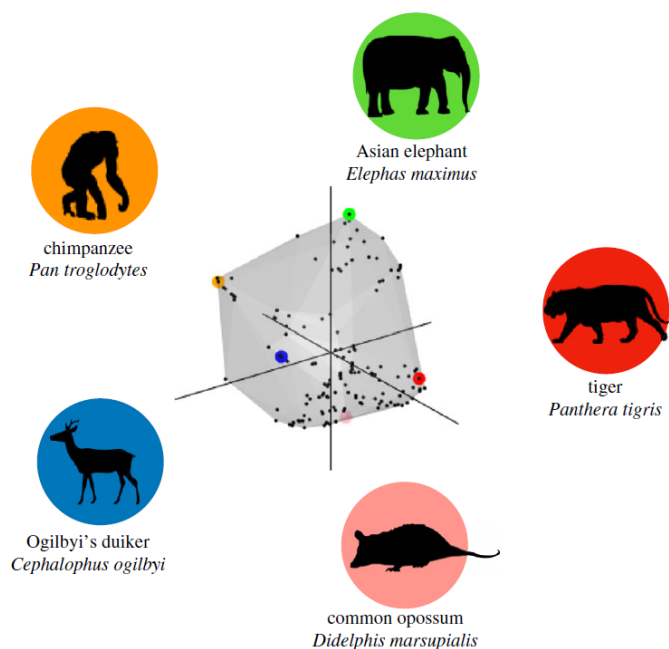
S funkcionalno pestrostjo vrednotimo bogastvo, raznolikost in poravnanost S vrst s T znaki v T -dimenzionalnem prostoru. Cenilke računamo bodisi na prezenci (podatek o prisotnosti izbrane vrste) bodisi jih utežimo z relativno abundanco vrst (podatek o številu osebkov izbrane vrste). Glede na število znakov vključenih v obravnavo, razlikujemo enodimenzionalni pristop, pri obravnavi enega samega funkcionalnega znaka, oz. multidimenzionalni pristop pri istočasni obravnavi dveh ali več znakov. V obeh primerih informacijo o funkcionalni pestrosti objekta strnemo v eno samo cenilko. Posledica redukcije razsežnosti je izguba informacije in obenem nezadostnost pri interpretaciji vrednosti cenilke. Večina raziskovalcev zato opozarja na previdnost pri interpretaciji izračunov. Na večjem, makroekološkem merilu lahko prikažemo odnos funkcionalne pestrosti vzdolž gradienta, npr. vzdolž geografske širine, nadmorske višine ali klimatskega gradienta, in podobno.

Funkcionalno pestrost prepoznavamo po treh dimenzijah: funkcionalno bogastvo, funkcionalna raznolikost in funkcionalna poravnanost, vsaka temelji na drugačni metriki. **Funkcionalno bogastvo** (ang. Functional Richness) je merilo celokupnega obsega funkcionalne pestrosti. Prikažemo ga z velikostjo poligona znakovnega prostora. Velikost poligona je pokazatelj funkcionalnega bogastva združbe. V primerjavah med skupinami nas zanima relativna velikost poligona posamezne skupine in prekrivanje med poligoni. **Funkcionalna pestrost** (ang. Functional Divergence) je merilo porazdelitve vrst in njihovih abundanc vzdolž osi funkcionalnih znakov. Navadno je izražena kot povprečna razdalja med objekti ali kot povprečna razdalja med objektom in centroidom celotne združbe. Večja razpršenost objektov običajno razlagamo kot posledico specializacije, in s tem razhajanj med vrstami, s čimer se zmanjšuje kompeticija med njimi. **Funkcionalna**

poravnano (ang. Functional Evenness or Functional Regularity) vključuje cenilke, ki vrednotijo funkcionalno pestrost, če bi bili vsi objekti zastopani v enakem številu (z enako abundanco).

Funkcionalno pestrost lahko razumemo tudi kot porazdelitev objektov v večrazsežnem prostoru (ang. Multidimensional Space) kjer so osi prostora opredeljene s funkcionalnimi znaki (T -razsežni prostor). Položaj posameznega objekta v prostoru je izkaz njegovih lastnosti. Pri vrstah, kjer je bila izmerjena le prezenca (prisotno/odsotno) znakovni prostor prikažemo dvodimenzionalno s poligonom oz. binarnim hipervolumnom (ang. Binary Hypervolume) oz. najmanjšim konveksnim polihedrom (ang. Convex Hull), ki so dvodimenzionalne predstavitve prostorske razsežnosti. Ob vključitvi abundance vrst njihovo razporeditev v prostoru prikažemo z n -razsežnim hipervolumnom (ang. n -dimensional Hypervolume). Tri-razsežno predstavitev predstavimo z verjetnostnim hipervolumnom (ang. Probabilistic Hypervolume) (Slika 5.1).

Pomen funkcionalne pestrosti je v razumevanju temeljnih fenomenov v makroevolucijskih in makroekoloških procesih, sestavi združb in njihovem odzivu na spreminjajoče se okolje ter funkcionalnih odnosih med biodiverzitetno in ekosistemi (de Bello in sod. 2021, Legras in sod. 2020, Mammola in sod. 2021, Wong in Carmona 2020). Bistven napredek v raziskavah funkcionalne pestrosti se je začel okrog leta 2010.



Slika 5.1 Funkcionalno bogastvo združbe je predstavljeno z znakovnim prostorom (minimalni konveksni poligon v sivi barvi), ki ga oblikuje nabor vrst. Vrste (na poligonu, označene s pikami) so razporejene v tridimenzionalnem prostoru (tri osi). Položaj izbranih vzorčnih vrst je na poligonu označen z barvnimi pikami in tudi silhuete živali so označene z isto barvo. Volumen poligona predstavlja funkcionalno bogastvo. Razporeditev vrst (pik) predstavlja različne usmeritve in stopnje specializacije funkcionalnih znakov. Vzeto iz Gorczynski in sod. 2021.

Preglednica 5.5 Tri multivariatne dimenzije funkcionalne pestrosti s pripadajočimi cenilkami, njihov opis in tip podatkov, s katerimi jih izračunamo

| Dimenzije funkcionalne pestrosti | Cenilka funkcionalne pestrosti | Opis | Podatki s katerimi jih izračunamo |
|----------------------------------|------------------------------------|---|---|
| Funkcionalno bogastvo | Funkcionalno bogastvo FRic | Volumen poligona | Vrednosti znakov ali glavnih koordinat. Na poligonih temelječe analize. |
| Funkcionalna pestrost | Funkcionalno razhajanje FDiv | Odstopanje vrste od centroida, uteženo z relativno abundanco | |
| | Funkcionalna razpršenost FDis | Povprečna razdalja vrst od centroida vseh vrst | Matrika razdalj. Na razdaljah temelječe analize, |
| | Raova kvadratna entropija Q | Vsota razdalj med pari vrst, utežena z relativno abundanco | |
| | Funkcionalna specializiranost FSpe | Evklidove razdalje vrst od centroida, utežene z relativno abundanco | |
| Funkcionalna poravnkost | Funkcionalna poravnkost FEve | Vsota razdalj MST, utežene z relativno abundanco | Vrednosti znakov ali glavnih koordinat. Na poligonih temelječe analize. |

Metodološki pristop analize funkcionalnih znakov je bodisi analiza surovih vrednosti znakov bodisi temelji izračun cenilk na razdaljah med objekti (ang. Dissimilarity-based Indices), bodisi na velikosti in zapolnjenosti poligona znakovnega prostora (ang. Multidimensional Spaces-based Indices). Izračuni cenilk v prikazih posameznih primerov uporabe in interpretacije izračunov smo naredili s pomočjo paketa »FD« v okolju R (Laliberté in sod. 2014).

Na podlagi metodološkega pristopa izračuna cenilk je vsebina razdeljena v tri poglavja:

- na surovih podatkih temelječe analize,
- na razdaljah temelječe analize,
- na poligonih temelječe analize.

5.1 Na surovih podatkih temelječe analize

Analizo samo enega funkcionalnega znaka izvedemo na surovih vrednostih (ang. Raw Data Analyses) z metodami opisne statistike. Vrednosti meritev znakov so glede na tip merjenja lahko zvezne ali diskretne, ordinalne, kategorične in opisne ali binarne. Glede na tip spremenljivke jo analiziramo bodisi s parametričnimi metodami: povprečje, standardni

odklon, minimalna ter maksimalna vrednost; bodisi z neparametričnimi metodami: mediana, frekvence in deleži. Rezultate analiz prikažemo tudi grafično npr. s frekvenčnim diagramom ipd.

Tehtano povprečje združbe CWM (ang. Community Weighted Mean) je enostavna cenilka za opis funkcionalne zgradbe združbe (v literaturi je navedena tudi z imenom »agregirani znak združbe« ali tudi »funkcionalni parameter združbe«). Cenilka vključuje povprečno vrednost znaka x_i za vrsto i , obteženo z njeno relativno abundanco p_i in jo izračunamo s formulo:

$$CWM = \sum_{i=1}^S p_i x_i$$

S je število vrst v združbi, relativna abundanca vrste p_i pa je kvocient med abundanco vrste a_i in skupnim številom osebkom v združbi N :

$$p_i = \frac{a_i}{N}$$

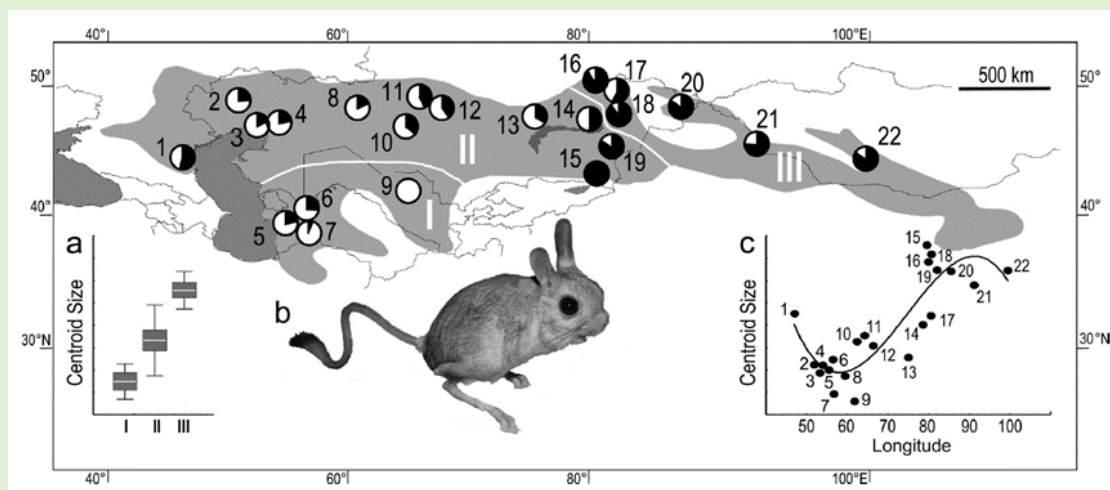
Cenilka CWM je uporabna predvsem v primerjavah med skupinami ali spremljanja združb vzdolž gradienta.



PRIMER 5.1

Funkcionalna znaka telesna velikost in tip hibernacije pri debelorepih skakačih

Kryštufek in sod. (2019) so analizirala ekomorfološko variabilnost spodnje čeljustnice pri debelorepem skakaču vrste *Pygeretmus pumilio* (glej Projekt 10.1). Debelorepi skakači so mali glodavci, ki živijo v puščavskem in pol puščavskem okolju centralne Azije. Iz celotnega areala vrste so pridobili 283 primerkov, ki so jih razvrstili v vzorce 22 populacij (Slika 5.2). Populacije so v nadaljevanju neodvisna spremenljivka oz. kriterij, glede na katerega je izvedena analiza funkcionalne variabilnosti. Del teh podatkov je vključen v ponovno analizo, vendar v novem kontekstu in z namenom prikaza variabilnosti dveh funkcionalnih znakov: 1. centroidna velikost spodnje čeljustnice (proxi za velikost živali) v nadaljevanju velikost; 2. tip hibernacije, razdeljen v tri kategorije I – ne hibernirajo, II – hibernirajo občasno ali s prekinitvami, III – intenzivno hibernirajo.

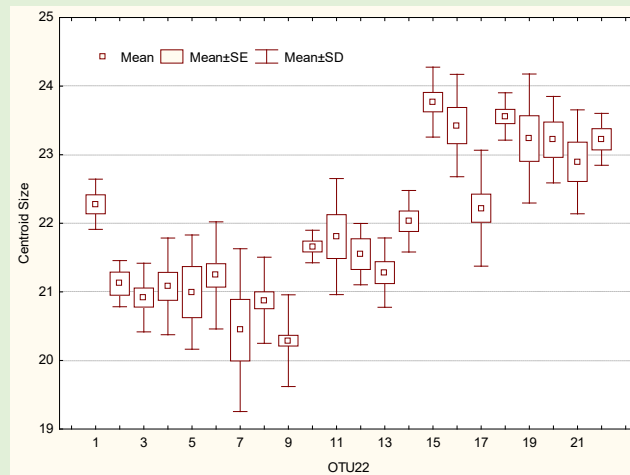


Slika 5.2 Fotografija živali in areal debelorepega skakača *Pygeretmus pumilio* (b). S krožnimi izseki je predstavljena velikost živali, črni krog – velike živali, beli krog – male živali. Z I je označeno območje nizke intenzitete hibernacije, II srednja intenziteta hibernacije in III intenzivno območje hibernacije. Variabilnost velikosti glede na tip hibernacije (a) in variabilnost velikosti vzdolž geografske dolžine (c). Vzeto iz Kryštufek in sod. 2019.

Morfometrični znak velikost čeljustnice je pridobljen z metodami geometrijske morfometrije (podrobneje razloženo v Poglavju 10) in ima lastnost intervalne spremenljivke, kar pomeni, da jo analiziramo z metodami parametrične statistike. Velikost vseh 283 čeljustnic strnemo v izračun povprečja, standardnega odklona ter minimalne in maksimalne vrednosti: $21,5 \pm 1,31$, 18,59 – 25,01. Glede na posamezno populacijo so vrednosti navedene v Preglednici 5.6 in grafično na Sliki 5.3. Telesno najmanjša je populacija 9 in največja populacija 15. Velikost je pomembna biološka lastnost živali, navadno jo vzpostavimo v odnos do drugih ekoloških ali evolucijskih dejavnikov. Pomembna lastnost populacij je tudi variabilnost v telesni velikosti, to razberemo iz standardnega odklona StD. Višji kot je StD, bolj variabilna je telesna velikost populacije. Najbolj uniformna, torej najmanj variabilna je populacija 10 (StdOdk 0,2) in najbolj variabilna je bila populacija 7 (StD 1,2) (Slika 5.3).

Preglednica 5.6 Izbrane vrednosti opisne statistike za funkcionalni znak velikost spodnje čeljustnice in velikost vzorca (n) iz posamezne populacije

| Populacija | Povprečje | StD | Min | Maks | n |
|------------|-----------|-----|------|------|-----|
| 1 | 22,3 | 0,4 | 21,8 | 23,0 | 7 |
| ... | ... | ... | ... | ... | ... |
| 22 | 23,2 | 0,4 | 22,7 | 23,7 | 6 |
| Skupaj | 21,5 | 1,3 | 18,6 | 25,0 | 283 |

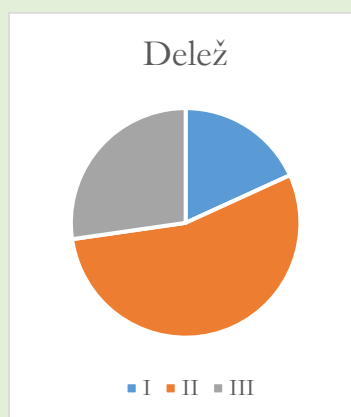


Slika 5.3 Grafični prikaz povprečne vrednosti in variabilnosti velikosti (Centroid Size) spodnje čeljustnice za 22 populacij (OTU22) debelorepega skakača.

Funkcionalni znak stopnja hibernacije je ekofiziološka prilagoditev živali na obdobje z nizko temperaturo in pomanjkanjem hrane. Posamezne populacije so razvile različno stopnjo intenzivnosti hibernacije v obdobju neugodnih razmer 18 % populacij ne hibernira, 27 % jih polno hibernira, dobra polovica oz. 54% so delni hibernatorji (Preglednica 5.7 in Slika 5.4).

Preglednica 5.7 Število populacij glede na tip hibernacije: I – ne hibernirajo, II – občasno hibernirajo, III – redno hibernirajo

| Tip hibernacije | Število populacij | Delež % |
|-----------------|-------------------|---------|
| I | 4 | 18,2 |
| II | 12 | 54,5 |
| III | 6 | 27,3 |



Slika 5.4 Deleži posameznih tipov hibernacije: I – ne hibernirajo, II – občasno hibernirajo, III – redno hibernirajo.



5.2 Na razdaljah temelječe analize

Koncept analiz, ki temeljijo na razdaljah, se naslanja na primerjave med vrstami v podobnosti oz. različnosti funkcionalnih znakov. Stanje oz. vrednost znakov za posamezno vrsto primerjamo glede na povprečje združbe in za vsako vrsto izračunamo njeno razdaljo od povprečja. Cenilke, temelječe na razdaljah, podajajo razporeditev vrst okrog povprečja. Večja variabilnost pomeni večjo funkcionalno pestrost združbe. Pomemben dodaten dejavnik je lahko abundanca vrst in glavne razlike med cenilkama so v metodi upoštevanja vpliva abundance. Ali so spremenljivke kvantitativnega ali kvalitativnega tipa ter ali sočasno obravnavamo oba tipa spremenljivk ali samo en tip, vpliva na izbor matematičnega pristopa računanja razdalje.

Iz metodološkega nabora na razdaljah temelječih cenilk so podrobneje predstavljene tri cenilke FDis, FSpe in Q. Vse tri so iz dimenzije funkcionalne raznolikosti.

5.2.1 Funkcionalna razpršenost FDis (ang. Functional Dispersion) je cenilka, s katero je podana povprečna razdalja \bar{c}_i posameznih vrst od centroida c vseh vrst (Slika 5.5). Vrednosti funkcionalnih znakov v matriki znakov T so lahko kvantitativne (zvezne ali diskretne) ali pa kvalitativne oz. kategorične. V primeru zveznih intervalnih vrednosti razdaljo med centroidom c in položajem vrste j računamo z Evklidovo razdaljo. V primeru kategoričnih vrednosti ali mešanega tipa podatkov vrednosti najprej vključimo v analizo glavnih koordinatnih PCoA (ang. Principal Coordinate Analysis), ki temelji na matriki Gowerjevih razdalj. Rezultat PCoA so glavne koordinate, ki jih kot nove intervalne spremenljivke vključimo v izračun Evklidovih razdalj med centroidom in posamezno vrsto. Vrednosti v matriki združb C so lahko abundance ali prezence. V nadaljevanju je podan postopek izračuna najprej za kvantitativne in potem za kvalitativne znake (Primeri 5.2 in 5.3), oboje ločeno za primer prezence in v primeru abundance vrst. FDis je konceptualno blizu naslednji cenilki FSpe, razlikujeta se v vrednotenju prispevka abundance posamezne vrste.

5.2.1.1 Izračun FDis na kvantitativnih podatkih

Izračun cenilke FDis na kvantitativnih vrednostih funkcionalnih znakov temelji na Evklidovih razdaljah. Centroid za FDis na podatkih o prezenci vrst (PA podatki; 0, 1; prisotno/odsočno) izračunamo po formuli:

$$c = [c_i] = \frac{\sum x_{ij}}{n}.$$

Vektor x_{ij} je položaj vrste j za znak i , n je število vrst (Slika 5.5a). Cenilko FDis za podatke o prezenci izračunamo:

$$FDis = \frac{\sum z_j}{n}.$$

V primeru izračuna cenilke FDis na podatkih z relativno abundanco vrst p_j (Slika 5.5b) centroid izračunamo:

$$c = [c_i] = \frac{\sum p_j x_{ij}}{\sum p_j}.$$

Relativno abundanco p_j vrste j izračunamo kot kvocient med abundanco osebkov vseh vrst N in abundanco osebkov vrste a_j :

$$p_j = \frac{N}{a_j}.$$

Cenilko FDis za podatke o abundanci računamo po formuli:

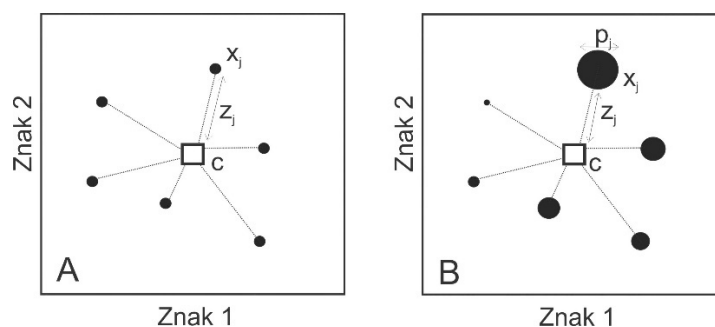
$$FDis = \frac{\sum p_j z_{cj}}{\sum p_j}.$$

Evklidova razdaljo z_j izračunamo s formulo:

$$z_{cj} = \sqrt{\sum_k (x_{kj} - c)^2},$$

kjer so: z_j – Evklidova razdalja med centroidom c in vrsto j ; c – vrednost centroida; x_{kj} – vrednost znaka k za vrsto j .

Cenilka FDis je multivariatna, kar pomeni, da moramo v analizo vključiti vsaj dva funkcionalna znaka, vendar število znakov naj ne presega števila vrst. Prav tako mora obravnavana združba vključevati vsaj dve vrsti. Vrednost cenilke je navzgor neomejena, v primeru obravnave samo ene vrste je njena vrednost 0. Višja vrednost FDis pomeni višjo razpršenost vrst. Vsebinska predstavitev FDis je povzeta po Villeger in sod. (2010).



Slika 5.5 Prikaz vrednosti za izračun funkcionalne razpršenosti FDis. Na dvodimenzionalni projekciji vrst v prostor funkcionalnih znakov 1 in 2 je razporejeno n število vrst, ki jih predstavljajo črni krogi. Velikost kroga je sorazmerne velikosti abundanci vrste. Vektor x_j predstavlja položaj vrste j v dvodimenzionalnem prostoru, vektor c (pravokotnik) je centroid vseh vrst n . z_j je razdalja med centroidom c in položajem vrste j in p_j je relativna abundanca vrste j . A. Vrste imajo enako abundanco, oz. gre za prezenco vrst (PA podatki). B. Vrste imajo različno abundanco (podatki o abundanci), razdalja med c in x_j se zmanjša sorazmerno glede na velikost p_j .

Slika je prirejena po Laliberte in Legendre (2010).



PRIMER 5.2

Funkcionalna razpršenost prezimujočih vodnih ptic v sladkovodnem in morskem habitatu – kvantitativni podatki

V analizo funkcionalne razpršenosti FDis smo vključili zimski popis vodnih ptic, izveden leta 2020 na dveh območjih v Sloveniji na reki Dravi in na območju Obale. Podatke smo povzeli iz Božič (2020). Skupaj je bilo popisanih 55 vrst ptic v skupnem številu 28.241 osebkov, od tega na Dravi 43 vrst in 21.876 osebkov in na Obali 42 vrst in 6.365 osebkov. Vseh 55 vrst smo opisali s 5 funkcionalnimi znaki. Vsi vključeni znaki so morfometrične spremenljivke in vse so kvantitativne. Povzeli smo jih po Storchová in Hořák (2018) (Preglednica 5.8).

Preglednica 5.8 Pregled 5 funkcionalnih znakov, vključenih v izračun cenilke FDis

| Oznaka | Funkcionalni znak, enota |
|--------------|------------------------------|
| LengthU_MEAN | Povprečna dolžina ptiča cm |
| WingU_MEAN | Povprečna dolžina peruti mm |
| TailU_MEAN | Povprečna dolžina repa mm |
| BillU_MEAN | Povprečna dolžina kljuna mm |
| TarsusU_MEAN | Povprečna dolžina tarzusa mm |

Matrika znakov T ima dimenzijo 55 vrst in 5 funkcionalnih znakov in matrika združbe C ima dimenzijo 2 območji in 55 vrst. Pripravili smo dve matriki C , eno z abundancami vrst, drugo s prezenco vrst.

V raziskavi smo preverjali hipotezo, ali je funkcionalna razpršenost združbe vodnih ptic na sladkovodnem habitatu reke Drave podobna tisti v morskem habitatu na Obali. Cenilke FDis v obeh habitatih sta precej podobni, zato rezultate lahko interpretiramo na način, da se znakovna prostora v morfometričnih lastnostih prezimujočih ptic zelo prekrivata. Vrednosti cenilke so nekoliko višje v primeru izračuna glede prezenca vrst. Cenilka FDis za celoten nabor vseh 55 vrst ima vrednost 1,812, vrednosti FDis izračunane na abundanci in na prezenci vrst so podane v tabeli 5.9. Vidimo, da vključitev abundanc zmanjša vrednost cenilke. Na Dravi so večja nesorazmerja med abundancami vrst, zato se vrednost zmanjša bolj kot na Obali.

Preglednica 5.9 Cenilke FDis za dve območji: Drava in Obala na zveznih podatkih, ločeno za izračun na abundancah in na prezenci

| Območje / FDis | Abundanca | Prezenca |
|----------------|-----------|----------|
| Drava | 1,338 | 1,913 |
| Obala | 1,356 | 1,699 |

```
R > package = »FD«
```

```
##FDis - prezimujoče ptice na Dravi in Obali
znaki #Matrika kvantitativnih znakov: 55 vrst in 5 znakov
FDis_skupaj<-dbFD(znaki) #Ukaz za izračun FDis za celoten nabor vrst
abundanca #Matrika abundance: 2 lokaliteti in 55 vrst
prezenca #Matrika prezenca: 2 lokaliteti in 55 vrst
FDis_a<-dbFD(znaki, abundanca) #Ukaz za izračun FDis po lokalitetah-abundanca
FDis_a
FDis_p<-dbFD(znaki, prezenca) #Ukaz za izračun FDis po lokalitetah - prezenca
FDis_p
```



5.2.1.2 Izračun FDis na kvalitativnih podatkih

Izračun cenilke funkcionalna razpršenost FDis na mešanem tipu podatkov ali v primeru manjkajočega podatka, kjer so istočasno vključene kvantitativne in kvalitativne vrednosti ali pa samo kvalitativne vrednosti funkcionalnih znakov, se prične s predhodno transformacijo kvalitativnih znakov v kvantitativne. Transformacijo vrednosti izvedemo z izračunom Gowerjevih razdalj med vsemi pari objektov. Rezultat je simetrična matrika Gowerjevih razdalj med objekti. Matriko razdalj vključimo v analizo glavnih koordinata PCoA. Nove izračunane koordinate, ki imajo kvantitativni značaj, uvedemo v izračun FDis.

Postopek transformacije kvalitativnih vrednosti v kvantitativne:

1. Oblikovanje pravokotne matrike znakov, vrste x znaki.
2. Vključitev pravokotne matrike znakov v izračun Gowerjevih razdalj in oblikovanje simetrične matrike razdalj, vrste x vrste.
3. Vključitev simetrične matrike razdalj v izračun PCoA, oblikovanje pravokotne matrike glavnih koordinat, vrste x koordinate.
4. Vključitev pravokotne matrike glavnih koordinat v izračun FDis (po postopku opisanem v predhodnem poglavju 2.1.1).

Izračun Gowerjevih razdalj in oblikovanje simetrične matrike razdalj

V izračunu razdalj primerjamo vsak objekt z vsakim drugim objektom. Običajno imamo opravka z vrstno sestavo združbe, zato v nadaljevanju govorimo o dveh vrstah Sp_i in Sp_j . Vrsti opišemo z naborom funkcionalnih znakov Z_1, Z_2, \dots, Z_p . Znak Z_j je spremenljivka, katere vrednost x_j je lahko: intervalna, ordinalna, razmernostna ali nominalna, lahko pa vrednost tudi manjka. Za sočasno obravnavo znakov različnih tipov ali z manjkajočimi vrednostmi uporabimo **Gowerjev splošni koeficient podobnosti**. Podobnost med Sp_i in Sp_j je povprečje podobnosti vseh znakov Z_p . Vrednost w_{12j} je tehtani prispevek posameznega znaka, s katero obravnavamo manjkajoče vrednosti in ima vrednost 0 ali 1. Kadar v paru primerjav manjka vrednost za en znak, ima vrednost 0 in tak znak ne prispeva k izračunu podobnosti. Vrednost 1 pomeni vključitev znaka. Za posamezen znak j se parcialna vrednost podobnosti s_{12j} med dvema Sp za vrednosti x_{1j} in x_{2j} izračuna, kot sledi glede na tip podatkov:

- Za vsak posamezen *binarni znak* velja, da je $s_j=1$, če sta stanji skladni, ali $s_j = 0$, če stanji nista skladni. Veljajo odnosi:
 - $w_{12j} = 1$, če sta x_1 in x_2 znana; če kateri od znakov manjka je $w_{12j} = 0$;
 - $s_{12j} = 0$, če je $x_{1j} \neq x_{2j}$;
 - $s_{12j} = 1$, če $x_{1j} = x_{2j}$ ali $x_{1j} = x_{2j} = 0$.
- Za posamezen *kvalitativni znak*, ko sta stanji skladni, je $s_j = 1$; oz. $s_j = 0$, če stanji nista skladni. Veljajo odnosi:
 - $w_{12j} = 1$, če sta x_1 in x_2 znana, če kateri od znakov manjka je $w_{12j} = 0$;
 - $s_{12j} = 0$, če $x_{1j} \neq x_{2j}$;
 - $s_{12j} = 1$, če $x_{1j} = x_{2j}$.

- Za posamezni *semikvantitativni* znak glede na tip znaka uporabimo:
 - »klasični« pristop, če so vrednosti kvantitativne, izračunamo po formuli, navedeni pri kvantitativnem znaku;
 - »metrični« pristop, vrednosti rangiramo kot 1., 2., 3. itd., uporabimo pa enačbo iz kvantitativnega pristopa.
- Za *kvantitativni* znak (zvezni ali diskretni) izračunamo absolutno razliko med dvema vrednostnima $|x_{1j} - x_{2j}|$. Dobljeno vrednost delimo z razponom R_j ($R = \max - \min$). Podobnost med objektoma v znaku s_{12j} izračunamo po formuli:

$$s_{12j} = 1 - \frac{|x_{1j} - x_{2j}|}{R_j}.$$

Gowerjev splošni koeficient podobnosti S_{15} (ob primerjanju S_{p1} in S_{p2}) ima obliko:

$$S_{15(1,2)} = \frac{\sum_{j=1}^p w_{12j} s_{12j}}{\sum_{j=1}^p w_{12j}}$$

z zalogo vrednosti od 0 do 1.

Koeficient podobnosti S_{15} preoblikujemo v Gowerjevo razdaljo D v postopku:

$$D = 1 - S_{15}.$$

Zaloga vrednosti Gowerjeve razdalje je od 0 do 1, pri čemer 0 pomeni popolno podobnost in vrednost 1 popolno različnost med dvema vrstama. Postopek izračuna smo povzeli po Legender in Legender (2012).

Razdalje med posameznimi pari vrst oblikujemo v simetrično matriko Gowerjevih razdalj. Dimenzija matrike je enaka številu vrst.

Analiza glavnih koordinat PCoA

Analiza glavnih koordinat (ang. Principal Coordinates Analysis, tudi metrično multidimenzionalno skaliranje, ang. Metric (Multidimensional) Scaling) je ordinacijska multivariatna statistična metoda, ki simetrično matriko razdalj preoblikuje v pravokotno matriko glavnih koordinat. To so nove, sintetične spremenljivke, ki na sukcesiven način razlagajo variabilnost osnovne simetrične matrike. Glavne koordinate vključimo v nadaljnje analize bodisi v 2D grafično projekcijo, s čimer vizualiziramo položaj objektov

v prostoru in ugotavljamo medsebojne odnose podobnosti oz. prepoznavamo skupine podobnih vrst; bodisi vrednosti glavnih koordinat vključimo v nadaljnje statistične analize. Sukcesivnost se nanaša na deleže variabilnosti, ki jih razlaga posamezna glavna koordinata. Prva glavna koordinata razlaga najvišji delež skupne variabilnosti, druga koordinata razlaga naslednji najvišji delež variabilnosti, ki ga ni razložila prva. Tretja koordinata razlaga naslednji delež variabilnosti, ki ga nista razložili prvi dve, in tako naprej do zadnje koordinate. Prvi dve glavni koordinati praviloma razložita okrog 50 % variabilnosti. Število glavnih koordinat je za ena manjše od števila objektov iz začetne simetrične matrike ($n - 1$). Ker prvih nekaj glavnih koordinat razloži večino variabilnosti večjega števila osnovnih spremenljivk, ordinacijsko preoblikovanje razlagamo kot prispevek k zmanjšanju števila spremenljivk in lažjemu prepoznavanju vzorcev v podatkovni strukturi. Podrobnejša razlaga in matematično ozadje PCoA je predstavljeno v Legendre in Legendre (2012).



PRIMER 5.3

Funkcionalna razpršenost prezimujočih vodnih ptic v sladkovodnem in morskem habitatu – mešani tip podatkov

V predstavitev analize funkcionalne razpršenosti FDis smo vključili isti nabor vrst kot v predhodnem poglavju o prezimujočih vodnih pticah. Popis je bil izveden leta 2020 na dveh območjih v Sloveniji, na reki Dravi in na območju Obale, podatke smo povzeli iz Božič (2020). Skupaj je bilo popisanih 55 vrst ptic v skupnem številu 28.241 osebkov, od tega na Dravi 43 vrst in 21.876 osebkov in na Obali 42 vrst in 6.365 osebkov. Vseh 55 vrst smo opisali s 14 funkcionalnimi znaki. Od tega so bile tri morfometrične spremenljivke numerične z zveznim tipom podatkov, dve spremenljivki sta bili povezani z gnezditveno strategijo in obliko združevanja izven gnezditvene sezone. Obe sta bili kategorični in vrednosti šestih spremenljivk so opisovale vrsto prehrane, bile so kategorične z binarnim kodiranjem. Podatkovni niz funkcionalnih znakov tvori mešani tip podatkov in smo jih povzeli po Storchová in Hořák (2018) (Preglednici 5.9a in b).

Izvorno podatkovno matriko 55 vrst x 14 funkcionalnih znakov smo vključili v izračun Gowerjevih razdalj. Primerjava dveh morfološko in ekološko podobnih vrst gosi sive in beločeje vrne Gowerjevo razdaljo 0,018, medtem ko je razdalja med dvema morfološko in ekološko različnima vrstama sivo gosjo in mokožem 0,710. Gowerjeve razdalje so organizirane v simetrično matriko dimenzije 55 x 55. Zaradi obsežnosti matrike je predstavljenih nekaj vrednosti Gowerjevih razdalj v Preglednici 5.8. Matriko Gowerjevih razdalj smo vključili v analizo glavnih koordinat PCoA. Rezultat te analize so glavne koordinate (Preglednica 5.10), ki povzemajo variabilnost iz simetrične matrike. Prvi dve glavni koordinati razložita 48,6 % variabilnosti.

Tako preoblikovane vrednosti smo nato vključili v izračun indeksa funkcionalne razpršenosti FDis. Vrednost FDis na podatkih vseh 55 vrst (brez upoštevanja lokalitet) ima vrednost 0,250. Vrednosti, ločeno za območje Drave in Obale ter z upoštevanjo abundanco ali prezenco, so navedene v Preglednici 5.11. Funkcionalna razpršenost je na obeh lokalitetah nizka $FDis < 0,3$, z vključitvijo abundanc pa se še zniža. Upad cenilke je izrazitejši na Dravi, k upadu prispeva nekaj zelo abundantnih vrst predvsem raca mlakarica.

Preglednica 5.8 Začetek simetrične matrike dimenzije 55 x 55 z Gowerjevimi razdaljami med posameznimi pari vrst

| | <i>Cygnus olor</i> | <i>Anser albifrons</i> | <i>Anser anser</i> | <i>Branta leucopsis</i> | ... |
|-------------------------|--------------------|------------------------|--------------------|-------------------------|-----|
| <i>Cygnus olor</i> | 0 | | | | |
| <i>Anser albifrons</i> | 0,08005 | 0 | | | |
| <i>Anser anser</i> | 0,06320 | 0,01684 | 0 | | |
| <i>Branta leucopsis</i> | 0,16831 | 0,08825 | 0,10510 | 0 | |
| ... | ... | ... | ... | ... | 0 |

Preglednica 5.9a Pregled 14 funkcionalnih znakov, vključenih v izračun cenilke FDis

| Oznaka | Funkcionalni znak, enota | Tip spremenljivke |
|---|---|--------------------|
| LengthU_MEAN | Povprečna dolžina ptiča cm | Intervalni, zvezna |
| WingU_MEAN | Povprečna dolžina peruti mm | Intervalni, zvezna |
| BillU_MEAN | Povprečna dolžina kljuna mm | Intervalni, zvezna |
| Mating system | Tip razmnoževalnega sistema; M = monogamija, PG = poligamija, PA = poliandrija, PM = promiskuiteta | Kategorična |
| Association outside the breeding season | Oblika združevanja odraslih izven gnezditvenega sezone; GR = agregirano, PA = v parih, SO = solitarno | Kategorična |
| Folivore_Y | Vsaj 10 % prehrane tekom celotnega leta sestavljajo trave, listje, majhne rastline ipd., 1: da; 0: ne | Numerična, binarna |
| Frugivore_Y | Vsaj 10 % prehrane tekom celotnega leta sestavljajo plodovi, 1: da; 0: ne | Numerična, binarna |
| Granivore_Y | Vsaj 10 % prehrane tekom celotnega leta sestavljajo zrnje, semena in oreščki, 1: da; 0: ne | Numerična, binarna |
| Arthropods_Y | Vsaj 10 % prehrane tekom celotnega leta sestavljajo členonožci, 1: da; 0: ne | Numerična, binarna |
| Other invertebrates_Y | Vsaj 10 % prehrane tekom celotnega leta sestavljajo nevretenčarji brez členonožcev, 1: da; 0: ne | Numerična, binarna |
| Fish_Y | Vsaj 10 % prehrane tekom celotnega leta sestavljajo ribe, 1: da; 0: ne | Numerična, binarna |
| Other vertebrates_Y | Vsaj 10 % prehrane tekom celotnega leta sestavljajo vretenčarji brez rib, 1: da; 0: ne | Numerična, binarna |
| Carrion_Y | Vsaj 10 % prehrane tekom celotnega leta sestavlja mrhovina, 1: da; 0: ne | Numerična, binarna |
| Omnivore_Y | Prehrana tekom celotnega leta vsebuje podoben delež rastlin in živali, 1: da; 0: ne | Numerična, binarna |

Preglednica 5.9b Začetni del matrike za 55 vrst ptic z vrednostmi mešanega tipa podatkov za 14 funkcionalnih znakov (razlago znakov glej v Preglednici 5.8a)

| Vrsta | Length | Wing | Bill | Mating | Association | Folivore | Frugivore | Granivore | Arthropods | Invertebrates | Fish | Vertebrates | Carrion | Omnivore |
|-------------------------|--------|--------|-------|--------|-------------|----------|-----------|-----------|------------|---------------|------|-------------|---------|----------|
| <i>Cygnus olor</i> | 152,5 | 605,5 | 78,35 | M | GR | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| <i>Anser albifrons</i> | 71,5 | 425,75 | 48,8 | M | GR | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| <i>Anser anser</i> | 82,5 | 453,5 | 64,05 | M | GR | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| <i>Branta leucopsis</i> | 64 | 401 | 29,1 | M | GR | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| <i>Tadorna tadorna</i> | 62,5 | 318,5 | 50,15 | M | GR | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| <i>Aix galericulata</i> | 45 | 230,5 | 28,4 | M | GR | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 |
| <i>Mareca penelope</i> | 48 | 258,5 | 34,25 | M | GR | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| <i>Mareca strepera</i> | 51 | 260,5 | 41,1 | M,PM | GR | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| <i>Anas crecca</i> | 36 | 183,5 | 35,65 | M | GR | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 |
| ... | | | | | | | | | | | | | | |

Preglednica 5.10 Začetek matrike glavnih koordinat z vrednostmi za 55 vrst ptic za prve tri glavne koordinate – GK

| | GK1 | GK2 | GK3 |
|-------------------------|-------------|--------------|--------------|
| <i>Cygnus olor</i> | 0,09754370 | -0,292920411 | 0,154181330 |
| <i>Anser albifrons</i> | 0,08552260 | -0,273007442 | 0,171105418 |
| <i>Anser anser</i> | 0,09097606 | -0,281008276 | 0,162105200 |
| <i>Branta leucopsis</i> | -0,04601584 | -0,307693864 | 0,168329579 |
| <i>Tadorna tadorna</i> | 0,02780917 | 0,141655965 | 0,107763626 |
| <i>Aix galericulata</i> | -0,28644004 | -0,189412393 | -0,005121924 |
| <i>Mareca penelope</i> | 0,07646021 | -0,228417887 | 0,210096421 |
| ... | ... | ... | ... |

Preglednica 5.11 Cenilke funkcionalne razpršenosti FDis za dve območji: Drava in Obala na mešanem tipu podatkov, ločeno za izračun na abundancah in na prezenci

| Območje / FDis | Abundanca | Prezenca |
|----------------|-----------|----------|
| Drava | 0,1975 | 0,2608 |
| Obala | 0,2261 | 0,2359 |

R > package = »dplyr«

#Pretvorba podatkov v drugačen tip iz: int v numerične

#1. kopija osnovnih podatkov

m1=mesani

#2. pretvorba kolon 6-14 v numerični tip

m1[6:14]=lapply(m1[6:14], FUN=function(y){as.numeric(y)})

```

R > package = »FD«

#FDis - prezimajoče ptice na Dravi in Obali
#Mašani znaki: 55 vrst in 14 znakov: 3 zvezni, 2 kategorična, 9
binarnih
mesani
#1. izračun Gowerjevih razdalj
gm<-gowdis(m1)
gm
#2. Izračun PCoA in indeksov - prezenca
fa_prez<-dbFD(gm, prezenca, w.abun = TRUE, print.pco = TRUE)
fa_prez
#3. Izračun PCoA in indeksov - abundanca
fa_abun<-dbFD(gm, abundanca, w.abun = TRUE, print.pco =
TRUE)
fa_abun
fa_abun$x.values
#4. Izračun indeksov brez abundance
fa_brez<-dbFD(gm)
fa_brez
# Izvoz podatkov
#1.Gowerjeve razdalje
mydf<-as.data.frame(as.matrix(gm))
write.table(mydf, file = "mydf.txt", sep="\t",
            row.names = TRUE, col.names = TRUE)
#2.Eigenvalue
mydf<-as.data.frame(as.matrix(fa_abun$x.values))
write.table(mydf, file = "mydf.txt", sep="\t",
            row.names = TRUE, col.names = TRUE)

```



5.2.2 Funkcionalna specializiranost FSpe (ang. Functional Specialization) je modifikacija predhodno predstavljene cenilke funkcionalne razpršenosti FDis. Ko vrsto S projiciramo v funkcionalni prostor glede na njene funkcionalne znake, lahko stopnjo njene specializacije podamo z Evklidovo razdaljo vrste od centroida (sin. centra gravitacije), ki ga oblikuje celoten nabor vrst v združbi. Bolj kot je vrsta specializirana, v kontekstu funkcionalnih znakov, bolj je oddaljena od centroida in večja je njena Evklidova razdalja. Izračun cenilke FSpe izračunamo po formuli:

$$FSpe = \sum_{j=1}^S (p_j z_{ij}) .$$

Izračun centroida združbe, oddaljenosti vrste od centroida in relativna abundanca posamezne vrste ter razlaga simbolov so predstavljeni v predhodnem Poglavju 5.2. Vsebinsko predstavitev FSpe smo povzeli po Villeger in sod. (2010).

5.2.1.3 Raova kvadratna entropija Q (ang. Rao's Quadratic Entropy, Rao Q) je konceptualno podobna cenilki funkcionalne razpršenosti FDis, po navadi sta obe cenilki v tesni pozitivni korelaciji. Nepričakovana lastnost cenilke Q je, da njena vrednost lahko upade ob dodatku nove vrste, ob dejanskem povečanju vrstnega bogastva. Do upada vrednosti Q pride v primeru, ko dodana nova vrsta poveča abundanco, ne poveča pa funkcionalne raznolikosti, in s tem zmanjša povprečno funkcionalno raznolikost med vrstami. Vhodni podatek za izračun cenilke Q je simetrična matrika razdalj. Izbira koeficienta razdalj je odvisna od tipa podatkov, npr. Evklidove razdalje v primeru intervalnih znakov, Gowerjeve razdalje v primeru mešanih znakov (obe sta predstavljeni v prejšnjem Poglavju 2.1.2) ali Podanijeve razdalje v primeru samo kategoričnega tipa znakov. Podanijevo razdaljo d_{ij} med vrstama i in j izračunamo s formulo:

$$d_{ij} = \frac{u_{ij}}{n},$$

kjer je n število znakov vključenih v obravnavo, u_{ij} je število znakov, ki se razlikujejo med vrstama i in j .

Cenilka Raov Q je definirana kot vsota razdalj d_{ij} med naključno izbranim parom vrst i in j , ki se ob vsaki primerjavi zamenjata, in je utežena z relativno abundanco obeh primerjanih vrst p_i in p_j , računamo jo s formulo:

$$Q = \sum_{i=1}^{S-1} \sum_{j=i+1}^S d_{ij} p_i p_j.$$

Višja vrednost Q pomeni višjo funkcionalno razpršenost združbe. Vsebino Raovega Q smo povzeli po Botta-Dukat (2005).



PRIMER 5.4

Raova kvadratna entropija prezimujočih vodnih ptic v sladkovodnem in morskem habitatu – kategorični tip podatkov

V predstavitev analize Raove kvadratne entropije Q smo vključili isti nabor vrst kot v predhodnem poglavju o prezimujočih vodnih pticah. Popis je bil izveden leta 2020 na dveh območjih v Sloveniji na reki Dravi in na območju Obale. Podatke smo povzeli iz Božič (2020). Skupaj je bilo popisanih 55 vrst ptic v skupnem številu 28.241 osebkov, od tega na Dravi 43 vrst in 21.876 osebkov in na Obali 42 vrst in 6.365 osebkov.

Vseh 55 vrst smo opisali s 5 funkcionalnimi znaki, ki so bili vsi kategorični, zasnovani po binarnem sistemu (0, 1), vrednosti vseh petih spremenljivk so se nanašale na njihov status oz. selitveni značaj (Preglednici 5.13a in 5.13b). Podatkovni niz funkcionalnih znakov tvori samo kvalitativni tip podatkov in smo jih povzeli po Storchová in Hořák (2018).

Pravokotno matriko 55 vrst x 5 funkcionalnih znakov smo vključili v izračun Podanijevih razdalj. Primerjava dveh statusno podobnih vrst gosi (sive in beločele) vrne Podanijevo razdaljo 0,2, medtem ko je razdalja med dvema selitveno različnima vrstama (konopnica in zelenonogi martinec) 0,8. Podanijeve razdalje so organizirane v simetrično matriko dimenzije 55 x 55, zaradi obsežnosti matrike je predstavljenih le nekaj vrednosti Podanijevih razdalj v Preglednici 5.14. Matriko Podanijevih razdalj smo vključili v izračun cenilke Raove kvadratne entropije Q . Vrednost Q na podatkih vseh 55 vrst (brez upoštevanja lokalitet) ima vrednost 4,909. Vrednosti (ločeno za območje Drave in Obale) in upoštevajoč abundanco ali prezenco so navedene v Preglednici 5.15. Vidimo, da vključitev abundance zelo zniža vrednost cenilke, na Dravi bistveno bolj, zaradi neenakomerno porazdeljene abundance med vrstami.

Preglednica 5.13a Pregled 5 funkcionalnih znakov, vključenih v izračun cenilke Q

| Oznaka | Funkcionalni znak | Tip spremenljivke |
|------------------------|---|---------------------------------|
| Territoriality | Vzdržuje teritorij (posamezni osebek, par ali socialna struktura brani teritorij), 1: da; 0: ne | Kategorična (numerična binarna) |
| Sedentary | Vrsta živi v istem območju ob gnezditvi in v negnezditvenem obdobju, 1: da; 0: ne | Kategorična (numerična binarna) |
| Facultative migrant | Vrsta naredi ne napovedljiv premik v izvengnezditvenem obdobju, 1: da; 0: ne | Kategorična (numerična binarna) |
| Short distance migrant | Vrsta v ne gnezditvenem obdobju migrira znotraj zahodnega Palearktisa, 1: da; 0: ne | Kategorična (numerična binarna) |
| Long distance migrant | Vrsta v ne gnezditvenem obdobju migrira izven zahodnega Palearktisa, 1: da; 0: ne | Kategorična (numerična binarna) |

Preglednica 5.13b Začetni del matrike za 55 vrst ptic s kategoričnimi vrednostmi podatkov za 5 funkcionalnih znakov (razlago znakov glej v Preglednici 5.13a)

| Vrsta | Territoriality | Sedentary | Facultative migrant | Short distance migrant | Long distance migrant |
|-------------------------|----------------|-----------|---------------------|------------------------|-----------------------|
| <i>Cygnus olor</i> | 1 | 1 | 0 | 1 | 0 |
| <i>Anser albifrons</i> | 0 | 0 | 0 | 1 | 0 |
| <i>Anser anser</i> | 1 | 0 | 0 | 1 | 0 |
| <i>Branta leucopsis</i> | 0 | 0 | 0 | 1 | 0 |
| <i>Tadorna tadorna</i> | 0 | 0 | 0 | 1 | 0 |
| ... | ... | ... | ... | ... | ... |

Preglednica 5.14 Začetek simetrične matrike dimenzije 55 x 55 s Podanijevimi razdaljami med posameznimi pari vrst

| | <i>Cygnus olor</i> | <i>Anser albifrons</i> | <i>Anser anser</i> | <i>Branta leucopsis</i> | ... |
|-------------------------|--------------------|------------------------|--------------------|-------------------------|-----|
| <i>Cygnus olor</i> | 0 | | | | |
| <i>Anser albifrons</i> | 0,4 | 0 | | | |
| <i>Anser anser</i> | 0,2 | 0,2 | 0 | | |
| <i>Branta leucopsis</i> | 0,4 | 0,0 | 0,2 | 0 | |
| ... | ... | ... | ... | ... | 0 |

Preglednica 5.15 Cenilke Q za dve območji: Drava in Obala na mešanem tipu podatkov, ločeno za izračun na abundanci in na prezenci

| Območje / FDis | Abundanca | Prezenca |
|----------------|-----------|----------|
| Drava | 1,876 | 5,217 |
| Obala | 2,689 | 4,255 |

R > package = »FD«

```
#Kategorični znaki
#Matrika mešanih znakov: 55 vrst in 5 znakov - binarnih
kategoricni
# Izračun Podanijevih razdalj
podani<-gowdis(kategoricni, ord = c("podani"))
podani
# Izračun Q na Podanijevih razdaljah
q_skupaj<-dbFD(kategoricni, ord = c("podani"))
q_skupaj
# Izvoz podatkov
```

#1.Podanijevih razdalje

```
podani<-as.data.frame(as.matrix(podani))
write.table(podani, file = "podani.txt", sep="\t",
            row.names = TRUE, col.names = TRUE)
#Izračun Q, za prezenco
q_p<-dbFD(kategoricni, prezenca, w.abun=TRUE,
ord=c("podani"))
q_p
#Izračun Q, za abundanco
q_a<-dbFD(kategoricni, abundanca, w.abun=TRUE,
ord=c("podani"))
q_a
```



5.3 Na poligonih temelječe analize

Analiza na poligonu temelječega vrednotenja funkcionalne pestrosti se naslanja na koncept ekološke niše, po kateri vrsta zaseda n -razsežni ekološki prostor. Funkcionalno pestrost združbe predstavimo v obliki n -dimenzionalnega poligona, glavna kazalnika sta velikost poligona in razporeditev vrst v poligonu. Dodatno utež pri izračunu cenilk daje razporeditev bolj abundantnih vrst, seveda v primeru analize podatkov z abundancami.

5.3.1 Funkcionalno razhajanje FDiv (ang. Functional Divergence) je prva cenilka iz komponente funkcionalno pestrosti. Izračun cenilke FDiv izhaja iz n -dimenzionalnega hipervolumna oz. poligona, to je znakovnega prostora, ki ga zaseda nabor vrst izbrane združbe. Utež cenilki dajejo vrste z visoko abundanco, ki so blizu roba poligona (Slika 5.6). Cenilka FDiv ima nabor vrednosti od 0 do 1. Vrednost cenilke se približuje 0 z manjšanjem poligona; in kadar so vrste z visoko abundanco razporejene blizu centroida poligona, in vrednost cenilke se pomika proti 1 z večanjem poligona in kadar se abundantne vrste pomikajo proti robu poligona, stran od centroida. Evolucijsko-ekološki kontekst indeksa je v tem, da poudari prisotnost uspešnih vrst, kar je povezano z visoko abundanco ob robu poligona in kar je odraz funkcionalne specializacije vrste.

Najprej definiramo funkcionalni prostor, posamezni funkcionalni znak predstavlja eno dimenzijo prostora. Centroid prostora G_V oblikuje V število oglišč (g_1, g_2, \dots, g_T) in jih računamo:

$$g_k = \frac{1}{V} \sum_{i=1}^V x_{ik},$$

kjer je vektor x_{ik} koordinata vrste i za znak k .

Evklidovo razdaljo posamezne vrste x_i od centroida G_V n -dimenzionalnega hipervolumna izračunamo:

$$dG_i = \sqrt{\sum_{k=1}^T (x_{ik} - g_k)^2}.$$

In povprečna razdalja vseh vrst S od centroida je:

$$\overline{dG} = \frac{1}{S} \sum_{i=1}^S dG_i.$$

V nadaljevanju za vsako vrsto izračunamo vsoto odstopanj, ki so utežene z abundanco Δd , in vsoto absolutnih odstopanj, ki so utežene z abundanco $\Delta |d|$:

$$\Delta d = \sum_{i=1}^s p_i (dG_i - \overline{dG})$$

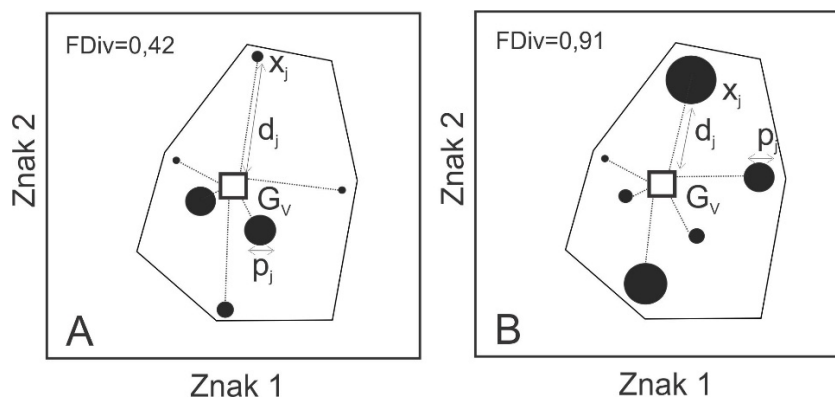
in

$$\Delta |d| = \sum_{i=1}^s p_i |dG_i - \overline{dG}|$$

Funkcionalno razhajanje je potem:

$$FDiv = \frac{\Delta d + \overline{dG}}{\Delta |d| + \overline{dG}}$$

Vsebina cenilke FDiv smo povzeli po Villeger in sod (2010).



Slika 5.6 Razporeditev vrst v 2-dimenzionalnem poligonu, ki ga oblikujeta znak 1 in 2. S krogcem je označen položaj posamezne vrste, velikost krogca pomeni bolj abundantno vrsto. Polna črta omejuje poligon – znakovni prostor, G_V je centroid poligona (kvadrat). X_j je položaj vrste j v 2-dimenzionalnem prostoru, d_j je razdalja med centroidom in vrsto j , p_j je relativna dominanca vrste j . A. Bolj abundantne vrste so bližje centroidu, zato nižja vrednost FDiv. B. Bolj abundantne vrste so ob robu poligona, zato višja vrednost FDiv. Slika je prirejena po Villeger in sod. (2010).



PRIMER 5.5

Funkcionalna razhajanje prezimujočih vodnih ptic v sladkovodnem in morskem habitatu – kvantitativni podatki

V analizo funkcionalnega razhajanja FDiv smo vključili zimski popis vodnih ptic, izveden leta 2020 na dveh območjih v Sloveniji (na reki Dravi in na območju Obale) podatke smo povzeli iz Božič (2020). Skupaj je bilo popisanih 55 vrst ptic v skupnem številu 28.241 osebkov, od tega na Dravi 43 vrst in 21.876 osebkov in na Obali 42 vrst in 6.365 osebkov.

Vseh 55 vrst smo opisali s 5 funkcionalnimi znaki, vsi vključeni znaki so morfometrične spremenljivke in vse so kvantitativne, povzeli smo jih po Storchová in Hořák (2018) (Preglednica 5.8).

Vrednost cenilke FDiv v sladkovodnem habitatu na reki Dravi je 0,460 in je nekoliko nižja kot v morskem habitatu na Obali, kjer ima vrednost 0,561. Razlika v razporeditvi vrst v znakovnem prostoru ni izrazita, kljub vsemu so bile na Dravi vrste z višjo abundanco v povprečju bližje centroidu poligona kot na Obali. Z drugimi besedami, abundantne vrste na Obali so morfološko bolj različne (Slika 5.7).

```
R > package = »FD«
```

```
#FDiv na kvantitativnih znakih
```

```
#1. kontrola podatkov 5 znakov + abundanca  
znaki
```

```
abundanca
```

```
#2. izračun Evklidovih razdalj
```

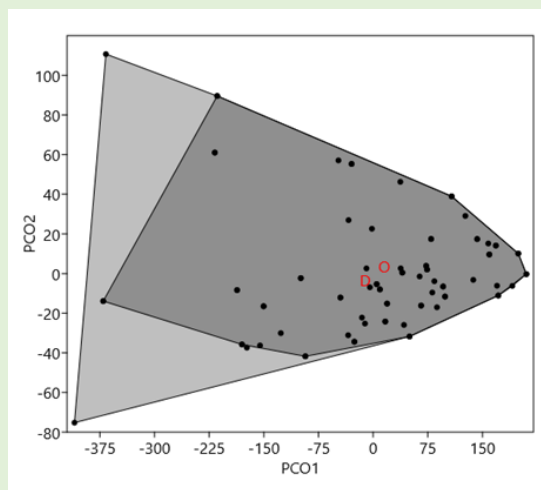
```
znaki_evk<-dist(znaki, method = "euclidean")
```

```
znaki_evk
```

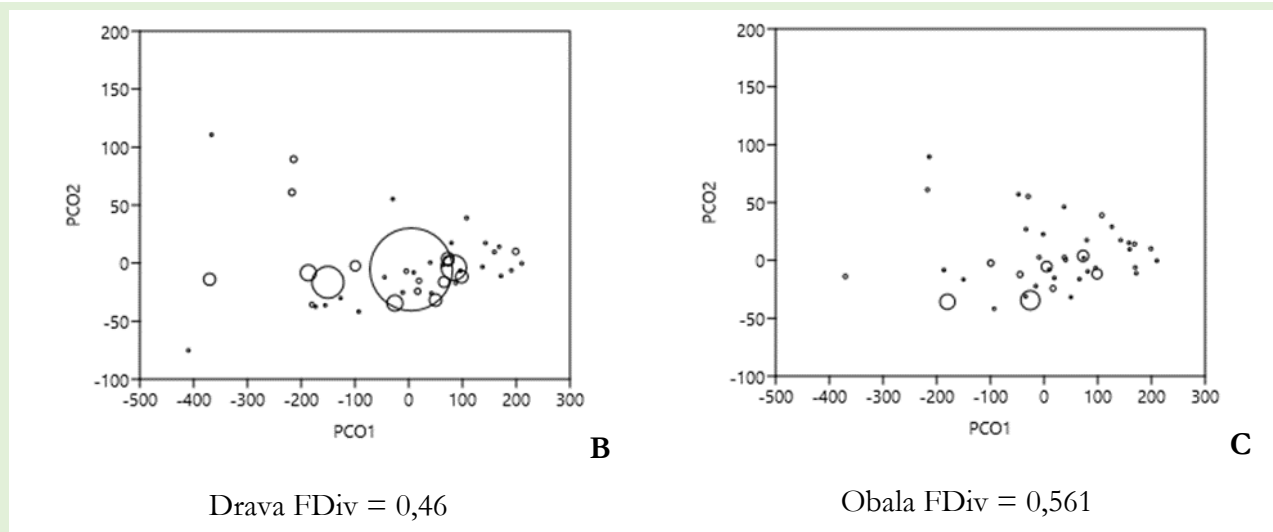
```
#3. izračun FDiv
```

```
FDiv<-dbFD(znaki_evk, abundanca, w.abun=TRUE)
```

```
FDiv
```



A



Slika 5.7 Projekcija znakovnega prostora na prvi dve koordinatni osi za združbi vodnih ptic na Dravi (svetlo sivi poligon) in na Obali (temno sivi poligon) (slika A). Centroida območij D-Drava in O-Obala sta blizu skupaj, poligona znakovnega prostora se v večinskem delu prekrivata, v negativnem delu PCO1 sta dve morfološko specializirani vrsti, ki bistveno povečata volumen, in sicer na pozitivnem delo PCO2 je štorklja, v negativnem belorepec. Velikost poligona je podana s cenilko funkcionalno bogastvo FRic, ki za Dravo znaša 4,832 in za Obalo 4,087. Ločeni sliki porazdelitve vrst v znakovnem prostoru za območje Drave in Obale (sliki B in C). Velikost kroga predstavlja abundanco vrste. Območje Drave vsebuje vrste z izrazito visoko abundanco, ki pa se nahajajo v samem centru, zato tudi nižja vrednost cenilke FDiv v primerjavi z Obalo, kjer ni tako velikih razlik v abundanci, ob tem so nekatere abundantne vrste umaknjene od centroida.



5.3.2 Funkcionalno bogastvo FRic (ang. Functional Richness) je prva cenilka iz komponente funkcionalno bogastvo (zato tudi identično ime komponente in cenilke). Izračun cenilke FRic izhaja iz velikosti volumna poligona, to je znakovnega prostora, ki ga zapolnjuje nabor vrst v izbrani združbi (Sliki 5.6 in 5.7). Velikost prostora oz. njegov volumen je brez enote, primerjave med združbami pa so seveda samo relativne. Volumen poligona – znakovnega prostora je definiran z razporeditvijo vrst v prostoru, pravzaprav najbolj z razporeditvijo ekstremnih vrst, abundanca vrst pa ni upoštevana. Poligon vizualiziramo na prvi dve osi glavnih koordinat iz analize PCoA.

Izračun FRic poteka po korakih:

1. Oblikovanje pravokotne matrike podatkov (vrste x znaki).
2. Vključitev pravokotne matrike znakov v izračun razdalj (kvantitativne spremenljivke: Evklidove razdalje, mešani tip: Gowerjeve razdale, kvantitativni tip Gowerjeve ali Podanijeve razdalje) in oblikovanje simetrične matrike razdalj, vrste x vrste.
3. Vključitev simetrične matrike razdalj v izračun PCoA, oblikovanje pravokotne matrike glavnih koordinat (vrste x koordinate).
4. Vključitev pravokotne matrike glavnih koordinat v izračun velikosti (površine oz. volumna) poligona znakovnega prostora, oz. to je cenilke FRic.

Predstavitve FRic smo povzeli po Legras in sod. (2018 in 2020).



PRIMER 5.6

Funkcionalno bogastvo prezimujočih vodnih ptic v sladkovodnem in morskem habitatu – kvantitativni podatki

V analizo funkcionalnega bogastva FRic smo vključili zimski popis vodnih ptic, izveden leta 2020 na dveh območjih v Sloveniji (na reki Dravi in na območju Obale) podatke smo povzeli iz Božič (2020). Skupaj je bilo popisanih 55 vrst ptic v skupnem številu 28.241 osebkov, od tega na Dravi 43 vrst in 21.876 osebkov in na Obali 42 vrst in 6.365 osebkov.

Vseh 55 vrst smo opisali s 5 funkcionalnimi znaki, vsi vključeni znaki so morfometrične spremenljivke in vse so kvantitativne, povzeli smo jih po Storchová in Hořák (2018) (Preglednica 5.8).

Cenilka FRic za skupni nabor vrst ima vrednost 8,062, v sladkovodnem habitatu na reki Dravi je 4,832 in je nekoliko višja kot v morskem habitatu na Obali, kjer ima vrednost 4,087. Razlika v velikosti znakovnega prostora je lepo vidna na Sliki 5.7, kjer je znakovni prostor za združbo ptic na Dravi večji na območju negativne vrednosti PCO1, zaznamujeta pa jo morfološko specializirani vrsti štoklja in belorepec.

R > package = »FD«

###FRic

#1.a Matrika kvantitativnih znakov: 55 vrst in 5 znakov
znaki

#1.b Matrika prezenca: 2 lokaliteti in 55 vrst
prezenca

#Ukaz za izračun FRic za celoten nabor vrst

FRic_skupaj<-dbFD(znaki)

FRic_skupaj

#Ukaz za izračun FRic ločeno po lokalitetah - prezenca

FRic_p<-dbFD(znaki, prezenca, stand.x = TRUE)

FRic_p



5.3.3 Funkcionalna poravnanoost FEve (ang. Functional Evenness) je prva cenilka iz komponente funkcionalne poravnanoosti (zato tudi identično ime komponente in cenilke).

Cenilka FEve izhaja iz koncepta poravnane porazdelitve vrst in abundanc v multidimenzionalnem znakovnem prostoru, dejansko pa kvantificira enakomernost porazdelitve vrst in poravnanoost abundanc vrst, ki zapolnjujejo znakovni prostor. Cenilka FEve vključuje oboje, porazdelitev vrst v prostoru in porazdelitev abundance vzdolž »skeleta MST« znotraj poligona. Vrednost cenilke FEev je od 0 do 1. Vrednost cenilke se zmanjšuje z zmanjševanjem poravnanoosti razdalje med vrstami oz. neenakomerne razporeditve vrst, ali z zmanjševanjem poravnanoosti abundanc med vrstami; npr. ko večina abundance pripada funkcionalno podobnim vrstam.

Postopek izračuna funkcionalne poravnanoosti FEev obsega več korakov:

1. Oblikovanje matrike znakov T (vrste x znaki) in matrike objektov O (prezenca/abundanca x vrste).
2. Nadaljujemo z:
3. 2.a V primeru kvantitativnih znakov nadaljujemo analizo z izračunom Evklidovih razdalj, -> točka 3.
4. 2.b V primeru mešanih ali kvalitativnih znakov oz. z manjkajočimi vrednostmi najprej izračunamo bodisi Gowerjeve bodisi Podanijeve razdalje in oblikujemo simetrično matriko (vrste x vrste). Sledi njihova vključitev v analizo PCoA in izračun glavnih koordinat. Pravokotno matriko kvantitativnih vrednosti glavnih koordinat vključimo v izračun cenilke -> točka 3.
5. Transformacija porazdelitve vrst iz T dimenzionalnega prostora v porazdelitev vzdolž samo ene osi. To dosežemo z metodo z imenom **drevo najkrajšega razvejanja MST**. MST je graf v obliki drevesa, ki poveže vse vrste, razporejene v T dimenzionalnem prostoru. Vrste S so med seboj povezane tako, da je vsota dolžin l vseh vej v drevesu minimalna. MST prikažemo v obliki grafa.
6. Izračun razdalje med vrstama i in j izvedemo s pomočjo formule za Evklidovo razdaljo d_{ij} :

$$d_{ij} = \sqrt{\sum_{k=1}^T (x_{ik} - x_{jk})^2},$$

kjer sta vektorja x_{ik} in x_{jk} položaja vrste i in j za znak k .

7. V naslednjem koraku posamezno vejo z razdaljo d_{ij} utežimo z relativno abundanco obeh vrst p_i in p_j , ki ju veja povezuje. Relativna abundanca vrste i je:

$$p_i = \frac{a_i}{N}.$$

Skupno število vseh osebkov v združbi je N in a_i je število osebkov vrste i , pri čemer dobimo uteženo razdaljo EW_i :

$$EW_i = \frac{d_{ij}}{p_i + p_j}.$$

8. Posamezno uteženo razdaljo delimo s celokupno vsoto razdalj:

$$PEW_i = \frac{EW_i}{\sum_{l=1}^{S-1} EW_l}.$$

Tako dobimo relativno utežene veje drevesa PEW_i .

9. Če bi bile vrste idealno (uniformno) razporejene vzdolž vej MST, bi bila dolžina posameznega PEW_i enaka $1/(S-1)$. Zato primerjamo posamezni PEW_i z vrednostjo $1/(S-1)$. In tako dobimo indeks funkcionalne poravnosti.

10. Cenilko funkcionalne poravnosti FEev izračunamo po formuli:

$$FEeve = \frac{\sum_{l=1}^{S-1} \min\left(PEW_l, \frac{1}{S-1}\right) - \frac{1}{S-1}}{1 - \frac{1}{S-1}}$$

Drevo najkrajšega razvejanja MST (ang. Minimum Spanning Tree) je oblika grafa, ki z vejami medsebojno poveže vsa oglišča z najmanjšim možnim številom vej in katerih skupna dolžina je najmanjša možna, pri tem ne tvori zank in ni križanja vej. Rezultat je razvejano drevo.



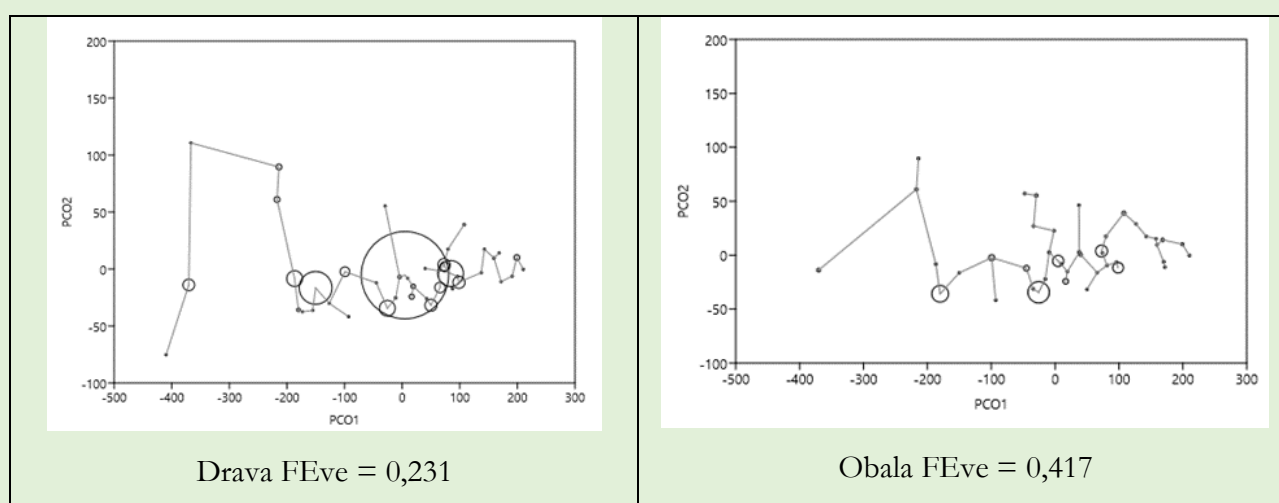
PRIMER 5.7

Funkcionalna poravnost prezimujočih vodnih ptic v sladkovodnem in morskem habitatu – kvantitativni podatki

V analizo funkcionalne poravnosti FEve smo vključili zimski popis vodnih ptic, izveden leta 2020 na dveh območjih v Sloveniji (na reki Dravi in na območju Obale) podatke smo povzeli iz Božič (2020). Skupaj je bilo popisanih 55 vrst ptic v skupnem številu 28.241 osebkov, od tega na Dravi 43 vrst in 21.876 osebkov in na Obali 42 vrst in 6.365 osebkov.

Vseh 55 vrst smo opisali s 5 funkcionalnimi znaki. Vsi vključeni znaki so morfometrične spremenljivke in vse so kvantitativne, povzeli smo jih po Storchová in Hořák 2018) (Preglednica 5.8).

Cenilka FEve za skupni nabor vrst ima vrednost 0,704, v sladkovodnem habitatu na reki Dravi je nizka 0,231 in je občutno nižja kot v morskem habitatu na Obali, kjer ima vrednost 0,417. Vrste so med seboj povezane s kraki MST. Razlika v enakomernosti porazdelitve vrst znotraj znakovnega prostora med Dravo in Obalo je lepo vidna na sliki 5.8. Izstopajo dolge povezave med vrstami v območju negativnih vrednosti PCO1, k nižanju vrednosti FEve na Dravi pa prispeva tudi izrazita neenakost med abundancami vrst, abundance so veliko bolj poravnane na Obali.



Slika 5.8 Projekcija vrst na prvi dve koordinatni osi (PCO) za združbi vodnih ptic na območju Drave in Obale. Položaj vrst v dvodimenzionalnem prostoru prvih dveh glavnih koordinat (PCO), vrste so predstavljene s krogi, velikost kroga ponazarja abudanco vrste. Vrste so med seboj povezane z vejami drevesa najkrajšega razvejanja. Neenakomerna razporeditve vrst v znakovnem prostoru in velike razlike v abudanci vrst, oba ta dejavnika znižujeta vrednost cenilke FEve, ki za Dravo znaša 0,231. Za Obalo ni tako izrazita neporavnost vrst v prostoru, veliko bolj enakomerna pa je poravnost abundanc med vrstami, zato je tudi primerjalno višja vrednost cenilke 0,417 glede na Dravo.

```
R > package = »FD«
```

```
#FEve
```

```
#1.a Matrika kvantitativnih znakov: 55 vrst in 5 zankov  
znaki
```

```
#1.b Matrika abundance: 2 lokaliteti in 55 vrst  
abundanca
```

```
#Ukaz za izračun FEve za celoten nabor vrst
```

```
FEve_skupaj<-dbFD(znaki)
```

```
FEve_skupaj
```

```
#Ukaz za izračun FEve ločeno po lokalitetah - abundanca
```

```
FEve_p<-dbFD(znaki, abundanca, stand.x = TRUE)
```

```
FEve_p
```



Zaključki

- Funkcionalna pestrost je pestrost morfoloških, fizioloških, vedenjskih in/ali fenoloških lastnosti na nivoju osebk, populacije, vrste ali življenjske združbe (objekt).
- Funkcionalni znak je merljiva značilnost osebk, s katero se bodisi odziva na biotske in abiotske dražljaje bodisi vpliva na ekosistemske funkcije ali oboje.
- Funkcionalno pestrost prepoznavamo po treh dimenzijah: funkcionalno bogastvo, funkcionalna raznolikost in funkcionalna poravnost.
- Funkcionalno bogastvo je merilo celokupne širine funkcionalne pestrosti. Prikažemo jo z velikostjo poligona znakovnega prostora in z istoimensko cenilko.
- Funkcionalna pestrost je merilo porazdelitve vrst in njihovih abundanc vzdolž osi funkcionalnih znakov. Za funkcionalno pestrost je več cenilk: funkcionalno razhajanje, funkcionalna razpršenost, Raova kvadratna entropija in funkcionalna specializiranost.
- Funkcionalna poravnost je komponenta, s katero vrednotimo funkcionalno pestrost, če bi bili vsi objekti zastopani v enakem številu.
- Cenilke vrstne pestrosti računamo bodisi s prezenco bodisi z abundanco vrst.
- Na večjem, makroekološkem merilu lahko prikažemo odnos funkcionalne pestrosti vzdolž gradienta, npr. vzdolž geografske širine, nadmorske višine ali klimatskega gradienta, in podobno; ali pa s cenilkami funkcionalne pestrosti primerjamo združbe med seboj.
- Analizo samo enega funkcionalnega znaka izvedemo s surovimi vrednostmi z metodami opisne statistike in s cenilko tehtano povprečje združbe.
- Multivariatno obravnavo funkcionalne pestrosti izvajamo bodisi na razdaljah bodisi na večdimenzionalnem hipervolumnu. Pomemben dejavnik, ki vpliva na izračun cenilk za oba koncepta, je abundanca vrst.

Literatura

- Botta-Dukát Z. 2005. Rao's quadratic entropy as a measure of functional diversity based on multiple traits. *Journal of Vegetation Science*, 16: 533–540.
- Božič L. 2020. Rezultati januarskega štetja vodnih ptic leta 2020 v Sloveniji. *Acrocephalus*, 41 (184/185): 41–53.
- de Bello F., Carmona C.P., Dias A.T.C., Götzenberger L., Moretti M., Berg M.P. 2021. *Handbook of Trait-Based Ecology. From Theory to R Tools*. University Press. Cambridge.
- Gorczyński D., Hsieh C., Tonos Luciano J., Ahumada J., Espinosa S., Johnson S., Rovero F., Santos F., Hugues Andrianarisoa H., Hurtado Astaiza J., Jansen A.P., Kayijamahe C., Guimarães Moreira Lima M., Salvador J., Beaudrot L. 2021. Tropical mammal functional diversity increases with productivity but decreases with anthropogenic disturbance. *Proceeding Royal Society B*, 288: 20202098. <https://doi.org/10.1098/rspb.2020.2098>.
- Kearney R.M., Jusup M., McGeoch A.M., Kooijman A.L.M.S., Chown L.S. 2021. Where do functional traits come from? The role of theory and models. *Functional Ecology*, 1–12.
- Kryštufek B., Janžekovič F., Shenbrot G., Ivajnsič D., Klenovšek T. 2019. Phenotypic plasticity under desert environment constraints: mandible variation in the dwarf fat-tailed jerboa, *Pygeretmus pumilio* (Rodentia: Dipodidae). *Canadian Journal of Zoology*, 97: 940–951.
- Laliberté E., Legendre P. 2010. A distance-based framework for measuring functional diversity from multiple traits. *Ecology*, 91: 299–305.
- Laliberté E., Legendre P., Shipley B. 2014. FD: measuring functional diversity from multiple traits, and other tools for functional ecology. R package version 1.0-12.
- Legendre P., Legendre L. 2012. *Numerical Ecology*. Elsevier. Amsterdam.
- Legras G., Loiseau N., Gaertner J.-C., Poggiale J.-C., Gaertner-Mazouni N. 2020. Assessing functional diversity: the influence of the number of the functional traits. *Theoretical Ecology*, 13: 117–126.
- Legras G., Loiseau N., Gaertner J.-C. 2018. Functional richness: Overview of indices and underlying concepts. *Acta Oecologica*, 87: 34–44.
- Mammola S., Carmona C. P., Guillerme T., Cardoso P. 2021. Concepts and applications in functional diversity. *Functional Ecology* 1–17. <https://doi.org/10.1111/1365-2435.13882>.
- Morel L., Jung V., Chollet S., Ysnel F., Barbe L. 2021. From taxonomic to functional dark diversity: Exploring the causes of potential biodiversity and its implications for conservation. *Journal of Applied Ecology*, 1–14.
- Storchová L., Hořák D. 2018. Life-history characteristics of European birds. *Global Ecology and Biogeography*, 27(4). DOI: 10.1111/geb.12709.
- Villegger S., Miranda J.R., Hernández D.F., Mouillot D. 2010. Contrasting changes in taxonomic vs. functional diversity of tropical fish communities after habitat degradation. *Ecol Appl*, 20: 1512–1522.
- Wong K.L.M., Carmona P.C. 2021. Including intraspecific trait variability to avoid distortion of functional diversity and ecological inference: Lessons from natural assemblages. *Methods in Ecology and Evolution*, 1–20.

6. poglavje



FILOGENETSKA PESTROST

Povzetek

6. poglavje: Filogenetska pestrost

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. Doi: 10.18690/um.fnm.2.2023

Filogenetska pestrost je vrednotenje biodiverzitete na podlagi filogenetskih odnosov med taksoni. Filogenetska pestrost je cenilka biodiverzitete, s katero izrazimo pestrost skupine organizmov na podlagi njihove evolucijske zgodovine, ki je predstavljena s filogenetsko rekonstrukcijo v obliki filogenetskega drevesa. V analizo filogenetske pestrosti vključimo nabor vrst in filogenetske odnose med njimi, ki jih ocenimo na podlagi podobnosti homolognih nukleotidnih zaporedij. Osnovna terminologija sestavin filogenetskega drevesa obsega korenino (izhodišče drevesa s hipotetičnim skupnim prednikom vseh vključenih vrst), veje (evolucijske linije med dvema razvejitvama), notranja razvejitev (evolucijski nastanek dveh taksonov – potomcev iz skupnega prednika) in zunanja razvejitev z listom (konec veje, ki se konča z vrsto). Vzorec razporeditev razvejitev in taksonov na drevesu predstavlja topologijo drevesa. Topologija drevesa, ki predstavlja vstopno informacijo za analizo filogenetske diverzitete, je podana v obliki Newick formata. Tip drevesa, ki ga analiziramo, je koreninjeno, ultrametrično drevo, ki je v celoti razrešeno. To pomeni, da so razdalje od korenino do listov enake za vse taksonne in da so vse razvejitve dihhotomna oz. drevo ne vsebuje politomij. Topologija z

razvejitvijo drevesa predstavlja akumulirane lastnosti v genomu med evolucijskimi linijami. Na tej podlagi je razvita raznolika filogenetska metrika za oceno filogenetske pestrosti za vse tri dimenzije: bogastvo, raznolikost in poravnost. Dimenzija bogastva temelji na vsoti akumuliranih filogenetskih razlik med vrstami. Dimenzija raznolikosti temelji na povprečni razdalji med vrstami. Dimenzija poravnosti temelji na varianci raznolikosti med vrstami. V statističnem kontekstu lahko poenostavimo, da bogastvo temelji na številu lastnosti, raznolikost na razlikah oz. razdaljah med lastnostmi in poravnost temelji na širini variabilnosti lastnosti.

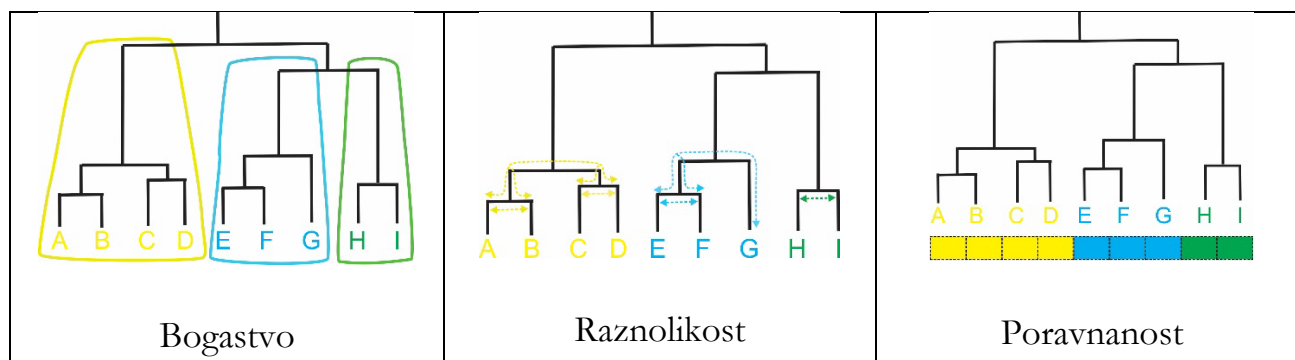
Taksonomska pestrost v vrednotenje biodiverzitete vnaša taksonomsko hierarhijo. Izhodišče za računanje cenilk taksonomske diverzitete je položaj vrst v Linnejevi taksonomski uvrstitvi v rodove, družine, redove, razrede, debla in kraljestva. Višjo taksonomsko pestrost izkazujejo vzorci z več vrstami in z njihovo večjo taksonomsko heterogenostjo. Metrika taksonomske pestrosti se prične z izračunom taksonomske razdalje med dvema vrstama, to je dolžina poti na hierarhičnem filogenetskem drevesu.

6.1 Filogenetska pestrost

Filogenetska informacija je v sodobni ekologiji ključnega pomena, še posebej v njenih disciplinah makroekologiji, ekologiji združb in varstveni biologiji. **Filogenetska pestrost** (ang. Phylogenetic Diversity) je vrednotenje biodiverzitete na podlagi filogenetskih odnosov med taksoni. Filogenetska pestrost je cenilka biodiverzitete s katero izrazimo pestrost skupine organizmov na podlagi njihove evolucijske zgodovine. Informacija o evoluciji organizmov je zajeta v njihovi filogeniji oz. v filogenetski rekonstrukciji (glej Poglavje 14 Filogenetika). V analizo filogenetske pestrosti vključimo nabor vrst (ang. Species Assemblage), filogenetske odnose med njimi ocenimo na podlagi podobnosti homolognih nukleotidnih zaporedij. Informacija o odnosih med taksoni je zajeta v obliki filogenetskega drevesa, ki je osnova za ocenjevanje filogenetske pestrosti. Filogenetsko drevo je graf v obliki drevesa, s katerim so predstavljeni evolucijski odnosi. Besedo »drevo« in terminologijo njegovih sestavin: »korenina, veje, razvejitev in listi« uporabljamo kot sinonime za opisovanje evolucijskih odnosov. Osnovna terminologija sestavin filogenetskega drevesa obsega korenino (izhodišče drevesa s hipotetičnim skupnim prednikom vseh vključenih vrst), veje (evolucijske linije med dvema razvejitvama), notranja razvejitev (evolucijski nastanek dveh taksonov – potomcev iz skupnega prednika) in zunanja razvejitev z listom (konec veje, ki se konča z vrsto). Vzorec razporeditev razvejitev in taksonov na drevesu je topologija drevesa. Topologija drevesa, ki predstavlja vstopno informacijo za analizo filogenetske diverzitete je podana v obliki Newick formata (glej Poglavje 14.4 Filogenetska drevesa), zapis formata je predstavljen v Primeru 6.1. Tip drevesa, ki ga analiziramo, je koreninjeno, ultrametrično drevo, ki je v celoti razrešeno. To pomeni, da so razdalje od korenino do listov enake za vse taksone in da so vse razvejitve dihotomne oz. drevo ne vsebuje politomij. Topologija z razvejanostjo drevesa predstavlja akumulirane lastnosti v genomu med evolucijskimi linijami. Na tej podlagi je razvita raznolika filogenetska metrika za oceno filogenetske pestrosti za vse tri dimenzije: bogastvo, raznolikost in poravnanoost. Dimenzija bogastva temelji na vsoti akumuliranih filogenetskih razlik med vrstami. Dimenzija raznolikosti temelji na povprečni razdalji med vrstami. Dimenzija poravnanoosti temelji na varianci raznolikosti med vrstami. V statističnem kontekstu lahko poenostavimo, da bogastvo temelji na številu lastnosti, raznolikost na razlikah oz. razdaljah med lastnostmi in poravnanoost temelji na širini variabilnosti lastnosti (Slika 6.1). Danes poznamo več kot 70 cenilk filogenetske pestrosti na vseh treh dimenzijah, vključujoč oceno alfa in beta diverzitete. Vsebino poglavja smo črpali iz Clark in sod. (2014), Nei in Kumar (2000), Scherson in Faith (2018), Swenson (2014) ter Tucker in sod. (2017).

Izračune in grafične izdelki, s katerimi je prikazana praktična uporaba postopkov, smo naredili s programskimi orodji v R okolju (R Core Team 2022) in s paketi: »phangorn« (Schliep 2011), »Picante« (Kembel in sod. 2010) in »vegan« (Oksanen in sod. 2019) in programom MEGA 11 (Tamura in sod. 2021).

Cenilke filogenetske pestrosti računamo za vsako dimenzijo posebej s pomočjo formul, predstavljenih v nadaljevanju. Pomen simbolov, uporabljenih v formulah, je naslednji: B je število vej v drevesu. B_i je celoten nabor vej v drevesu. B_{ii} je niz vej, ki povezuje vrsto i s korenino drevesa. L_b je dolžina veje b . R_b je rang veje b (rang, ki ga zavzema takson ali taksoni v naboru vrst, za katere računamo stopnjo endemizma). A_b je abundanca taksona ali skupna abundanca taksonov, ki so na veji b . S je število vrst oz. vrstno bogastvo nabora vrst. S_b je število vrst na veji b . d_{ij} je filogenetska razdalja med vrstama i in j . c_{ij} je vsota dolžin vej od prvega skupnega prednika taksona i in j do korenine.



Slika 6.1 Konceptualni diagrami izračuna posamezne od treh dimenzij filogenetske pestrosti.

Filogenetsko drevo je koreninjeno, ultrametrično in razrešeno. Topologija drevesa z razvejitevami predstavlja evolucijske odnose med vrstami od A do I. Vrste so združene v tri skupine: A – D rumena skupina, E – G modra skupina in H – I zelena skupina. Tri skupine predstavljajo akumulirane razlike med taksoni. Pri dimenziji bogastvo so s poligoni označene veje, katerih dolžino seštejemo. Pri dimenziji raznolikost so s puščicami označene razdalje med vrstami, ki jih ovrednotimo. Pri dimenziji poravnanosť so s pravokotniki označena enakovredna razmerja med vrstami in z barvo vrednost, ki pripada posameznemu kladu.

Prerejeno po Tucker in sod. 2017.

6.1.1 Dimenzija bogastva

S cenilkami iz dimenzije filogenetskega bogastva merimo velikost filogenetske strukture. Računamo jih posebej za alfa in posebej za beta filogenetsko diverzitetu. Cenilke alfa diverzitetete temeljijo na vsoti dolžin vej (PD, PE), taksoni lahko obtežimo tudi z njihovo abundanco (ΔnPD). Računamo lahko filogenetsko razdaljo med pari (PSR) ali vrednotimo filogenetsko izolacijo taksonov (ED). Cenilke, s katerimi primerjamo beta diverzitetu med nabori vrst, temeljijo na deležu skupnih vej v topologiji drevesa (Unifrac). Praktični izračun filogenetske pestrosti je naveden v Primeru 6.1.

Filogenetska pestrost PD (ang. Phylogenetic Diversity):

$$PD = \sum_{b \in B_t} L_b .$$

Filogenetska pestrost obtežena z abundanco ΔnPD , po Barkerju (ang. Abundance-weighted Phylogenetic Diversity sensu Barker):

$$\Delta nPD = \sum_{b \in Bt} L_b A_b .$$

Filogenetski endemizem PE (ang. Phylogenetic Endemism):

$$PE = \sum_{b \in Bt} \frac{L_b}{R_b} .$$

Filogenetsko vrstno bogastvo PSR (ang. Phylogenetic Species Richness):

$$PSR = \sum_i \left(\frac{1}{S-1} \sum_j d_{ij} \right),$$

kjer je d_{ij} :

$$d_{ij} = 0,5(c_{ii} + c_{jj} - c_{ij}) \text{ ali}$$

$$d_{ij} = \frac{1-c_{ij}}{\sqrt{c_{ii}c_{jj}}} .$$

Vsota evlucijskih posebnosti ED (ang. Summed Evolutionary Distinctivnes):

$$ED = \sum_i ED_i \text{ in}$$

$$ED_i = \sum_{b \in Bti} \frac{L_b}{S_b}$$

S cenilko **Unifrac** ocenjujemo vrstni obrat med naboroma vrst i in j . Kjer so: BL_{ij} , je dolžina vej skupna obema naboroma vrst i in j ; BL_i je celotna dolžina vej v naboru i in BL_j je celotna dolžina vej v naboru j .

$$Unifrac = \frac{BL_{ij}}{BL_i + BL_j - BL_{ij}} .$$

6.1.2 Dimenzija raznolikost

S cenilkami iz dimenzije filogenetske raznolikosti merimo povprečne vrednosti pri porazdelitvi vrst v filogenetskem drevesu. Računamo jih posebej za alfa in posebej za beta filogenetsko diverzitetu. Cenilke temeljijo na relativni dolžini vej (avPD), ki jih lahko obtežimo z abundanco (avPD_{Ab}). Lahko jih računamo na razdaljah med pari (MNTD) ali merimo filogenetsko izolacijo (mean(ED)). Cenilke za oceno beta diverzitetu med dvema naboroma vrst temeljijo na relativni dolžini vej (Dpw).

Povprečna filogenetska pestrost avPD (ang. Average Phylogenetic Diversity):

$$avPD = \frac{PD}{S}.$$

Povprečna filogenetska pestrost obtežena z abundanco avPD_{Ab} (ang. Abundance-weighted Phylogenetic Diversity):

$$avPD_{Ab} = \frac{1}{S} \left(B \frac{\sum_{b \in Bt} A_b L_b}{\sum_{b \in Bt} A_b} \right).$$

Povprečna razdalja najbližjega taksona MNTD (ang. Mean Nearest Taxon Distance):

$$MNTD = \frac{1}{S} \sum_i d_{i \min}.$$

Povprečna evolucijska posebnost mean(ED) (ang. Mean Evolutionary Distinctivness):

$$mean(ED) = \frac{\sum_i ED_i}{S}.$$

Beta diverziteta temelječa na povprečni filogenetski razdalji Dpw (ang. Mean Phylogenetic Distance-based β -diversity) (sin. COMDIST, Rao D):

$$Dpw = \sum_{i=1}^{S_1} \frac{1}{S_1} \sum_{j=1}^{S_2} \frac{1}{S_2} d_{ij}.$$

6.1.3 Dimenzija poravnosti

S cenilkami iz dimenzije filogenetske poravnosti merimo odstopanje filogenetskega drevesa od zvezdaste filogenije (v kateri so vsi kladi enakomerno razporejeni). Z drugimi besedami, s to metriko kvantificiramo poravnost vrst na filogenetskem drevesu. Računamo jih samo za alfa filogenetsko diverziteto. Cenilke temeljijo na topologiji drevesa (I_c), razdalji med pari (VPD) in filogenetski izolaciji ($var(ED)$).

Collesov indeks I_c (ang. Colles Index) podaja vsoto absolutnih razlik v vrstnem bogastvu med sestrskima kladoma. Na vsakem notranjem razvejišču definiramo dva sestrška klada. S_{1k} je število vrst, ki izvirajo iz prvega klada, definirane z razvejiščem k , in S_{2k} je število vrst, ki izvirajo iz drugega klada. v je število razvejišč v filogenetskem drevesu. Collesov indeks računamo s formulo:

$$I_c = \sum_{k=1}^v |S_{1k} - S_{2k}|.$$

Varianca razdalj med pari VPD (ang. Variance in Pairwise Distances) (sin. VarTD, Λ^+):

$$VPD = \frac{1}{S(S-1)} \left(\sum_i \sum_{j \neq i} (d_{ij} - Dpw)^2 \right).$$

Varianca evolucijske posebnosti $\text{var}(ED)$ (ang. Variance in Evolutionary Distinctivness):

$$\text{var}(ED) = \frac{1}{s-1} \sum_{i=1}^s \left(ED_i - \frac{\sum_{i=1}^s ED_i}{s} \right)^2.$$



PRIMER 6.1

Filogenetska pestrost v združbah malih sesalcev

Ovrednotili smo filogenetsko pestrost PD v 14 združbah malih sesalcev na Smrekovcu in Peci (Projekt 7.1). Raziskovalno vprašanje smo oblikovali na način ali se filogenetska pestrost združb malih sesalcev iz pogorja Pece razlikujejo od pestrosti iz pogorja Smrekovca.

Za vrednotenje filogenetske pestrosti 14 združb malih sesalcev smo uporabili cenilko Faithova PD. Izračun PD smo izvedli v programskem okolju R s pomočjo paketov »picante« (Kembel in sod. 2010) in »phangorn« (Schliep 2011).

Pripravili smo dva niza podatkov. Prvi niz podatkov so bila nukleotidna zaporedja mitohondrijskega gena za citokrom *b*, v dolžini > 1.000 baznih mest. Podatke smo pridobili iz NCBI GenBank za zaporedja osebkov posameznih vrst z naslednjimi dostopnimi kodami: SORARA: MH332346, SORMIN: MH332368, SORALP: EF636526, CLEGLA: DQ472348, ARVAMP: LR746493, MICAGR: AY167212, MICSUB: FR869887, APOFLA: AY158454 in MUSAVE: AJ225117. Nukleotide smo poravnali z metodo MUSCLE, končni nabor nukleotidov je obsegal 970 mest. Zaporedja smo vključili v izračun drevesa po metodi največjega verjetja in ga skupaj z dolžinami vej shranili v Newick formatu: (((CLEGLA:0.05316457,ARVAMP:0.11132668):0.01922190,(MICAGR:0.06600642,MICSUB:0.06038873):0.03111076):0.09109136,APOFLA:0.09712749,(MUSAVE:0.14240664,(SORMIN:0.06292406,(SORARE:0.04847948,SORALP:0.09705585):0.02520811):0.09038435):0.03713610); Poravnavo nukleotidov in izračun drevesa smo opravili v programu MEGA 11 (Tamura in sod. 2021).

Drugi niz podatkov je obsegal abundance vrst malih sesalcev za 14 združb, matrika je imela velikost 14 vrstic (združbe) x 9 stolpcev (OTU), v poljih so bile abundance (Preglednica 7.3b).

Oba niza podatkov, to je filogenetsko drevo in matriko abundanc, smo uvozili v okolje RStudio in jih analizirali s paketom »picante« in »phangorn«. Izračun Faithovega PD se prične s spremembo nekoreninjenega drevesa Newick formata v koreninjeno drevo, kar smo izvedli z ukazom *midpoint()*. Izračun PD smo izvedli z ukazom *pd()*.

Filogenetska pestrost PD za združbe iz pogorja Smrekovca je obsegala vrednost od 0,45 do 0,66 in povprečno PD = 0,54, vrednosti za združbe iz pogorja Pece so bile od 0,53 do 0,81 in s povprečjem P = 0,66 (Preglednica 6.1). Izračun sklenemo z ugotovitvijo, da združbe malih sesalcev iz Pece izkazujejo višjo filogenetsko pestrost v primerjavi z združbami iz Smrekovca.

Preglednica 6.1 Vrednosti filogenetske pestrosti PD in vrstnega bogastva SP za 7 združb malih sesalcev iz pogorja Smrekovca (S1-S7) in 7 iz pogorja Pece (P8-P14)

| Združba | PD | SR |
|---------|------|----|
| S1 | 0,45 | 4 |
| S2 | 0,59 | 5 |
| S3 | 0,54 | 5 |
| S4 | 0,47 | 3 |
| S5 | 0,66 | 6 |
| S6 | 0,45 | 4 |
| S7 | 0,59 | 5 |
| P8 | 0,59 | 5 |
| P9 | 0,66 | 6 |
| P10 | 0,57 | 5 |
| P11 | 0,53 | 5 |
| P12 | 0,66 | 5 |
| P13 | 0,81 | 7 |
| P14 | 0,81 | 7 |

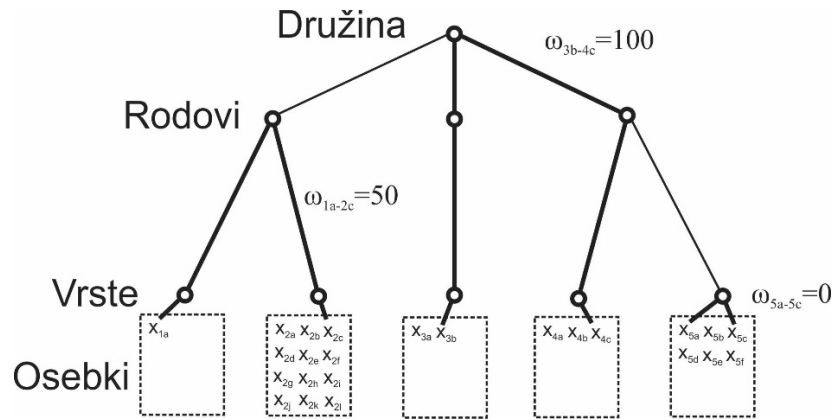
```
R > package = »picante«, »phangorn«
#1. uvoz podatkov, filogenetsko drevo =
# »drevosmrekovec«, združbe = »lok«
drevo<-read.tree(file = "drevosmrekovec")
drevo
lok
#2. koreninjenje drevesa
korendrevo<-midpoint(drevo)
#3. izračun filogenetske pestrosti PD
pd<-pd(lok, korendrevo, include.root = TRUE)
pd
```



6.2 Taksonomska pestrost

Clarke in sod. (1998, 2014) so uvedli vrednotenje biodiverzitete, ki temelji na taksonomski hierarhiji, in ga poimenovali **taksonomska pestrost TD** (ang. Taxonomic Diversity). Izhodišče za računanje cenilk taksonomske pestrosti je položaj vrst v Linnejevi taksonomski klasifikaciji v rodove, družine, redove, razrede, debela in kraljestva. Višjo taksonomsko diverziteteto izkazujejo vzorci z več vrstami in z njihovo večjo taksonomsko heterogenostjo.

Metrika taksonomske pestrosti obsega več cenilk, njihov izračun se prične z vrednostjo **taksonomske razdalje ω_{ij}** (ang. Taxonomic Distance). Taksonomsko razdaljo ω_{ij} (omega) med vrstama x_i in x_j določimo z dolžino poti med položajem x_i in x_j na hierarhičnem filogenetskem drevesu. Dolžina poti je standardizirana z določitvijo vrednosti najdaljše poti, ki ima vrednost 100. Potek in dolžina poti med osebki je predstavljena na Sliki 6.2. Najdaljšo pot ima vrednost $\omega_{34} = 100$, npr. med osebkom x_{3b} in x_{4c} .



Slika 6.2 Razporeditev osebkov petih vrst x_1 , x_2 , x_3 , x_4 in x_5 v taksonomsko hierarhijo: vrsta, rod in družina. Osebk so označeni s črkami, večje število osebkov pomeni višjo abundanco vrste, npr. vrsta x_1 je zastopana s samo enim osebkom x_{1a} . Merilo za izračun taksonomske razdalje je maksimalna dolžina, ki ima vrednost $\omega = 100$ npr. med osebkoma x_{3b} in x_{4c} . Predstavljen je primer s tremi hierarhičnimi nivoji. Maksimalno štirje odseki poti so potrebni od enega do drugega osebka, ki ne pripadata isti vrsti. To pomeni, da je relativna dolžina poti od enega do drugega nivoja $\omega = 25$. Dolžina poti med osebkoma iste vrste npr. x_{5a} in x_{5c} je $\omega = 0$. S krepko črto je označena pot v posameznih primerih. S krogcem je označen posamezni hierarhični nivo. Prirejeno po Clarke in sod. 2014.

Povprečna taksonomska pestrost Δ (ang. Average Taxonomic Diversity) je povprečje taksonomskih razdalj ω med vsemi osebkoma N v vzorcu, ki pripadajo S vrstam. Delto Δ izračunamo s formulo:

$$\Delta = \frac{\sum \sum_{i < j} \omega_{ij} x_i x_j}{\frac{N(N-1)}{2}}$$

Povprečno taksonomsko razdaljo interpretiramo kot povprečje taksonomskih razdalj med vsemi pari osebkov v vzorcu, oz. kot pričakovano dolžino poti med dvema naključno izbranimi osebkoma v vzorcu.

V primeru, da je taksonomsko drevo omejeno na samo en nivo, če združbo sestavljajo vrste iz samo enega rodu, takrat Δ postane Δ^o in dobi obliko Simpsonovega indeksa diverzitete:

$$\Delta^o = \frac{1 - \sum_i p_i^2}{1 - \frac{1}{N}},$$

kjer je p_i delež vrste x_i : $p_i = \frac{x_i}{N}$.

Povprečno taksonomsko pestrost Δ lahko razumemo kot razširitev Simpsonovega indeksa. Z drugimi besedami, Δ je v močni povezavi s Simpsonovim indeksom. Odstranitev vpliva vrstne abundance in natančnejšo predstavitev taksonomske hierarhije dosežemo s kvocientom med Δ in Δ^o . S tem postopkom odstranimo vpliv Simpsonovega

indeksa in dobimo cenilko imenovano **povprečno taksonomsko razlikovanje** Δ^* (ang. Average Taxonomic Distinctness):

$$\Delta^* = \frac{\sum \sum_{i < j} \omega_{ij} x_i x_j}{\sum \sum_{i < j} x_i x_j}.$$

Delta Δ^* interpretiramo kot pričakovano taksonomsko razdaljo med dvema naključno izbranimi osebkom v vzorcu, ki ne pripadata isti vrsti.

Taksonomsko razlikovanje na podatkih o prezenci vrst izračunamo s cenilko **taksonomska širina** Δ^+ (ang. Taxonomic Breadth). Delta plus izračunamo s formulo:

$$\Delta^+ = \frac{\sum \sum_{i < j} \omega_{ij}}{\frac{S(S-1)}{2}}.$$

Taksonomska širina Δ^+ daje povprečno taksonomsko razdaljo med vsemi pari vrst v vzorcu.

Variabilnost v taksonomskem razlikovanju (ang. Variation in Taxonomic Distinctness) podamo s cenilko Λ^+ (lambda plus):

$$\Lambda^+ = \frac{\sum \sum_{i < j} \omega_{ij}^2}{\frac{N(N-1)}{2}} - (\Delta^+)^2.$$

Ko taksonomska razlikovanje Δ^+ pomnožimo z vrstnim bogastvom S , dobimo **popolno taksonomsko razlikovanje** $s\Delta^+$ (ang. Total Taxonomic Distinctness):

$$s\Delta^+ = \sum_i \frac{\sum_{j \neq i} \omega_{ij}}{S-1}.$$

Popolno taksonomsko razlikovanje $s\Delta^+$ daje vsoto vseh povprečnih taksonomskih razdalj od vrste i do katerekoli druge vrste. Cenilka $s\Delta^+$ daje najprimernejšo oceno taksonomske širine vzorca. Na višino cenilke ne vplivata samo vrstno bogastvo in abundanca vrst, ampak je bistvena utež na njihovi taksonomski heterogenosti. Npr. vrstno revnejši vzorec z taksonomsko bolj različnimi vrstami ima lahko višjo oceno kot vrstno bogatejši vzorec s taksonomsko sorodnimi vrstami.

Izračun cenilk taksonomske pestrosti za združbe malih sesalcev je predstavljen v Primeru 6.2.

Linnejeva taksonomija

Klasifikacija živih bitij v hierarhični sistem se po njenem utemeljitelju imenuje Linnejeva taksonomija (ang. Linnean Taxonomy). S filogenetskimi rekonstrukcijami raziskujemo in ponazorimo potek evolucije. V sodobni taksonomski analizi združeno obravnavamo Linnejevo taksonomsko razvrstitev v hierarhične kategorije, ki jih nato prikažemo s filogenetskim drevesom. Klasifikacija živih bitij v hierarhični sistem in gradnja filogenetskega drevesa še ni stabilna. K premeščanju vrst oz. taksonov prispevajo predvsem spoznanja s pomočjo metod molekularnega vrednotenja dednine in odkritja fosilnih taksonov. Prerazporeditve organizmov potekajo tako v vertikali smeri med različnimi taksonomskim ravni kot po horizontali med skupinami istega hierarhičnega ranga. Izračun cenilk taksonomske pestrosti je zato odvisen tudi od stanja v taksonomskih raziskavah. Pred računanjem cenilk taksonomske pestrosti priporočamo posodobitev taksonomije obravnavanih taksonov. Celosten in aktualiziran pogled na taksonomijo živih bitij ponuja Katalog življenja **Catalogue of Life** dostopen na: <https://www.checklistbank.org/> (Bánki in sod. 2022, Catalogue of Life Checklist 2022).



PRIMER 6.2

Taksonomska pestrost v združbah malih sesalcev

Na koroških gorah Smrekovec in Peca smo vzorčili male sesalce iz skupin rovk Soricomorpha in glodavcev Rodentia. V gozdnih habitatih smo na vsaki gori odvzeli 7 vzorcev, skupaj 14. Potrdili smo skupno 9 različnih vrst, 3 iz redu rovk in 6 iz rodu glodavcev. Taksonomski položaj vrst je predstavljen v Preglednici 6.2, abundanca vrst po posameznem vzorcu je predstavljena v Preglednici 6.3.

V raziskavi biodiverzitete malih sesalcev nas je zanimala njihova taksonomska pestrost. Natančneje ali se taksonomska pestrost vzorcev iz Smrekovca razlikuje od vzorcev iz Pece.

Analizo taksonomske diverzitete smo izvedli v okolju R s paketom »vegan« (Oksanen 2019). Z ukazom *taxa2dist()* smo izračunali taksonomske razdalje ω_{ij} med vsemi pari vrst. Najdaljša razdalja je vedno standardizirana na $\omega = 100$, tolikšno razdaljo izkazujejo vse vrste rovk proti vsem vrstam glodavcev. Najkrajšo razdaljo $\omega = 20$ izkazujejo vrste iz istega rodu (Preglednica 6.4). Vključitev taksonomskih razdalj v klastrsko analizo predstavi taksonomske odnose med vrstami v obliki drevesa (Slika 6.3).

V naslednjem koraku, ki smo ga izvedli z ukazom *taxondive()*, smo v analizo vključili taksonomske razdalje iz Preglednice 6.4 in abundance vrst iz Preglednice 6.2. Rezultat je izračun cenilk taksonomske pestrosti: vrstnega bogastva S , delte Δ , delte Δ^* , delte Δ^+ ,

lambde Λ^+ in $s\Delta^+$ (Preglednica 6.5). Najvišjo povprečno taksonomsko pestrost je imel vzorec P13 iz Pece $\Delta = 68,46$, najmanjšo pa vzorec S7 iz Smrekovca $\Delta = 44,97$. Povprečno taksonomsko razlikovanje vrne drugačen rezultat, najvišjo vrednost $\Delta^* = 94,55$ je dosegel vzorec S4 iz Smrekovca, ki je imel tudi najnižje vrstno bogastvo $S = 3$ (Preglednica 6.5).

Popolno taksonomsko razlikovanje $s\Delta^+$ je najprimernejša cenilka taksonomske širine vzorca. Vzorci iz Pece so imeli vrednost $s\Delta^+$ v razponu od 365 do 550 s povprečjem $s\Delta^+ = 447,3$, vzorci iz Smrekovca pa od 240 do 456 s povprečjem $s\Delta^+ = 363,5$ (Preglednica 6.5).

Analizo taksonomske pestrosti na primeru vzorcev malih sesalcev lahko sklenemo z ugotovitvijo, da so združbe malih sesalcev na Peci bolj taksonomsko pestre kot na Smrekovcu.

Preglednica 6.2 Taksonomska hierarhija 9 vrst malih sesalcev

| Koda | Vrsta | Rod | Družina | Red | Razred |
|--------|---------------------------------|----------------------|------------|--------------|----------|
| SORARA | <i>Sorex araneus</i> | <i>Sorex</i> | Soricidae | Soricomorpha | Mammalia |
| SORMIN | <i>Sorex minutus</i> | <i>Sorex</i> | Soricidae | Soricomorpha | Mammalia |
| SORALP | <i>Sorex alpinus</i> | <i>Sorex</i> | Soricidae | Soricomorpha | Mammalia |
| CLEGLA | <i>Clethrionomys glareolus</i> | <i>Clethrionomys</i> | Cricetidae | Rodentia | Mammalia |
| ARVTER | <i>Arvicola terrestris</i> | <i>Arvicola</i> | Cricetidae | Rodentia | Mammalia |
| MICSUB | <i>Microtus subterraneus</i> | <i>Microtus</i> | Cricetidae | Rodentia | Mammalia |
| MICAGR | <i>Microtus agrestis</i> | <i>Microtus</i> | Cricetidae | Rodentia | Mammalia |
| APOFLA | <i>Apodemus flavicollis</i> | <i>Apodemus</i> | Muridae | Rodentia | Mammalia |
| MUSAVE | <i>Muscardinus avellanarius</i> | <i>Muscardinus</i> | Gliridae | Rodentia | Mammalia |

Preglednica 6.3 Abundance 9 vrst malih sesalcev v 14 vzorcih (koda vrste je razložena v Preglednici 6.2; SU – vzorec; sp – vrsta)

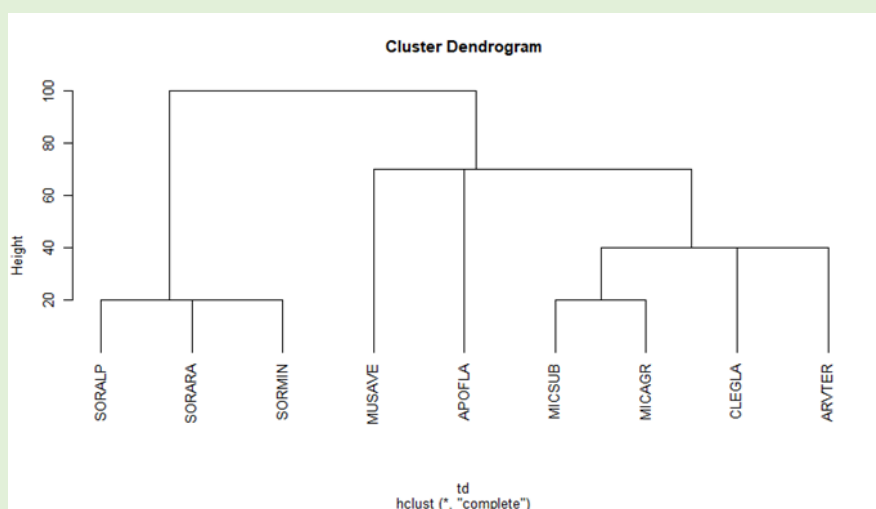
| Gora | SU/sp | SORARA | SORMIN | SORALP | CLEGLA | ARVTER | MICAGR | MICSUB | APOFLA | MUSAVE |
|-----------|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Smrekovec | S1 | 4 | 6 | 0 | 12 | 0 | 0 | 0 | 9 | 0 |
| Smrekovec | S2 | 1 | 6 | 0 | 5 | 0 | 2 | 0 | 0 | 1 |
| Smrekovec | S3 | 3 | 1 | 0 | 4 | 0 | 0 | 1 | 2 | 0 |
| Smrekovec | S4 | 0 | 5 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| Smrekovec | S5 | 2 | 7 | 1 | 5 | 0 | 0 | 2 | 3 | 0 |
| Smrekovec | S6 | 3 | 5 | 0 | 12 | 0 | 1 | 0 | 0 | 0 |
| Smrekovec | S7 | 4 | 21 | 0 | 7 | 0 | 0 | 0 | 1 | 1 |
| Peca | P8 | 7 | 5 | 0 | 3 | 0 | 0 | 0 | 5 | 1 |
| Peca | P9 | 18 | 5 | 1 | 6 | 0 | 0 | 4 | 10 | 0 |
| Peca | P10 | 15 | 7 | 1 | 2 | 0 | 0 | 0 | 7 | 0 |
| Peca | P11 | 2 | 0 | 0 | 9 | 0 | 0 | 1 | 11 | 1 |
| Peca | P12 | 0 | 1 | 1 | 6 | 0 | 0 | 3 | 5 | 0 |
| Peca | P13 | 5 | 2 | 1 | 4 | 0 | 0 | 1 | 3 | 1 |
| Peca | P14 | 3 | 5 | 1 | 5 | 0 | 0 | 1 | 3 | 1 |

Preglednica 6.4 Taksonomske razdalje ω med vsemi pari vrst (koda vrste je razložena v Preglednici 6.2)

| | SORARA | SORMIN | SORALP | CLEGLA | ARVTER | MICSUB | MICAGR | APOFLA | MUSAVE |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| SORARA | 0 | | | | | | | | |
| SORMIN | 20 | 0 | | | | | | | |
| SORALP | 20 | 20 | 0 | | | | | | |
| CLEGLA | 100 | 100 | 100 | 0 | | | | | |
| ARVTER | 100 | 100 | 100 | 40 | 0 | | | | |
| MICSUB | 100 | 100 | 100 | 40 | 40 | 0 | | | |
| MICAGR | 100 | 100 | 100 | 40 | 40 | 20 | 0 | | |
| APOFLA | 100 | 100 | 100 | 70 | 70 | 70 | 70 | 0 | |
| MUSAVE | 100 | 100 | 100 | 70 | 70 | 70 | 70 | 70 | 0 |

Preglednica 6.5 Cenilke taksonomske pestrosti za vzorce iz Smrekovca in Pece (razlaga vzorcev – SU je v Preglednici 6.3)

| SU | S | Delta | Delta* | Lambda+ | Delta+ | S Delta+ |
|----------|---|--------|--------|---------|--------|----------|
| S1 | 4 | 62,5 | 84,9 | 880,6 | 81,7 | 326,7 |
| S2 | 5 | 63,0 | 83,7 | 780,0 | 80,0 | 400,0 |
| S3 | 5 | 67,6 | 82,7 | 780,0 | 80,0 | 400,0 |
| S4 | 3 | 49,5 | 94,5 | 800,0 | 80,0 | 240,0 |
| S5 | 6 | 64,9 | 80,1 | 1064,0 | 76,0 | 456,0 |
| S6 | 4 | 53,2 | 85,3 | 1122,2 | 76,7 | 306,7 |
| S7 | 5 | 45,0 | 77,9 | 621,0 | 83,0 | 415,0 |
| P8 | 5 | 62,4 | 79,0 | 621,0 | 83,0 | 415,0 |
| P9 | 6 | 61,5 | 81,2 | 1064,0 | 76,0 | 456,0 |
| P10 | 5 | 48,8 | 69,6 | 1281,0 | 73,0 | 365,0 |
| P11 | 5 | 50,5 | 75,7 | 369,0 | 79,0 | 395,0 |
| P12 | 5 | 55,8 | 72,7 | 780,0 | 80,0 | 400,0 |
| P13 | 7 | 68,5 | 80,3 | 840,8 | 78,6 | 550,0 |
| P14 | 7 | 67,5 | 79,7 | 840,8 | 78,6 | 550,0 |
| Expected | | 65,052 | 63,4 | | 75,3 | |



Slika 6.3 Drevo po maksimalni metodi taksonomskih odnosov med vrstami rovk in glodavcev (razlaga oznak za vrste je v Preglednici 6.2).

```
R > package = »vegan«  
library(vegan) # Aktivacija paketa »vegan«  
vzorci_smrekovec # Podatki o združbah  
taxa_smrekovec # Podatki o taksonomiji  
td<-taxa2dist(taxa_smrekovec, varstep = TRUE)  
td  
plot(hclust(td), hang = -1)  
tdi<-taxondive(vzorci_smrekovec, td)  
tdi
```



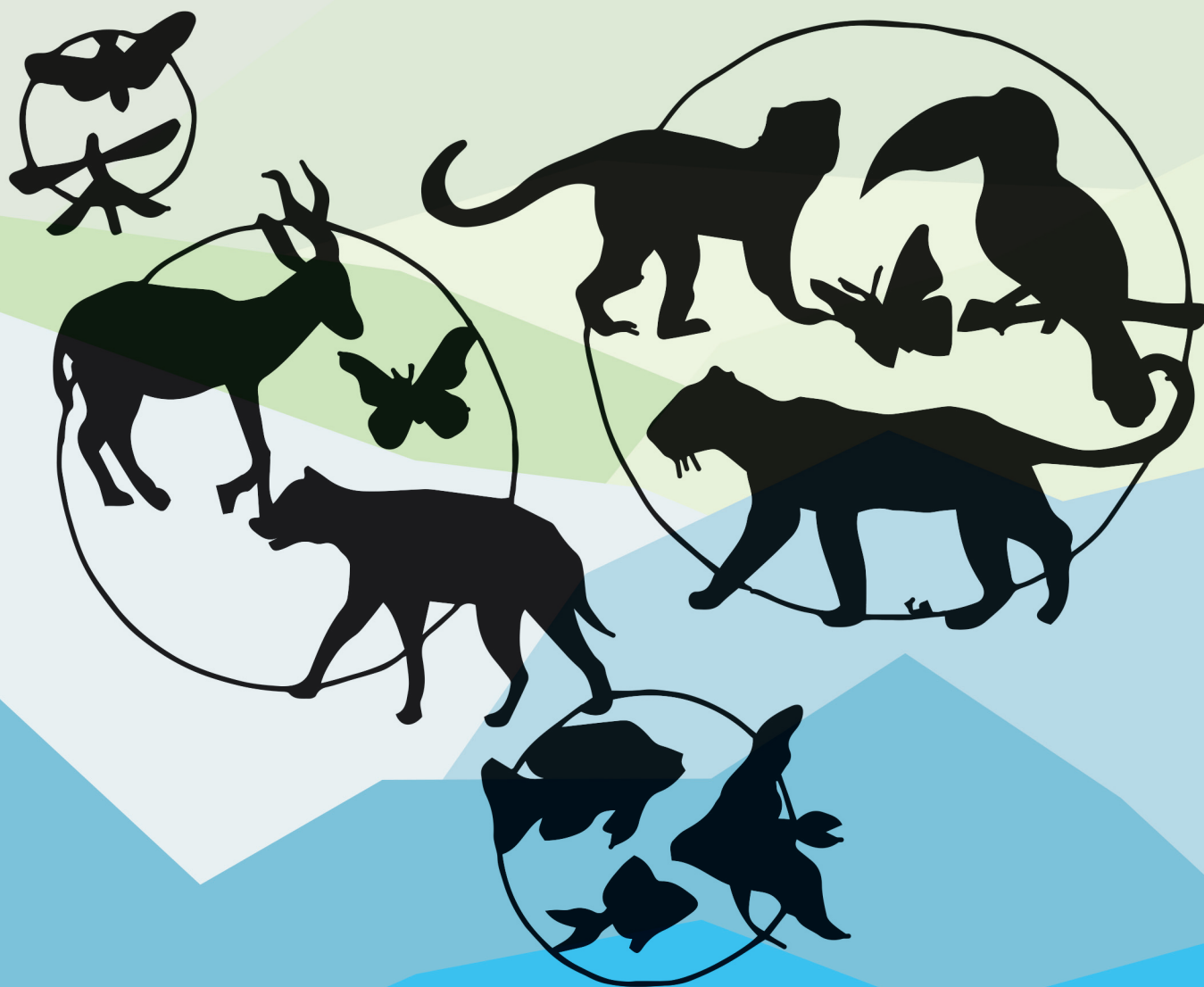
Zaključki

- Filogenetska pestrost je vrednotenje biodiverzitete na podlagi filogenetskih odnosov med taksoni.
- Filogenetska pestrost je cenilka biodiverzitete, s katero izrazimo pestrost skupine organizmov na podlagi njihove evolucijske zgodovine, ki je predstavljena v obliki filogenetskega drevesa.
- Terminologija sestavin filogenetskega drevesa obsega korenino (izhodišče drevesa s hipotetičnim skupnim prednikom vseh vključenih vrst), veje (evolucijske linije med dvema razvejitvama ali med razvejitvijo in listom), razvejitev (evolucijski nastanek dveh taksonov – potomcev iz skupnega prednika) in liste (konec veje, ki se konča z vrsto). Vzorec razporeditev razvejitev in taksonov na drevesu je topologija drevesa, ki je podana v obliki Newick formata.
- Tip drevesa, ki ga analiziramo, je koreninjeno, ultrametrično drevo in v celoti razrešeno. To pomeni, da so razdalje od korenino do listov enake za vse taksone in da so vse razvejitve dihrotomne oz. drevo ne vsebuje politomij.
- Filogenetsko pestrost prepoznavamo na treh dimenzijah (bogastvo, raznolikost, poravnano). Dimenzija bogastva temelji na vsoti akumuliranih filogenetskih razlikah med vrstami. Dimenzija raznolikosti temelji na povprečni razdalji med vrstami. Dimenzija poravnano temelji na varianci raznolikosti med vrstami.
- Taksonomska pestrost v vrednotenje biodiverzitete vnaša taksonomsko hierarhijo. Izhodišče za računanje cenilk taksonomske diverzitete je položaj vrst v Linnejevi taksonomski uvrstitvi v rodove, družine, redove, razrede, debla in kraljestva.
- Višjo taksonomsko pestrost izkazujejo vzorci z več vrstami in z njihovo večjo taksonomsko heterogenostjo. Metrika taksonomske pestrosti se prične z izračunom taksonomske razdalje med dvema vrstama, to je dolžina poti na hierarhičnem filogenetskem drevesu.

Literatura

- Clarke K.R., Gorley R.N., Somerfield P.J., Warwick R.M. 2014. *Change in Marine Communities: An Approach to Statistical Analysis and Interpretation*. 3rd edition. PRIMER-E, Plymouth.
- Clarke K.R., Warwick R.M. 1998. A taxonomic distinctness index and its statistical properties. *Journal of Applied Ecology*, 35: 523–531.
- Kembel S.W., Cowan P.D., Helmus M.R., Cornwell W.K., Morlon H., Ackerly D.D., Blomberg S.P., Webb C.O. 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26: 1463–1464.
- Nei M., Kumar S. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
- Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., Wagner H. 2019. *vegan: Community Ecology Package*. R package version 2.5-6. <https://CRAN.R-project.org/package=vegan>.
- R Core Team. 2022. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Scherson R., Faith P.D. 2018. *Phylogenetic Diversity. Application and Challenges in Biodiversity Science*. Springer, Cham (Switzerland).
- Schliep K.P. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics*, 27(4) 592–593.
- Swenson G.N. 2014. *Functional and Phylogenetic Ecology in R*. Springer, New York.
- Tamura K., Stecher G., Kumar S. 2021. MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution* <https://doi.org/10.1093/molbev/msab120>.
- Tucker C.M., Cadotte W.M., Carvalho B.S., Jonathan Davies T.J., Ferrier S., Fritz A.S., Grenyer R., Helmus R.M., Jin S.L., Mooers O.A., Pavoine S., Purschke O., Redding W.D., Rosauer F.D., Winter W., Mazel F. 2017. A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biological Reviews*, 92: 698–715. doi: 10.1111/brv.12252.

7. poglavje



ANALIZA EKOLOŠKIH ZDRUŽB

Povzetek

7. poglavje: Analiza ekoloških združb

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

Analize ekoloških združb izvajamo z namenom opisa združbe, primerjanja združb med seboj ali spremljanje odziva združbe vzdolž ekološkega gradienta. Združbe med seboj primerjamo s podatki o prezenci ali abundanci vrst. Osnovni niz podatkov predstavlja pravokotna matrika, v kateri so združbe v vrsticah, vrste organizmov so v stolpcih (združbe x vrste), vrednosti v celicah so podatki o prezenci, o abundanci, o deležih ali rangih vrst.

Analiza z vidika združb (ali lokalitet) se imenuje Q analiza. Analiza z vidika vrst je R analiza. Primerjave med združbami izvajamo bodisi z izračunom podobnosti S bodisi različnosti D. Dve združbi z enakim naborom vrst in z enako številčnostjo izkazujeta največjo podobnost. Koeficient podobnosti je 1, razdalja med združbama je 0, z večanjem razlik v vrstni sestavi in abundanci se podobnost približuje vrednosti 0 in večja se razdalja med njima. Odnose med združbami predstavljamo s klasterskimi in ordinacijskimi metodami, vse delujejo na izračunu razdalje med združbami. Mer za podajanje razdalj med združbami je več, vgrajene so v programsko okolje R paket »vegan«. Po teoriji niš imajo vrste ekološke preference. Iz nje izhaja napoved, da imajo unimodalno porazdelitev abundance vzdolž gradienta okolja. Izven tolerančnega območja ekološke preference vrsta ni prisotna, zato matrika podatkov združbe x vrste praviloma vsebuje številne ničle. Delež ničel v podatkovnem nizu o sestavi združbe na splošno narašča z višanjem variabilnosti dejavnikov okolja. Odsotnost vrste v dveh združbah privede do pojava dvojne ničle. Vključevanje dvojnih ničel v izračun podobnosti navidezno poveča podobnost med njimi, to je posledica računskih postopkov. Dvojne ničle v analizi združbe niso informativne, sploh pri večkratnem pojavljanju. Njihova interpretacija

generira problem dvojne ničle, ker vplivajo na izračun koeficientov. Pomembna je prava izbira koeficienta, glede na strukturo podatkov in namen analize. Asimetrični koeficienti ne upoštevajo vrednosti dvojne ničle. Simetrični koeficienti vrednosti z dvojno ničlo obravnavajo na enak način kot vse druge vrednosti. Rezultati so seveda posledično različni, zato moramo biti pri izbiri koeficientov pozorni na problem dvojne ničle, ki pripeljejo do paradoksa abundance. V njem so si združbe brez skupnih vrst bolj podobne kot združbe s skupnimi vrstami. Porazdelitev abundance vrst v združbi je po navadi asimetrično porazdeljena, ali pa so števila osebkov med združbami zelo različna. Obe statistični lastnosti korigiramo, da zmanjšamo njun vpliv na izračun razdalje. Zaradi zmanjšanja asimetrije abundance vrednosti transformiramo s korenjenjem ali logaritmiranjem. Vpliv absolutne razlike v abundanci med lokalitetami, odpravimo s tetivno transformacijo ali Hellingerjevo transformacijo, pri tem se ohrani variabilnost v relativni sestavi vrst. Odnose med združbami predstavimo s hierarhično klastersko analizo in multidimenzionalnim skaliranjem. Obe metodi temeljita na simetrični matriki razdalj med združbami. Rezultat klasterske analize je dendrogram, s katerim vizualiziramo odnose med združbami v obliki drevesa. Multidimenzionalno skaliranje je ordinacijska metoda, s katero reduciramo multidimenzionalno informacijo, razporeditev združb predstavimo v dvodimenzionalnem prostoru. Rezultate po obeh metodah interpretiramo, da so združbe razporejene na razsevnem grafu ali na dendrogramu bližje skupaj, tudi bolj podobne oz. imajo podobno vrstno sestavo.

Analizo ekoloških združb (ang. Analysis of Ecological Communities) izvajamo z namenom opisa združbe, razumevanja sestave združbe, primerjanja združb med seboj in preverjanja vpliva biotskih in/ali abiotskih dejavnikov na sestavo združbe. Združbo praviloma popišemo na izbrani lokaliteti (nahajališču) ali več lokalitetah. Posamezen popis združbe je **vzorec**, označimo ga s SU (ang. Sampling Units). Termina združba in lokaliteta uporabljamo kot sopomenki (včasih tudi termin nahajališče), iz konteksta zapisa razberemo, za kaj pravzaprav gre.

V analizo ekološke združbe lahko vključujemo različne kategorije podatkov: prisotnost (sin. prezenco) in številčnost (sin. abundanco) vrst, okoljske spremenljivke (fizikalne, kemijske ipd.), prostorske vrednosti (imena lokalitet in njihove geografske koordinate), funkcionalne lastnosti vrst (biometrične dimenzije, ekološke lastnosti, spol, starost ...) in druge kategorij podatkov. Osnovni niz podatkov v analizi združbe predstavlja pravokotna matrika, kjer so združbe v vrsticah (tudi vzorčene enote, lokalitete, SU) in vrste organizmov v stolpcih (združbe x vrste). Vrednosti v celicah so bodisi PA podatki bodisi abundance vrst, bodisi deleži ali rangi. Vsaki posamezni združbi ali lokaliteti lahko priredimo okoljske spremenljivke, te so v pravokotni matriki, vrstice so lokalitete, okoljske vrednosti so stolpci (združbe x okoljske). Vsaki posamezni vrsti lahko priredimo funkcionalne znake, vrstice so funkcionalni znaki, stolpci so vrste (funkcionalno x vrste) (Preglednica 7.1).

Preglednica 7.1 Primer podatkovnega niza ekoloških združb, dodani so podatki o okoljskih spremenljivkah in funkcionalnih znakih. Polje A – šest vzorčenih mest s tremi vrstami, posamezna vrsta je označena s šest črkovno kodo, začetnico rodovnega in vrstnega imena. Vsaka vrstica je popis posamezne združbe z navedbo abundance za posamezno vrsto. Polje E – šest vzorčenih mest s tremi okoljskimi spremenljivkami. Prvi dve (nadmorska in temperatura) sta intervalni okoljski meritvi, tretja je nominalna klasifikator za habitat. Polje S – tri funkcionalne lastnosti treh vrst

| A | | Vrste | | | E | Okoljske spremenljivke | | |
|---|-------------------------|--------|--------|---------|---|------------------------|-------------|---------|
| | vzorčena mesta | SORARA | SORMIN | CLETGLA | | nadmorska | temperatura | habitat |
| | SU1 | 4 | 6 | 12 | | 160 | 10 | 1 |
| | SU2 | 1 | 6 | 5 | | 290 | 9,6 | 1 |
| | SU3 | 3 | 1 | 4 | | 300 | 8,5 | 1 |
| | SU4 | 0 | 5 | 1 | | 1000 | 4,4 | 2 |
| | SU5 | 2 | 7 | 5 | | 550 | 8,8 | 2 |
| | SU6 | 3 | 5 | 12 | | 1240 | 6,7 | 2 |
| S | Funkcionalni znaki vrst | | | | | | | |
| | Maks. starost | 2 | 2 | 2 | | | | |
| | Leglo | 6 | 5 | 4 | | | | |
| | Masa | 9 | 4 | 22 | | | | |

Q in R analize

Matriko podatkov o združbah in vrstah, ki jo sestavljajo, lahko obravnavamo z dvema pristopoma. Z vidika združbe, to je objekt analize (tudi lokalitete, vzorčne enote), bodisi z vidika vrst, to so lastnosti združbe (prisotnost ali abundanca vrst). Izpostavljam, da

glede na pristop k analizi uporabimo različne cenilke podobnosti in različnosti med primerjanimi objekti. Mere povezanosti med vrstami organizmov v glavnem temeljijo na korelacijskih koeficientih in takšnemu pristopu rečemo **R analiza**. Analiza podatkovne matrike, v kateri podajamo odnose med združbami, imenujemo **Q analiza**. Najprej moramo sprejeti odločitev o namenu analize, tj. ali analize in primerjave izvajamo z vidika združb ali z vidika vrst. S **Q** obliko analize primerjamo med seboj združbe (v tem primeru so vrste spremenljivke, vrednosti so prisotnost oz. abundanca vrst). Z **R** obliko analize primerjamo med seboj vrste organizmov (v tem primeru so spremenljivke združbe). Glede na namen analize podatke ustrezno organiziramo. V primeru **Q** analize so združbe v vrsticah in vrste v stolpcih. V primeru **R** analize matriko transponiramo, tako da so vrste organizmov v vrsticah in združbe v stolpcih.

V tem poglavju so predstavljene vsebine:

- Ekološka podobnost – **Q** analiza, primerjave med združbami.
 - Koeficienti podobnosti.
 - Koeficienti razdalj.
- Ekološka podobnost – **R** analiza, primerjave med vrstami.
 - Podatki o prisotnosti oz. odsotnosti vrst.
 - Podatki o abundanci vrst.
- Problem dvojne ničle in paradoks abundance.
- Transformacija podatkov o sestavi združbe.
- Predstavitev odnosov med združbami.
 - Hierarhično razvrščanje v skupine.
 - Multidimenzionalno skaliranje.

Poglavje o analizi ekoloških združb je narejeno po literaturi: Jesenko in Jesenko 2007, Kotar 2011, Legendre in Legendre 2012 in Oksanen in sod. 2019.



PROJEKT 7.1

Združbe malih sesalcev v alpskih gozdovih

Janžekovič in Čas (2001) sta raziskala združbe malih sesalcev v alpskih gozdovih na Smrekovcu in Peci. Na nadmorski višina od 1100 do 1450 m sta na vsaki gori na sedmih lokalitetah (Preglednica 7.2) vzorčila talne vrste malih sesalcev iz skupin: rovke, voluharice, miši in polhi (Slika 7.1). Prezenco in abundanco vrst na posamezni lokaliteti (tj. vzorčeni enoti) sta ugotavljala s pomočjo pasti za naključni lov talnih malih sesalcev (metoda lovnih jam; ang. Pitfall Trap). Metode zbiranja podatkov so bile razdeljene na: terensko delo, laboratorijsko delo in delo v kabinetu oz. z računalnikom. Na terenu sta postavila 14 linij

pasti, eno linijo je sestavljalo pet lovnih jam. Pasti so bile aktivne eno sezono v obdobju brez snega, to je od začetka junija do konca oktobra. Pasti sta pregledovala v enakomernih časovnih presledkih na vsakih 15 dni. V laboratoriju sta s pomočjo literature determinirala osebke do vrste. Kumulativne podatke vzorčenja ene cele sezone, sta v Excelovi preglednici uredila ločeno za prezenco in abundanco posamezne vrste v posamezni vzorčeni enoti. Izračunala sta deleže posamezne vrste po združbah in vrednosti uredila v range – kabinetno delo. Te preglednice s primarnimi podatki omogočajo različne vidike analiz ekološke združbe in na tem podatkovnem nizu so predstavljeni principi in postopki obdelave podatkov o združbah v tem poglavju.



Slika 7.1 Podlesek, *Muscardinus avellanarius*.

Foto. Dušan Klenovšek.

Preglednica 7.2 Seznam lokalitet z ekspozicijo in preteklo rabo: S-sever, J-jug, V-vzhod, Z-zahod, T-trajna raba, P-pašniki

| Smrekovec | Ekspozicija | Raba | Peca | Ekspozicija | Raba |
|---------------------|-------------|------|-----------------------|-------------|------|
| S1 Podrta bajta sp. | S | T | P8 Javorje | JZ | T |
| S2 Podrta bajta zg. | S | T | P9 Najbrževo sp. | JV | P |
| S3 Črnačeva bajta | S | T | P10 Najbrževo zg. | JV | P |
| S4 Sleme | S | P | P11 Bela peč jug | JV | T |
| S5 Kolarica I zg. | J | T | P12 Bela peč S I zg. | S | T |
| S6 Kolarica II sp. | J | T | P13 Bela peč S II sp. | S | T |
| S7 Ramšak | J | P | P14 Olševa | S | P |





PRIMER 7.1

Tipi podatkov za analizo združbe – podatki

Preglednico s podatki za namen Q analize orientiramo z združbami oz. lokalitetami (SU) v vrsticah in vrstami (V) v stolpcih. Preglednica je dimenzije 14 vrstic krat 9 stolpcev. Skupaj smo zbrali 314 osebkov 9 različnih vrst (Preglednica 7.4), od tega 3 vrste žužkojedov Eulipotyphla in 6 vrst glodavcev Rodentia.

Podatki so predstavljeni na štiri načine:

1. Kot podatki o prisotnosti vrste V_i v združbi Z_j po načelu prisotno–odsotno. Vrednost 1 v posamezni celici Preglednice 7.3a pomeni, da smo prisotnost vrste na tej lokaliteti potrdili, vrednost 0 pa, da prisotnosti nismo potrdili, t. i. PA podatki.
2. Kot podatki o abundanci vrste V_i v združbi Z_j . Vrednost v posamezni celici Preglednice 7.3b pomeni, da je bila določena vrsta zastopana s tolikšnim številom osebkov na tej lokaliteti. Vrednost 0 pomeni, da vrste na tej lokaliteti nismo potrdili oz. v numeričnem žargonu, da je vrsta zastopana z 0 osebki.
3. Kot podatki o deležu vrste V_i v združbi Z_j . Vrednost v posamezni celici Preglednice 7.3c pomeni delež števila določene vrste glede na skupno število osebkov te vrste v vseh združbah.
4. Kot podatki o rangi vrste V_i v združbi Z_j . Vrednost v posamezni celici Preglednice 7.3č pomeni rang, v katerega je bila vrsta uvrščena.

Preglednica 7.3 Prezenca (a) in abundanca (b) 9 vrst malih sesalcev na 14 lokalitetah na območju Smrekovca (S1–7) in Pece (P8–14); razlaga za kodo vrste je v Preglednici 7.4

a. Prezenca (prisotnost vrst)

| SU / V | SORARA | SORMIN | SORALP | CLEGLA | ARVTER | MICAGR | MICSUB | APOFLA | MUSAVE |
|----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| S1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| S2 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 |
| S3 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| S4 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| S5 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 |
| S6 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 |
| S7 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| P8 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| P9 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 |
| P10 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 |
| P11 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 |
| P12 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 |
| P13 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| P14 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |

b. Abundanca (številčnost vrst)

| SU / V | SORARA | SORMIN | SORALP | CLEGLA | ARVTER | MICAGR | MICSUB | APOFLA | MUSAVE |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| S1 | 4 | 6 | 0 | 12 | 0 | 0 | 0 | 9 | 0 |
| S2 | 1 | 6 | 0 | 5 | 0 | 2 | 0 | 0 | 1 |
| S3 | 3 | 1 | 0 | 4 | 0 | 0 | 1 | 2 | 0 |
| S4 | 0 | 5 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| S5 | 2 | 7 | 1 | 5 | 0 | 0 | 2 | 3 | 0 |
| S6 | 3 | 5 | 0 | 12 | 0 | 1 | 0 | 0 | 0 |
| S7 | 4 | 21 | 0 | 7 | 0 | 0 | 0 | 1 | 1 |
| P8 | 7 | 5 | 0 | 3 | 0 | 0 | 0 | 5 | 1 |
| P9 | 18 | 5 | 1 | 6 | 0 | 0 | 4 | 10 | 0 |
| P10 | 15 | 7 | 1 | 2 | 0 | 0 | 0 | 7 | 0 |
| P11 | 2 | 0 | 0 | 9 | 0 | 0 | 1 | 11 | 1 |
| P12 | 0 | 1 | 1 | 6 | 0 | 0 | 3 | 5 | 0 |
| P13 | 5 | 2 | 1 | 4 | 0 | 0 | 1 | 3 | 1 |
| P14 | 3 | 5 | 1 | 5 | 0 | 0 | 1 | 3 | 1 |

c. Delež

| SU / V | SORARA | SORMIN | SORALP | CLEGLA | ARVTER | MICAGR | MICSUB | APOFLA | MUSAVE |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| S1 | 0,060 | 0,079 | 0,000 | 0,148 | 0,000 | 0,000 | 0,000 | 0,153 | 0,000 |
| S2 | 0,015 | 0,079 | 0,000 | 0,062 | 0,000 | 0,667 | 0,000 | 0,000 | 0,167 |
| S3 | 0,045 | 0,013 | 0,000 | 0,049 | 0,000 | 0,000 | 0,077 | 0,034 | 0,000 |
| S4 | 0,000 | 0,066 | 0,000 | 0,012 | 1,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| S5 | 0,030 | 0,092 | 0,167 | 0,062 | 0,000 | 0,000 | 0,154 | 0,051 | 0,000 |
| S6 | 0,045 | 0,066 | 0,000 | 0,148 | 0,000 | 0,333 | 0,000 | 0,000 | 0,000 |
| S7 | 0,060 | 0,276 | 0,000 | 0,086 | 0,000 | 0,000 | 0,000 | 0,017 | 0,167 |
| P8 | 0,104 | 0,066 | 0,000 | 0,037 | 0,000 | 0,000 | 0,000 | 0,085 | 0,167 |
| P9 | 0,269 | 0,066 | 0,167 | 0,074 | 0,000 | 0,000 | 0,308 | 0,169 | 0,000 |
| P10 | 0,224 | 0,092 | 0,167 | 0,025 | 0,000 | 0,000 | 0,000 | 0,119 | 0,000 |
| P11 | 0,030 | 0,000 | 0,000 | 0,111 | 0,000 | 0,000 | 0,077 | 0,186 | 0,167 |
| P12 | 0,000 | 0,013 | 0,167 | 0,074 | 0,000 | 0,000 | 0,231 | 0,085 | 0,000 |
| P13 | 0,075 | 0,026 | 0,167 | 0,049 | 0,000 | 0,000 | 0,077 | 0,051 | 0,167 |
| P14 | 0,045 | 0,066 | 0,167 | 0,062 | 0,000 | 0,000 | 0,077 | 0,051 | 0,167 |

č. Rang. Vrednosti abundance za meje ranga; Rang 0 = 0, Rang 1 = od 1 do 2, Rang 2 = od 3 do 5, Rang 3 = od 6 do 10, Rang 4 = od 11 do 20, Rang 6 = več od 20 osebkov

| SU / V | SORARA | SORMIN | SORALP | CLEGLA | ARVTER | MICAGR | MICSUB | APOFLA | MUSAVE |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| S1 | 2 | 3 | 0 | 4 | 0 | 0 | 0 | 3 | 0 |
| S2 | 1 | 3 | 0 | 2 | 0 | 1 | 0 | 0 | 1 |
| S3 | 2 | 1 | 0 | 2 | 0 | 0 | 1 | 1 | 0 |
| S4 | 0 | 2 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| S5 | 1 | 3 | 1 | 2 | 0 | 0 | 1 | 2 | 0 |
| S6 | 2 | 2 | 0 | 4 | 0 | 1 | 0 | 0 | 0 |
| S7 | 2 | 5 | 0 | 3 | 0 | 0 | 0 | 1 | 1 |
| P8 | 3 | 2 | 0 | 2 | 0 | 0 | 0 | 2 | 1 |
| P9 | 4 | 2 | 1 | 3 | 0 | 0 | 2 | 3 | 0 |
| P10 | 4 | 3 | 1 | 1 | 0 | 0 | 0 | 3 | 0 |
| P11 | 1 | 0 | 0 | 3 | 0 | 0 | 1 | 4 | 1 |
| P12 | 0 | 1 | 1 | 3 | 0 | 0 | 2 | 2 | 0 |
| P13 | 2 | 1 | 1 | 2 | 0 | 0 | 1 | 2 | 1 |
| P14 | 2 | 2 | 1 | 2 | 0 | 0 | 1 | 2 | 1 |

Preglednica 7.4 Nabor vrst malih sesalcev, vključenih v analizo ekološke združbe

| koda | latinsko ime | slovensko ime |
|--------|---------------------------------|----------------------|
| SORARA | <i>Sorex araneus</i> | gozdna rovka |
| SORMIN | <i>Sorex minutus</i> | mala rovka |
| SORALP | <i>Sorex alpinus</i> | gorska rovka |
| CLEGLA | <i>Clethrionomys glareolus</i> | gozdna voluharica |
| ARVTER | <i>Arvicola terrestris</i> | veliki voluhar |
| MICAGR | <i>Microtus agrestis</i> | travniška voluharica |
| MICSUB | <i>Microtus subterraneus</i> | vrtna voluharica |
| APOFLA | <i>Apodemus flavicollis</i> | rumenogrla miš |
| MUSAVE | <i>Muscardinus avellanarius</i> | podlesek |



7.1 Ekološka podobnost – Q analiza, primerjave med združbami

Imamo p združb Z in p vrst V , ki te združbe sestavljajo. Združbe so objekt analize, vrste so lastnosti združb. Primerjave med združbami izvajamo v parih torej tako, da primerjamo vsako združbo z vsako. Podatke organiziramo, da so združbe v vrsticah in vrste v stolpcih (Preglednica 7.5).

Preglednica 7.5 Organizacija matrike podatkov za dve združbi Z_1 in Z_2 s p vrstami V v Q analizi

| vrste / združbi | V_1 | V_2 | V_j | ... | V_p |
|--------------------|----------|----------|----------|-----|----------|
| Z_1 | y_{11} | y_{21} | y_{j1} | ... | y_{p1} |
| Z_2 | y_{12} | y_{22} | y_{j2} | ... | y_{p2} |

Izračun ekološke podobnosti med dvema združbama in predstavitev odnosov med njima je osnovni pristop v analiziranju multivariatnih biodiverzitetnih podatkov. Primerjave med združbama izvajamo z izračunom mere podobnosti in mere različnosti med njima. Dve združbi z enakim naborom vrst in z enako abundanco vrst izkazujeta največjo podobnost oz. najnižjo različnost. Podobnost med združbama se zmanjšuje in različnost se povečuje z večanjem razlik v vrstni sestavi in/ali abundanci. Vse klastrske in ordinacijske statistične metode, s katerimi predstavljamo odnose med združbami, delujejo na izračunu podobnosti ali različnosti med njimi. Koeficiente podobnosti in različnosti razdelimo v dve skupini; na simetrične in asimetrične. **Simetrični koeficienti** vrsto brez predstavnikov v dveh primerjanih združbah vrednotijo na enak način kot vrsto, ki ima predstavnike v vsaj eni združbi oz. v obeh združbah. **Asimetrični koeficienti** iz izračuna izločijo vrsto, ki nima predstavnika v nobeni od primerjanih združb (za več glej Poglavje 7.3). Cenilk za izračun podobnosti ali različnosti med vzorci je mnogo, njihova uporaba je odvisna od namena analize in strukturiranosti podatkov, podrobno jih obravnavata

Legendre in Legendre (2012). Vgrajene so tudi v različne pakete **programskega okolja R**, tukaj je predstavljena uporabo paketov »vegan« in »stats«.

Vrednosti, s katerimi podajamo zastopanost vrst v združbi in jih vključimo v izračun koeficientov, so:

- abundanca vrste kot diskretna vrednost (npr. vrsta V_j je v združbi Z_1 zastopana s 35 osebk);
- delež vrste (npr. vrsta V_j v združbi Z_1 zavzema 45 % delež oz. 0,45);
- rang (npr. vrsta V_j je v združbi Z_1 zastopana v rangi 10–50 oseb);
- v binarni obliki (npr. vrsta V_j je v združbi Z_1 prisotna, kar označimo z 1 ali vrsta V_j v združbi Z_1 ni prisotna, kar označimo z 0).

V primeru manjkajočih podatkov celice pustimo prazne, ne vpisujemo vrednosti 0. Vrednost 0 v matematičnem kontekstu pomeni, da smo v združbi Z_p prešteli nič osebkov vrste V_p .

7.1.1 Koeficienti podobnosti

S **koeficienti podobnosti** podamo mero povezanosti med dvema združbama, označimo jih z veliko črko S (ang. S – Similarity, indeks ob črki S se nanaša na oznake v Legendre in Legendre (2012)). Podobnost zavzema vrednost v razponu od 0 do 1, vrednost 1 imata popolnoma identični združbi z istim naborom vrst in/ali enako abundanco posamezne vrste, oz. če objekt primerjamo s samim seboj. Združbi imata vrednost koeficienta podobnosti enako 0, kadar sta popolnoma različni oz. si ne delita nobenih skupnih vrst.

7.1.1.1 Binarni koeficienti

Podobnosti med dvema združbama izračunamo na podatkih o prisotnosti oz. odsotnosti vrst, t.i. **PA podatki** (PA – prisotno/odsotno). Z binarno lastnostjo opišemo prisotnost ali odsotnost vrste. Dve združbi primerjamo tako, da primerjamo zastopanost vsake posamezne vrste V_j v združbi Z_1 in Z_2 po metodologiji:

| | | Združba Z_2 | | |
|---------------|---|---------------|-----|-----------|
| | | 1 | 0 | |
| Združba Z_1 | 1 | a | b | a+b |
| | 0 | c | d | c+d |
| | | a+c | b+d | p=a+b+c+d |

kjer vrednost 1 pomeni prisotnost vrste V_j v združbi, vrednost 0 pa odsotnost vrste V_j v združbi. Oznaka a pomeni število vrst, ki so prisotne v obeh združbah, in d število vrst, ki niso prisotne nit v prvi niti v drugi združbi. Vrednosti b in c sta števili vrst glede na kombinacijo prisotnosti ali v eni ali v drugi združbi. Število p so vse vrste.

V skupino *simetričnih binarnih koeficientov* uvrščamo tiste, ki v izračunu upoštevajo učinek dvojne ničle (za razlago glej Poglavje 7.3), navajamo samo **koeficient enostavnega ujemanja** – S_1 in ga računamo po formuli:

$$S_1 = \frac{a+d}{p}.$$

Skupina *asimetričnih binarnih koeficientov* ne upošteva učinek dvojne ničle, najbolj znana med njimi sta **Jaccardov koeficient** – S_7 :

$$S_7 = \frac{a}{a+b+c},$$

pri katerem imajo vse lastnosti enako utež in **Sørensenov koeficient** – S_8 , kjer je utež podvojena na dvojni enki:

$$S_8 = \frac{2a}{2a+b+c}.$$



PRIMER 7.2

Jaccardov koeficient podobnosti

Q analiza ekološke združbe na PA podatkih. Glede na prisotnost posamezne vrste primerjamo vsako združbo z vsako in izračunamo vrednosti a , b , c in d , ki jih vključimo v izračun koeficienta. Naveden je primer primerjave med lokalitetama S1–P8 (Preglednica 7.6), kjer ima Jaccardov koeficient podobnosti vrednost 0,8.

Preglednica 7.6 Primerjava združb malih sesalcev na lokaliteti S1 in P8

| SU / V | SORARA | SORMIN | SORALP | CLEGLA | ARVTER | MICAGR | MICSUB | APOFLA | MUSAVE | vsota | p |
|----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|-------|-----|
| S1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | | |
| P8 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | | |
| a | 1 | 1 | | 1 | | | | 1 | | 4 | |
| b | | | | | | | | | | 0 | |
| c | | | | | | | | | 1 | 1 | |
| d | | | 1 | | 1 | 1 | 1 | | | 4 | 9 |

Koeficient enostavnega ujemanja – S_7 : $S_{(S_1, P_8)} = (a + d) / p = (4 + 4) / 9 = 8 / 9 = 0,889$

Sorensenov koeficient – S_8 : $S_{(S_1, P_8)} = 2a / (2a + b + c) = (2 * 4) / (2 * 4 + 0 + 1) = 8 / 9 = 0,889$

Jaccardov koeficient – S_7 : $S_{(S_1, P_8)} = a / (a + b + c) = 4 / (4 + 0 + 1) = 4 / 5 = 0,800$

Izračun vseh Jaccardovih koeficientov med 14 lokalitetami je predstavljen v Preglednici 7.7 dimenzije 14 x 14. Na diagonali so primerjave objektov s samim seboj, zato imajo vrednosti 1. Izračun izvedemo z ukazom *vegdist()* v paketu »vegan«. Jaccardov indeks je podan kot razdalja med dvema lokalitetama S'_7 . Da dobimo podobnost med lokalitetami moramo vrednost odšteti od 1 ($S_7 = 1 - S'_7$). Najbolj podobni so pari lokalitet S5–P9, S7–P8 in P13–P14, pri katerih je vrednost koeficienta 1, to pomeni, da imajo popolnoma identično vrstno sestavo malih sesalcev.

Preglednica 7.7 Vrednosti Jaccardovega koeficienta podobnosti med lokalitetami

| SU | S1 | S2 | S3 | S4 | S5 | S6 | S7 | P8 | P9 | P10 | P11 | P12 | P13 | P14 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----|
| SU | 1 | | | | | | | | | | | | | |
| S2 | 0,500 | 1 | | | | | | | | | | | | |
| S3 | 0,800 | 0,429 | 1 | | | | | | | | | | | |
| S4 | 0,400 | 0,333 | 0,333 | 1 | | | | | | | | | | |
| S5 | 0,667 | 0,375 | 0,833 | 0,286 | 1 | | | | | | | | | |
| S6 | 0,600 | 0,800 | 0,500 | 0,400 | 0,429 | 1 | | | | | | | | |
| S7 | 0,800 | 0,667 | 0,667 | 0,333 | 0,571 | 0,500 | 1 | | | | | | | |
| P8 | 0,800 | 0,667 | 0,667 | 0,333 | 0,571 | 0,500 | 1,000 | 1 | | | | | | |
| P9 | 0,667 | 0,375 | 0,833 | 0,286 | 1,000 | 0,429 | 0,571 | 0,571 | 1 | | | | | |
| P10 | 0,800 | 0,429 | 0,667 | 0,333 | 0,833 | 0,500 | 0,667 | 0,667 | 0,833 | 1 | | | | |
| P11 | 0,500 | 0,429 | 0,667 | 0,143 | 0,571 | 0,286 | 0,667 | 0,667 | 0,571 | 0,429 | 1 | | | |
| P12 | 0,500 | 0,250 | 0,667 | 0,333 | 0,833 | 0,286 | 0,429 | 0,429 | 0,833 | 0,667 | 0,429 | 1 | | |
| P13 | 0,571 | 0,500 | 0,714 | 0,250 | 0,857 | 0,375 | 0,714 | 0,714 | 0,857 | 0,714 | 0,714 | 0,714 | 1 | |
| P14 | 0,571 | 0,500 | 0,714 | 0,250 | 0,857 | 0,375 | 0,714 | 0,714 | 0,857 | 0,714 | 0,714 | 0,714 | 1,000 | 1 |

```
R > package = »vegan«
#Q analiza ekološke združbe PA, data = qpa
jacc<-vegdist(qpa, method = "jaccard")
jacc
s_jacc<-1-jacc
s_jacc
```



7.1.1.2 Kvantitativni koeficienti

Podobnost med dvema združbama lahko izračunamo na podatkih o abundanci vrst, na velikostnem rangju, v katerega je vrsta uvrščena ali deležu, ki ga vrsta zavzema. Kvantitativni koeficient ujemanja temeljijo na skupnem številu ujemajočih se vrst glede na celokupno število vrst, računamo ga:

$$S = \frac{\sum_{n=1}^p k}{p},$$

kjer je k ujemanje vrednosti vrste V_j med obema združbama, p je skupno število vrst. Primer izračuna podobnosti med dvema objektoma z desetimi vrstami:

| Združbe | Vrste | | | | | | | | | | k | p |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|---|----|
| | V ₁ | V ₂ | V ₃ | V ₄ | V ₅ | V ₆ | V ₇ | V ₈ | V ₉ | V ₁₀ | | |
| Z ₁ | 9 | 3 | 7 | 3 | 4 | 9 | 5 | 4 | 1 | 6 | | |
| Z ₂ | 2 | 3 | 2 | 1 | 2 | 9 | 3 | 2 | 1 | 6 | | |
| Ujemanje | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 4 | 10 |

$$S = 4 \text{ ujemanja} / 10 \text{ vrst} = 0,4.$$

Podobno kot pri binarnih tudi pri kvantitativnih koeficientih poznamo dve skupini koeficientov. *Simetrični kvantitativni koeficienti* upoštevajo učinek dvojne ničle in *asimetrični kvantitativni koeficienti*, pri katerih je učinek dvojne ničle izločen (za razlago dvojne ničle glej Poglavje 7.3). Predstavljeni so asimetrični koeficienti: Steinhausov koeficient, Kulczynskijev koeficient, Modificirani Gowerjev koeficient in Hi-kvadrat X^2 koeficient podobnosti.

Steinhausov koeficient – S_{17} je iz skupine *asimetričnih kvantitativnih koeficientov*. Izračunamo ga iz surovih podatkov o abundanci vrst. Steinhausov koeficient primerja dva objekta na podlagi minimalne abundance posamezne vrste:

$$S_{17} = \frac{2W}{A+B},$$

kjer je W vsota minimalnih abundanc za posamezne vrste, A in B pa sta vsoti vseh abundanc za posamezno združbo.

Kulczynskijev koeficient – S_{18} je asimetrični koeficient, ki temelji na minimalni abundanci:

$$S_{18} = \frac{1}{2} \left(\frac{W}{A} + \frac{W}{B} \right),$$

kjer je W vsota minimalnih abundanc za posamezne vrste, A in B sta vsoti vseh abundanc za posamezno združbo.

Modificiran Gowerjev koeficient – S_{19} ja asimetrični koeficient, ki izključi vrste z dvojno ničlo. Prednost koeficienta je v istočasni obravnavi bodisi numeričnih bodisi mešanega tipa podatkov, saj podatkovni niz lahko hkrati sestavljajo intervalne, ordinalne in nominalne spremenljivke. Podobnost med združbama računamo po formuli:

$$S_{19} = \frac{\sum_{j=1}^p w_{12j} s_{12j}}{\sum_{j=1}^p w_{12j}},$$

kjer je $w_{12j} = 0$ v primeru dvojne ničle ($y_{1j} + y_{2j} = 0$), oz. $w_{12j} = 1$ v primeru, ko je zastopana vsaj ena vrsta ($y_{1j} + y_{2j} > 0$). V primeru intervalne ali ordinalne spremenljivke se s_{12j} računa:

$$s_{12j} = 1 - \frac{|y_{1j} - y_{2j}|}{R_j},$$

y_{1j} in y_{2j} sta vrednosti abundance ali vrednosti razpona za vrste y_j v združbi Z_1 in v združbi Z_2 , R_j je razpon vrednosti abundance ali ranga za vrsto V_j in se računa:

$$R_j = \max y_j - \min y_j.$$

V primeru nominalne oz. binarne spremenljivke ima lahko s_{12j} vrednost $s = 1$ v primeru ujemanja vrednosti med združbama Z_1 in Z_2 ali pa $s = 0$ v primeru neujemanja vrednosti za vrsto V_j v združbi Z_1 in v združbi Z_2 .

Hi-kvadrat X^2 koeficient podobnosti (ang. X^2 Similarity) – S_{21} je asimetrični koeficient, ki izključi dvojno ničlo, izračunamo ga iz X^2 po formuli:

$$S_{21} = 1 - D_{15}$$

Izračun D_{15} glej v poglavju 7.2.2.



PRIMER 7.3

Modificirani Gowerjev koeficient podobnosti

Q analiza ekoloških združb, primer izračuna Gowerjevega koeficienta z mešanim tipom podatkov: rang, binarni, abundance in deleži. Koeficient podobnosti med združbama S_1 in P_9 znaša 0,478 (Preglednica 7.8).

Preglednica 7.8 Izračun Modificiranega Gowerjevega koeficienta podobnosti med združbama S1 in P9 na štirih tipih spremenljivk

| tip | rang | | abundanca | | binarni | | | delež | | |
|--------|--------|--------|-----------|--------|---------|--------|--------|--------|--------|-------|
| SU/V | SORARA | SORMIN | SORALP | CLEGLA | ARVTER | MICAGR | MICSUB | APOFLA | MUSAVE | vsota |
| S1 | 2 | 3 | 0 | 12 | 0 | 0 | 0 | 0,153 | 0 | |
| P9 | 4 | 2 | 1 | 6 | 0 | 0 | 1 | 0,169 | 0 | |
| w | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 6 |
| R_j | 6 | 6 | 1 | 11 | | | | 0,186 | 0 | |
| s | 0,667 | 0,833 | 0 | 0,454 | 0 | 0 | 0 | 0,914 | 0 | |
| $w*s$ | 0,667 | 0,833 | 0 | 0,454 | 0 | 0 | 0 | 0,914 | 0 | 2,869 |

$$S_{19} = 2,87 / 6 = 0,478$$

R_j je razpon za vrednosti abundance, ranga ali deleža.

Postopek izračuna za posamezni tip podatka:

$$s_{SORARA} \{S1, P9, rang\} = 1 - \{|2 - 4|\} / 6 = 0,667$$

$$w_{SORARA} \{S1, P9, rang\} = 1$$

$$s_{CLEGLA} \{S1, P9, abundanca\} = 1 - \{|12 - 6|\} / 11 = 0,454$$

$$w_{CLEGLA} \{S1, P9, abundanca\} = 1$$

$$s_{ARVTER} \{S1, P9, binarni\} = 0$$

$$w_{ARVTER} \{S1, P9, binarni\} = 0$$

$$s_{APOFLA} \{S1, P9, delež\} = 1 - \{|0,123 - 0,169|\} / 0,186 = 0,914$$

$$w_{APOFLA} \{S1, P9, delež\} = 1$$

Splošni zapis izračuna vrednosti s in w v Modificiranem Gowerjevem koeficientu podobnosti:

$s\{vrsta V_j \text{ v združbah } Z_1 \text{ in } Z_2\}$ je definiran kot podobnost v vrednosti za vrsto V_j med Z_1 in Z_2 .

$w\{združba Z_1, Z_2, vrsta V_j\}$ je definiran z Boolovim operatorjem, ki ima vrednost 1, ko je vsaj v eni združbi vrsta prisotna in ima vrednost 0, ko vrsta ni prisotna v nobeni od primerjanih združb ali pa v primeru ča v eni od združb, podatek manjka.





PRIMER 7.4

Steinhausov koeficient podobnosti

Primer Q analiza ekološke združbe na podatkih o abundancah. Steinhausov koeficient primerja dve lokaliteti na podlagi manjše abundance. Koeficient podobnosti med združbama S1 in S2 znaša 0,522 (Preglednica 7.9).

Preglednica 7.9 Izračuna Steinhausovega koeficienta podobnosti med združbama S1 in S2

| SU / sp | SORARA | SORMIN | SORALP | CLEGLA | ARVTER | MICAGR | MICSUB | APOFLA | MUSAVE | A | B | W |
|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|----|----|----|
| S1 | 4 | 6 | 0 | 12 | 0 | 0 | 0 | 9 | 0 | 31 | | |
| S2 | 1 | 6 | 0 | 5 | 0 | 2 | 0 | 0 | 1 | | 15 | |
| Min | 1 | 6 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | | | 12 |

Steinhausov koeficient – S_{17} : $S_{(S1-S2)} = 2W / A + B = (2 * 12) / (31 + 15) = 24 / 46 = 0,522$
 Vrednosti Steinhausovih koeficientov med vsemi lokalitetami je predstavljen pri predstavitvi Bray-Curtisovih razdalj v Preglednici 7.11.



7.1.2 Koeficienti razdalj

Koeficienti razdalj oz. mere različnosti temeljijo na dveh konceptih. Na koeficientih podobnosti in na geometriji v ravnini koordinatnega sistema. Vrednost 0 koeficienta razdalje pomeni, da je razdalja med združbama enaka nič, $d = 0$, oz. da sta združbi popolnoma identični, in obratno. Ko vrednost koeficienta narašča, se večja razdalja med združbama. Večja je razdalja, bolj različni sta združbi. Razdaljo med združbami označujemo z veliko črko D (ang. Distance); (indeks ob črki D se nanaša na oznake v Legender in Legender (2012)).

7.1.2.1 Na koeficientih podobnosti temelječe razdalje

izračunamo kot:

$$D = 1 - S,$$

kjer so D – koeficient razdalje, S – koeficient podobnosti (npr. Jaccardov ali Steinhausov koeficient podobnosti).

Bray-Curtisova razdalja – D_{14} temelji na Steinhausovem koeficientu in velja:

$$D_{14} = 1 - S_{17} .$$

Izračun S_{17} glej v Poglavlju 7.1.2.



PRIMER 7.5

Steinhausov in Bray-Curtisov koeficient

Primer Q analiza ekološke združbe na podatkih o abundancah. Steinhausov koeficient podobnosti (Preglednica 7.11) primerja dve lokaliteti na podlagi manjše abundance, njegova nasprotna vrednost je Bray-Curtisov koeficient razdalje (Preglednica 7.10).

Preglednica 7.10 Vrednosti Bray-Curtis koeficientov razdalj med lokalitetami

| SU | S1 | S2 | S3 | S4 | S5 | S6 | S7 | P8 | P9 | P10 | P11 | P12 | P13 | P14 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----|
| S1 | | | | | | | | | | | | | | |
| S2 | 0,478 | | | | | | | | | | | | | |
| S3 | 0,524 | 0,538 | | | | | | | | | | | | |
| S4 | 0,684 | 0,455 | 0,778 | | | | | | | | | | | |
| S5 | 0,373 | 0,314 | 0,355 | 0,556 | | | | | | | | | | |
| S6 | 0,231 | 0,333 | 0,500 | 0,571 | 0,415 | | | | | | | | | |
| S7 | 0,446 | 0,469 | 0,600 | 0,707 | 0,444 | 0,455 | | | | | | | | |
| P8 | 0,346 | 0,444 | 0,438 | 0,571 | 0,366 | 0,476 | 0,491 | | | | | | | |
| P9 | 0,360 | 0,627 | 0,600 | 0,765 | 0,438 | 0,569 | 0,590 | 0,385 | | | | | | |
| P10 | 0,397 | 0,617 | 0,628 | 0,692 | 0,423 | 0,623 | 0,576 | 0,283 | 0,211 | | | | | |
| P11 | 0,273 | 0,641 | 0,486 | 0,935 | 0,500 | 0,511 | 0,621 | 0,511 | 0,441 | 0,607 | | | | |
| P12 | 0,489 | 0,613 | 0,407 | 0,826 | 0,333 | 0,622 | 0,680 | 0,514 | 0,467 | 0,625 | 0,400 | | | |
| P13 | 0,458 | 0,500 | 0,214 | 0,750 | 0,297 | 0,526 | 0,529 | 0,263 | 0,475 | 0,469 | 0,463 | 0,394 | | |
| P14 | 0,360 | 0,294 | 0,267 | 0,538 | 0,128 | 0,350 | 0,434 | 0,250 | 0,429 | 0,451 | 0,442 | 0,371 | 0,167 | |

Preglednica 7.11 Vrednosti Steinhausovih koeficientov podobnosti med lokalitetami

| SU | S1 | S2 | S3 | S4 | S5 | S6 | S7 | P8 | P9 | P10 | P11 | P12 | P13 | P14 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----|
| S1 | | | | | | | | | | | | | | |
| S2 | 0,522 | | | | | | | | | | | | | |
| S3 | 0,476 | 0,462 | | | | | | | | | | | | |
| S4 | 0,316 | 0,545 | 0,222 | | | | | | | | | | | |
| S5 | 0,627 | 0,686 | 0,645 | 0,444 | | | | | | | | | | |
| S6 | 0,769 | 0,667 | 0,500 | 0,429 | 0,585 | | | | | | | | | |
| S7 | 0,554 | 0,531 | 0,400 | 0,293 | 0,556 | 0,545 | | | | | | | | |
| P8 | 0,654 | 0,556 | 0,563 | 0,429 | 0,634 | 0,524 | 0,509 | | | | | | | |
| P9 | 0,640 | 0,373 | 0,400 | 0,235 | 0,563 | 0,431 | 0,410 | 0,615 | | | | | | |
| P10 | 0,603 | 0,383 | 0,372 | 0,308 | 0,577 | 0,377 | 0,424 | 0,717 | 0,789 | | | | | |
| P11 | 0,727 | 0,359 | 0,514 | 0,065 | 0,500 | 0,489 | 0,379 | 0,489 | 0,559 | 0,393 | | | | |
| P12 | 0,511 | 0,387 | 0,593 | 0,174 | 0,667 | 0,378 | 0,320 | 0,486 | 0,533 | 0,375 | 0,600 | | | |
| P13 | 0,542 | 0,500 | 0,786 | 0,250 | 0,703 | 0,474 | 0,471 | 0,737 | 0,525 | 0,531 | 0,537 | 0,606 | | |
| P14 | 0,640 | 0,706 | 0,733 | 0,462 | 0,872 | 0,650 | 0,566 | 0,750 | 0,571 | 0,549 | 0,558 | 0,629 | 0,833 | |

```
R > package = »vegan«
#Izračun Bray-Curtisove razdalje in
Stenihausovega koeficienta
bray<-vegdist(qab, method = "bray")
bray
stein<-1-bray
stein
```



7.1.2.2 Na geometriji temelječe razdalje

Iz nabora na geometriji temelječih koeficientov navajamo samo asimetrične koeficiente, torej tiste, ki v izračun ne vključijo parov z dvojno ničlo (glej Poglavje 7.3). To pomeni, da dvojne ničle ne vplivajo na izračun. Z njimi primerjamo abundance vrst med združbami.

Računamo razdaljo med združbama Z_1 in Z_2 , kjer sta y_{1j} in y_{2j} abundanci vrste V_j v Z_1 in Z_2 ; p so vse vrste. Vrednost pp je število parov, ki niso dvojne ničle, oz. število parov, ki ostanejo po odstranitvi dvojnih ničel. Oznaka y_+ pomeni vsoto vseh osebkov za izbrano vrsto ali v izbrani združbi.

Tetivna razdalja – D_3 (ang. Chord Distance) je asimetrični koeficient, priporočen za primerjanje združb v abundancah vrst. V izračun vključimo surove podatke o abundancah in lahko izračunamo po dveh formulah:

$$D_3 = \sqrt{\sum_{j=1}^p \left(\frac{y_{1j}}{\sqrt{\sum_{j=1}^p y_{1j}^2}} - \frac{y_{2j}}{\sqrt{\sum_{j=1}^p y_{2j}^2}} \right)^2}$$

ali

$$D_3 = \sqrt{2 \left(1 - \frac{\sum_{j=1}^p y_{1j} y_{2j}}{\sqrt{\sum_{j=1}^p y_{1j}^2} \sqrt{\sum_{j=1}^p y_{2j}^2}} \right)}$$

(Opomba. V uporabi so različne modifikacije tetivne razdalje, predstavljeni formuli sta povzeti po Legender in Legender (2012)).

Whittakerjev indeks asociacije – D_9 je asimetrični koeficient, priporočen za podatke o abundancah:

$$D_9 = \frac{1}{2} \sum_{j=1}^p \left| \frac{y_{1j}}{y_{1+}} - \frac{y_{2j}}{y_{2+}} \right|$$

Canberrova razdalja – D_{10} je asimetrični koeficient, priporočen za podatke o abundancah:

$$D_{10} = \sum_{j=1}^p \left[\frac{|y_{1j} - y_{2j}|}{(y_{1j} + y_{2j})} \right]$$

Clarkova razdalja – D_{11} oz. koeficient divergence je modificirana Canberrova razdalja in ga računamo po formuli:

$$D_{11} = \sqrt{\frac{1}{pp} \sum_{j=1}^p \left(\frac{y_{1j} - y_{2j}}{y_{2j} + y_{2j}} \right)^2}$$

Hellingerjeva razdalja – D_{17} je asimetrični koeficient, priporočen v ordinacijskih metodah:

$$D_{17} = \sqrt{\sum_{j=1}^p \left[\sqrt{\frac{y_{1j}}{y_{1+}}} - \sqrt{\frac{y_{2j}}{y_{2+}}} \right]^2}$$

Razdalja med vrstnimi profili – D_{18} je asimetrični koeficient, priporočen za podatke o abundancah:

$$D_{18} = \sqrt{\sum_{j=1}^p \left(\frac{y_{1j}}{y_{1+}} - \frac{y_{2j}}{y_{2+}} \right)^2}$$

Hi-kvadrat X^2 – D_{15} je asimetrična cenilka, ki v izračunu razdalje bolj uteži redke vrste kot pogoste. Hi-kvadrat je priporočena v analizah kjer obravnavamo redke vrste kot indikatorje ekoloških razmer, izračunamo ga po formuli:

Tetivna razdalja – D_3 : $D_{(S1-P9)} = 0,951$

Razdalje med objekti računamo z ukazom `vegdist()` v paketu `vegan`

Vgrajeni so naslednji postopki izračunov: "manhattan", "euclidean", "canberra", "clark", "bray", "kulczynski", "jaccard", "gower", "altGower", "morisita", "horn", "mountford", "raup", "binomial", "chao", "cao" in "mahalanobis".



7.2 Ekološka podobnost – R analiza – primerjave med vrstami

Namen **R analize** je predstavitev odnosov med dvema vrstama V_1 in V_2 iz p združb Z_p .

Imamo dve vrsti V_1 in V_2 , ki sta razporejeni v p združbah Z in imajo abundanco y_{ji} . Vrste obravnavanih organizmov so objekt analize in združbe so lastnosti vrst. Primerjave med vrstami izvajamo v parih, torej tako, da primerjamo vsako vrsto z vsako drugo. Podatke organiziramo tako, da so vrste v vrsticah in združbe v stolpcih (Preglednica 7.13).

Preglednica 7.13 Organizacija matrike podatkov za dve vrsti V_1 in V_2 v p združbah Z_p v R analizi

| združbe / vrsti | Z_1 | Z_2 | Z_j | ... | Z_p |
|-----------------|----------|----------|----------|-----|----------|
| V_1 | y_{11} | y_{21} | y_{j1} | ... | y_{p1} |
| V_2 | y_{12} | y_{22} | y_{j2} | ... | y_{p2} |

Podobnost v številčnosti ali prisotnosti vrst računamo na abundancah ali na PA podatkih. Primerjave med vrstami na PA podatkih izvajamo z izračunom koeficienta podobnosti. Primerjave na abundancah izračunamo s korelacijskim koeficientom. Pri dveh vrstah z enako ali zelo podobno abundanco oz. prezenco vzdolž obravnavanih združb se vrednost koeficienta približuje 1, z razhajanjem vrednosti abundanc oz. prezenc se vrednost koeficienta približuje vrednosti 0.

7.2.1 Podatki o prisotnosti oz. odsotnosti vrst

Podobnost v prezenci med dvema vrstama izračunamo na podatkih o njuni prisotnosti oz. odsotnosti, t. i. **PA podatki** (PA – prisotno/odsotno). Z binarno lastnostjo opišemo prisotnost ali odsotnost vrste. Dve vrsti primerjamo tako, da primerjamo zastopanost dveh primerjanih vrst V_1 in V_2 v posamezni združbi Z_p po metodologiji:

| | | Vrsta V ₂ | | |
|----------------------|---|----------------------|-------|-------------------|
| | | 1 | 0 | |
| Vrsta V ₁ | 1 | a | b | a+b |
| | 0 | c | d | c+d |
| | | a + c | b + d | p = a + b + c + d |

kjer vrednost 1 pomeni prisotnost vrste V_j v obeh združbah in vrednost 0 odsotnost vrste V_j v združbi. Oznaka a pomeni število združb, v kateri sta prisotni obe vrsti, in d število, v kateri ni prisotna niti prva niti druga vrsta. b in c sta števili združb glede na kombinacijo prisotnosti ene ali druge vrste. Število p je število vseh združb.

Mere povezanosti med vrstami vedno izključujejo vrednost lokalitet z dvojno odsotnostjo – d . Navajamo **Jaccardov koeficient – S_7** in **Sørensenov koeficient – S_8** , ki ju računamo po formuli:

$$S_7 = \frac{a}{a + b + c}$$

$$S_8 = \frac{2a}{2a + b + c}$$

7.2.2 Podatki o abundanci vrst

Pri kvantitativnih podatkih za prikaz odnosov med vrstami uporabimo **korelacijski koeficient**. Glede na porazdelitev podatkov uporabimo parametrični **Pearsonov korelacijski koeficient r** ali neparametrični **Spearmanov koeficient korelacije r_s** . Za linearizacijo odnosa pri izračunu korelacije lahko podatke o abundancah predhodno **transformiramo** (Poglavje 7.4).



PRIMER 7.7

Jaccardov koeficient podobnosti med vrstami

Primer R analize ekološke združbe na PA podatkih. Matriko malih sesalcev, na kateri smo izvajali Q analize transponiramo, dobimo matriko, v kateri so vrste v vrsticah in združbe v stolpcih. Na posamezni lokaliteti preverjamo prisotnost dveh vrst in izračunamo vrednosti a , b , c in d , ki jih vključimo v izračun koeficienta. V Preglednici 7.14 je predstavljen primer izračuna Jaccardovega koeficienta med vrstama SORARA-SORMIN in ima vrednost = 0,786, Sørensenov koeficient ima vrednost = 0,88.

Preglednica 7.14 Izračun vrednosti Jaccardovega in Sørensenov koeficienta med vrstama AORARA in SORMIN

| V/SU | S1 | S2 | S3 | S4 | S5 | S6 | S7 | P8 | P9 | P10 | P11 | P12 | P13 | P14 | vsota | n |
|--------|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-------|----|
| SORARA | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | | |
| SORMIN | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | | |
| a | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | 1 | | | 1 | 1 | 11 | |
| b | | | | 1 | | | | | | | | 1 | | | 2 | |
| c | | | | | | | | | | | 1 | | | | 1 | |
| d | | | | | | | | | | | | | | | 0 | 14 |

Sørensenov koeficient – $S_8: S_{(SORARA, SORMIN)} =$

$$2a / 2a + b + c = 2 * 11 / 2 * 11 + 2 + 1 = 22 / 25 = 0,880$$

Jaccardov koeficient – $S_7: S_{(SORARA, SORMIN)} = a / a + b + c = 11 / 11 + 2 + 1 = 11 / 14 = 0,786$

Izračun vseh primerjav je predstavljen v simetrični matriki dimenzije 9 x 9. Izvedemo ga z ukazom `vegdist()` v paketu »vegan«. Jaccardov koeficient je podan kot indeks, torej kot razdalja med dvema vrstama. Da dobimo podobnost med vrstami, moramo vrednost odšteti od 1 ($S_7 = 1 - S'_7$). Najvišja vrednost koeficienta 0,929 v skupni prisotnosti je med vrstama SORMIN in CLEGLA (Preglednica 7.15).

Preglednica 7.15 Vrednosti Jaccardovega koeficienta podobnosti med vrstami

| V | SORARA | SORMIN | SORALP | CLEGLA | ARVTER | MICAGR | MICSUB | APOFLA | MUSAVE |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| SORARA | 1 | | | | | | | | |
| SORMIN | 0,786 | 1 | | | | | | | |
| SORALP | 0,385 | 0,462 | 1 | | | | | | |
| CLEGLA | 0,857 | 0,929 | 0,429 | 1 | | | | | |
| ARVTER | 0,000 | 0,077 | 0,000 | 0,071 | 1 | | | | |
| MICAGR | 0,167 | 0,154 | 0,000 | 0,143 | 0,000 | 1 | | | |
| MICSUB | 0,462 | 0,429 | 0,625 | 0,500 | 0,000 | 0,000 | 1 | | |
| APOFLA | 0,769 | 0,714 | 0,545 | 0,786 | 0,000 | 0,000 | 0,636 | 1 | |
| MUSAVE | 0,500 | 0,357 | 0,200 | 0,429 | 0,000 | 0,143 | 0,300 | 0,417 | 1 |

R > package = »vegan«

```
#R analiza, Jaccard
r_jacc<-vegdist(rpa, method =
"jaccard")
r_jacc
r_jacc_s<-1-r_jacc
r_jacc_s
```

R >

```
#Transpose PA data za R analizo
rpa<-t(qpa)
rpa
#Transpose abundanc za R analizo
rab<-t(qab)
rab
```





PRIMER 7.8

Pearsonov korelacijski koeficient

Abundance nekaterih vrst so med seboj povezane (korelirane), kar pomeni, da s povečevanjem števila osebkov ene vrste narašča (pozitivna korelacija) ali upada (negativna korelacija) število osebkov druge vrste. V primeru negativnega odnosa, torej ob naraščanju števila osebkov prve vrste, število osebkov druge vrste upada. Variabilnost abundanc nekaterih vrst je lahko neodvisna, kar pomeni, da njihova številčnost ni povezana s številčnostjo drugih vrst. Medsebojne odnose med abundancami posameznih vrst preverimo s korelacijsko analizo. V našem podatkovnem nizu smo potrdili samo dve tesnejši pozitivni korelaciji, ki jo izkazujeta para SORARA-APOFLA in SORALP-MICSUB (Preglednica 7.16).

Preglednica 7.16 Vrednosti Pearsonovega korelacijskega koeficienta med vrstami, s krepkim tiskom sta označena para z močnejšo korelacijo ($r > 0,5$)

| V | SORARA | SORMIN | SORALP | CLEGLA | ARVTER | MICAGR | MICSUB | APOFLA | MUSAVE |
|--------|-------------|--------|-------------|--------|--------|--------|--------|--------|--------|
| SORARA | 1 | | | | | | | | |
| SORMIN | 0,11 | 1 | | | | | | | |
| SORALP | 0,40 | -0,17 | 1 | | | | | | |
| CLEGLA | -0,16 | 0,05 | -0,30 | 1 | | | | | |
| ARVTER | -0,26 | -0,02 | -0,24 | -0,42 | 1 | | | | |
| MICAGR | -0,23 | 0,02 | -0,33 | 0,19 | -0,11 | 1 | | | |
| MICSUB | 0,30 | -0,31 | 0,64 | -0,02 | -0,21 | -0,29 | 1 | | |
| APOFLA | 0,52 | -0,29 | 0,23 | 0,26 | -0,32 | -0,45 | 0,41 | 1 | |
| MUSAVE | -0,19 | 0,19 | -0,17 | -0,08 | -0,24 | 0,18 | -0,30 | -0,09 | 1 |

R >

Korelacija med abundancami vrst, zaokrožena na dve decimalki, izvoz

```
corr <- round((cor(qab)), digits = 2)
```

```
write.table(corr, file = "corr.txt", row.names = TRUE)
```

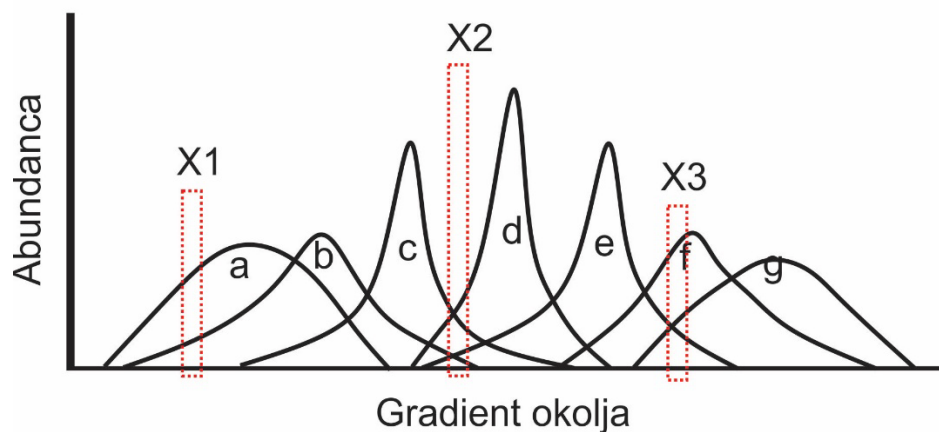


7.3 Problem dvojne ničle in paradoks abundance

Posamezna vrsta organizma ima specifično ekološko preferenco, to pomeni, da je razvila evolucijske adaptacije na specifične okoljske razmere, vključno s prilagoditvami na druge vrste organizmov. Pravimo, da ima ekološko nišo. Vrste praviloma najdemo na lokalitetah, ki izkazujejo razmere okolja v skladu z njihovimi ekološkimi preferencami. Iz teorije niš izhaja napoved, da imajo vrste unimodalno porazdelitev abundance vzdolž gradienta okolja (npr. v obliki Gaussove krivulje), z drugimi besedami, v optimalnem odseku gradienta, abundanca vrste močno naraste, proti ekstremom njene ekološke preference pa se zmanjšuje. Izven tolerančnega območja ekološke preference vrsta ni prisotna (Slika 7.2).

Ob popisu združbe vzdolž gradienta okolja podatkovna matrika, lokalitete krat vrste, lahko vsebuje številne ničle (0). Ilustrativen je spodnji primer vzorčenja vzdolž gradienta na treh lokalitetah (Slika 7.2). Skupni nabor šteje 7 vrst, od tega 5 vrst izkazuje dvojno ničlo (Preglednica 7.17).

Primerjave med lokalitetami oz. med združbami temeljijo na podatkih o abundanci vrst (ali prezenci). Če je vrsta prisotna na obeh primerjanih lokalitetah, je to indikator podobnih razmer med lokalitetama, saj so ekološke razmere za vrsto ugodne ali vsaj tolerantne. Če je vrsta prisotna samo na lokaliteti 1, na lokaliteti 2 pa ne, je to indikator različnih ekoloških razmer (seveda, če ne gre za napako v vzorčenju). V ekološki praksi obravnave več lokalitet ali združb in s širokim razponom gradientov okoljskih spremenljivk se srečujemo s primeri, ko je vrsta odsotna na dveh primerjanih lokalitetah. To je posledica za vrsto netolerantnih razmer za njeno preživetje oz. je lokaliteta izven njene ekološke niše. Vzroki za odsotnost vrste so lahko tudi v zasedenosti niše z drugo vrsto, disperziji vrste, naključnem lokalnem izumrtju, zgodovinskih dogodkih ali drugimi stohastičnimi procesi (glej Poglavlje 3.4 o ekološkem filtriranju).



Slika 7.2 Ekološka niša – stopnja prilagoditve posameznih vrst, vrste so označene s črkami (a-g). Širine krivulj opisujejo tolerančno območje posamezne vrste, njihove višine pa abundanco. Rdeči trakovi predstavljajo tri vzorčenja vrst vzdolž gradienta X1, X2 in X3; PA popis je v Preglednici 7.17.

Preglednica 7.17 PA podatki vzorčenja iz modela ekoloških niš na Sliki 7.2, z 00 so označene vrste, ki v medsebojnih primerjavah med lokalitetami izkazujejo dvojno ničlo

| <i>SU / V</i> | a - 00 | b | c - 00 | d - 00 | e | f - 00 | g - 00 |
|---------------|--------|---|--------|--------|---|--------|--------|
| X1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| X2 | 0 | 1 | 1 | 1 | 1 | 0 | 0 |
| X3 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |

Delež ničel v podatkovnem nizu o sestavi združbe na splošno narašča z variabilnostjo dejavnikov okolja med vzorčenimi lokalitetami. Če je vzorčenje izvedeno vzdolž enega ali celo več gradientov okolja, se sestava združbe bolj ali manj razlikuje od lokalitete do lokalitete, kar privede do večkratnega **pojava dvojnih ničel**. Nekritično vključevanje dvojnih ničel v izračun primerjav med lokalitetami lahko privede do višje vrednosti izračuna podobnosti med njimi. V mnogoterih primerjavah med lokalitetami (> 2) z več dvojnimi ničlami (včasih celo z več, kot je primerjav z abundanco) to navidezno (oz. neresnično) poveča podobnost med njimi. Zaključki, izpeljani iz takšnih analiz, niso ekološko utemeljeni, gre namreč za posledico računskih postopkov v analizi.

Dvojne ničle v analizi združbe niso informativne, njihova interpretacija generira **problem dvojnih ničel** (ang. Double-zero Problem), ki se mu moramo izogniti. Zato moramo biti pri izbiri koeficientov podobnosti oz. različnosti pozorni ali formula za izračun vključuje dvojne ničle ali jih izloči. Koeficiente, ki ne upoštevajo vrednosti dvojnih ničel, imenujemo **asimetrični koeficienti**, saj lokalitete z dvojno ničlo obravnavajo drugače – jih izločijo, kot lokalitete z dvojno prisotnostjo. Koeficiente, ki vrednosti z dvojno ničlo obravnavajo na enak način, kot vse druge vrednosti iz podatkovne matrike, imenujemo **simetrični koeficienti**. Rezultati so seveda posledično različni, zato matrike s popisi združb, ki vsebujejo dvojne ničle, analiziramo z asimetričnimi koeficienti. Poglejmo preprost primer s problemom dvojne ničle, ki pripelje do učinka, imenovanega **paradoks abundance**. Primerjamo abundance treh vrst V_1 , V_2 in V_3 med tremi združbami Z_1 , Z_2 in Z_3 , predstavljene v matriki:

| <i>Z / V</i> | V_1 | V_2 | V_3 |
|--------------|-------|-------|-------|
| Z_1 | 0 | 4 | 8 |
| Z_2 | 0 | 1 | 1 |
| Z_3 | 1 | 0 | 0 |

Izračunamo Evklidovo razdaljo med posameznimi združbami:

| <i>Z</i> | Z_1 | Z_2 | Z_3 |
|----------|-------|-------|-------|
| Z_1 | 0 | | |
| Z_2 | 7,616 | 0 | |
| Z_3 | 9,000 | 1,732 | 0 |

Evklidova razdalja med združbama Z_2 in Z_3 , ki nimata skupnih vrst, je krajša kot med združbama Z_1 in Z_2 , čeprav imata skupni vrsti V_2 in V_3 . S stališča ekološke interpretacije je to nekorekten izračun primerjav med združbama. Evklidska razdalja ni ustrezno merilo za primerjanje med združbami, ker so v izračun vključena dvojne ničle.

Asimetrični koeficienti, primerni za analizo ekoloških združb z dvojnimi ničlami, so: S_7 , S_8 , S_{17} , D_3 , D_4 , D_9 , D_{10} , D_{11} , D_{14} , D_{15} , D_{16} , D_{17} , D_{18} in D_{19} (Legender in Legendre 2012).

Ničla ali prazna celica

Pri popisu združbe zbiramo podatke o prezenci in/ali abundanci vrst, ki jih proučujemo. V biodiverzitetni praksi to pomeni, da oblikujemo seznam prisotnih vrst oz. seznam s številom osebkov posamezne vrste, ki sestavljajo združbo. Za predstavitev vrstnega bogastva je to dovolj. V analizah združbe, kjer z matematičnimi orodji primerjamo dve združbi in njun celoten nabor vrst, postanejo pomembne vrednosti, kjer vrste nismo potrdili oz. kjer prisotnost vrste ovrednotimo z 0 (vrsta ni prisotna) oz. kjer je abundanca vrste 0 (vrsta je zastopana z nič osebki). Pri oblikovanju matrike podatkov za namen analize združbe, kjer vrste nismo potrdila (PA podatki), vnesemo v celico vrednost 0, oz. kjer smo prešteli 0 osebkov (podatki o abundanci). Prazna celica brez navedbe vrednosti pomeni, da tam popisa nismo izvedli oz., da gre za manjkajoče podatke. Postopek analize in primerjanja združb se seveda razlikuje, če so podatki nepopolni oz. vsebuje celice z manjkajočimi podatki ali pa je vrednost v celici 0. Posledično so tudi zaključki drugačni.

7.4 Transformacije podatkov o sestavi združb

Porazdelitev vrednosti abundanc je lahko asimetrična. Z namenom zmanjšanja asimetrije v porazdelitvi lahko abundance y posamezne vrste **transformiramo** v y' . Transformacija poteka med vzorčnimi enotami. Najpogostejši transformaciji sta **korenjenje**:

$$y' = \sqrt{y}$$

in **logaritmiranje**. V primeru, ko matrika podatkov vsebuje vrednosti 0, $y = 0$ (log vrednosti 0 ne obstaja), vsem vrednosti prištejemo konstantno vrednost c , po navadi je to vrednost 1, $c = 1$:

- $y' = \log(y)$ kadar so vsi $y > 0$, oziroma
- $y' = \log(y+1)$ kadar v podatkovni matriki obstajajo vrednosti $y = 0$.

Za zmanjšanje absolutne razlike v abundanci med vzorci in ohranitvi variabilnost v relativni sestavi vrst med njimi uporabimo:

tetivno transformacijo (ang. Chord Transformation):

$$y'_{ij} = \frac{y_{ij}}{\sqrt{\sum_{j=1}^p y_{ij}^2}},$$

ali **Hellingerjevo transformacijo**:

$$y'_{ij} = \sqrt{\frac{y_{ij}}{y_{i+}}},$$

kjer je y_{ij} abundanca vrste i na lokaliteti j , y_{i+} je vsota abundanc vrste i na vseh lokalitetah.



PRIMER 7.9

Tetivna transformacija

Skupno število osebkov oz. abundance vrst se med lokalitetami lahko zelo razlikujejo. Absolutno večje vrednosti lahko zakrijejo variabilnost absolutno manjših vrednosti. V podatkovnem nizu malih sesalcev ima lokaliteta S4 samo 7 osebkov. Največ osebkov 44 ima lokaliteta P9 (Preglednica 7.18). Za odpravo tega nesorazmerja in ohranitev variabilnosti v odnosih med vrstami, abundance transformiramo po lokalitetah. Izračun koeficientov podobnosti ali razdalj naredimo na transformiranih vrednostih. Postopek tetivne transformacije za lokaliteto S1 je predstavljen v Preglednici 7.19.

Preglednica 7.18 Skupna vsota abundanc po lokalitetah

| <i>SU</i> | vsota |
|-----------|-------|
| S1 | 31 |
| S2 | 15 |
| S3 | 11 |
| S4 | 7 |
| S5 | 20 |
| S6 | 21 |
| S7 | 34 |
| P8 | 21 |
| P9 | 44 |
| P10 | 32 |
| P11 | 24 |
| P12 | 16 |
| P13 | 17 |
| P14 | 19 |

Preglednica 7.19 Postopek transformacije surovih vrednosti za lokaliteto S1 s tetivno transformacijo

| V | S1 | S1 ² | tetivna S1 |
|--------|----|-----------------|------------|
| SORARA | 4 | 16 | 0,961 |
| SORMIN | 6 | 36 | 2,163 |
| SORALP | 0 | 0 | 0,000 |
| CLEGLA | 12 | 144 | 8,652 |
| ARVTER | 0 | 0 | 0,000 |
| MICAGR | 0 | 0 | 0,000 |
| MICSUB | 0 | 0 | 0,000 |
| APOFLA | 9 | 81 | 4,867 |
| MUSAVE | 0 | 0 | 0,000 |
| vsota | | 277 | |
| koren | | 16,643 | |

```
R > package = »vegan«
```

```
#Standardizacija s tetivno transformacijo in izračun Bray-Curtis razdalje
```

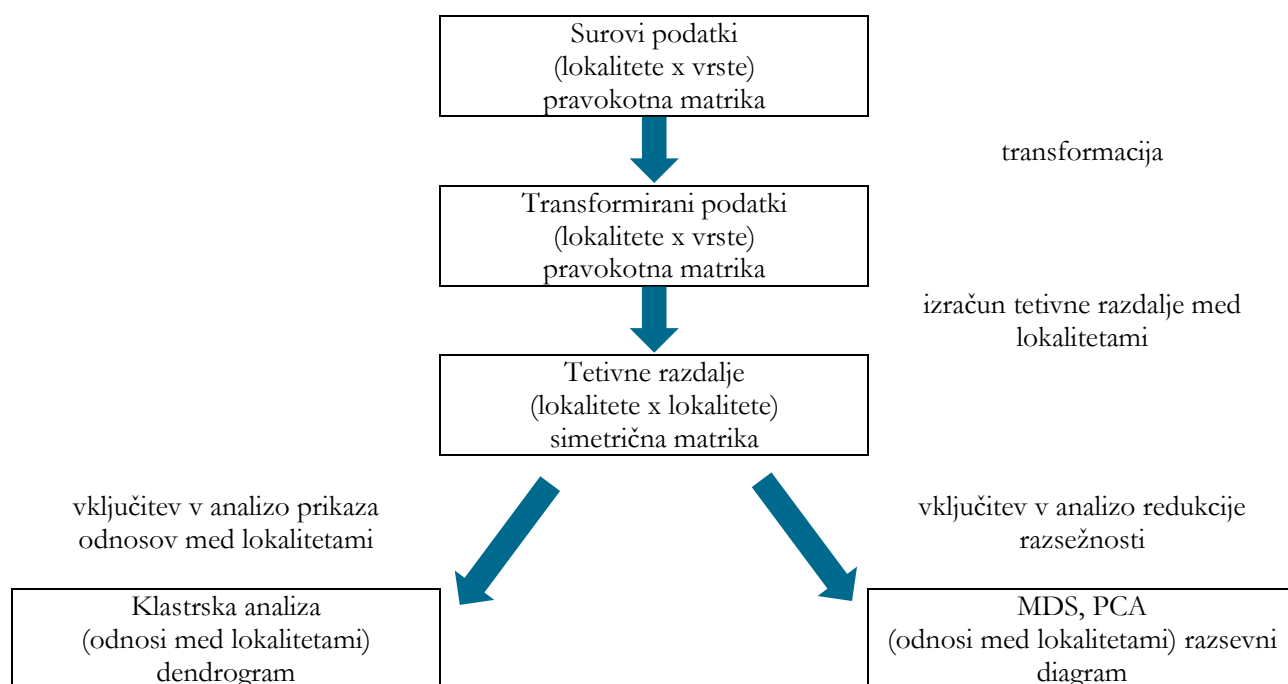
```
bray_st<-vegdist(decostand(qab, "norm"), "bray")
```

```
bray_st
```



7.5 Predstavitev odnosov med združbami

Temeljni izziv pri predstavljanju odnosov med združbami je zmanjšanje dimenzionalnosti oz. poenostavitev (redukcija) multivariatnega podatkovnega niza. Cilj analize je grafični izdelek, s katerim na nazoren in enostaven način predstavimo odnose med združbami v dvodimenzionalnem prostoru. V tem poglavju sta predstavljena dva metodološka pristopa za prikaz odnosov med združbami. Oba temeljita na izračunu razdalj med združbami, to sta klastrska analiza in ordinacija. Iz družine klasterskih metod je predstavljena **hierarhična klastrska analiza HCA** (sin. hierarhično združevanje v skupine; ang. Hierarchical Cluster Analysis). Iz družine ordinacijskih metod pa **multidimenzionalno skaliranje MDS** (ang. Multidimensional Scaling). Obe metodi imata skupni začetni del izračuna, ki temelji na izdelavi simetrične matrike razdalj (predstavljeno v Poglavju 7.2). Rezultat klastrske analize je **dendrogram**, s katerim vizualiziramo odnose med združbami v obliki drevesa. Dendrogram interpretiramo, kot da so si združbe z bližjimi razvejitvami bolj in združbe z bolj oddaljenimi razvejitvami manj podobne. Multidimenzionalno skaliranje je ordinacijska metoda, s katero objekte razporedimo v dvodimenzionalnem prostoru. Rezultat je **razsevni diagram** (ang. Bi-plot), s katerim vizualiziramo odnose med združbami. Bližnje združbe na biplotu so si bolj podobne, bolj oddaljene pa manj. Proces analize od pravokotne matrike podatkov do dendrograma oz. diagrama je predstavljen na Sliki 7.3.

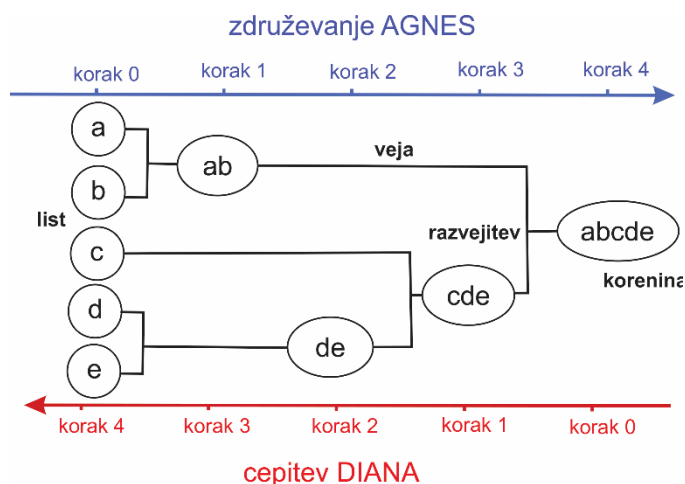


Slika 7.3 Diagram procesa Q analize ekološke združbe s klasterško in ordinacijsko metodo.

7.5.1 Razvrščanje v skupine

Hierarhično združevanje v skupine (sin. hierarhična klasterška analiza, klasterška analiza; ang. Hierarchical Cluster Analysis) je multivariatni statistični pristop za prepoznavanje, bodisi dovolj podobnih objektov, da jih lahko razvrstimo v isto skupino, bodisi dovolj različnih objektov, da jih lahko razvrstimo v različne skupine (Jesenko in Jesenko 2007, Legendre in Legendre 2012).

Razlikujemo dva osnovna tipa razvrščanja: **cepitev** (ang. Divisive Hierarchical Clustering, DIANA – Divise Analysis) in **združevanje** (ang. Agglomerative Clustering, AGNES – Agglomerative nesting). Združevanje oz. AGNES algoritem poteka od spodaj navzgor. Vsak objekt, ki je samostojna enota (list; ang. Leaf) primerjamo z drugimi objekti in jih na podlagi kriterija združujemo v skupine (v razvejitvah; ang. Nodes). Proces ponavljamo, dokler ne združimo vseh enot v eno skupino (korenina; ang. Root). Algoritem DIANA ima nasprotni potek. Analiza se prične pri korenini, ki vključuje vse enote. Na podlagi kriterija skupino cepimo v vse manjše skupine do razcepitve na posamezne liste. Končni rezultat hierarhične analize je grafična predstavitev odnosov med objekti v obliki drevesa (sin. dendrograma; ang. Tree) (Slika 7.4).



Slika 7.4 Shema metodološkega pristopa izgradnje drevesa. Z združevanjem – AGNES od spodaj navzgor, oz. s cepitvijo – DIANA, od zgoraj navzdol (Prirejeno iz Hierarchical Cluster Analysis, 2020).

V tem poglavju so predstavljene samo metode iz skupine združevanja skupin – **AGNES pristop**. Metoda DIANA je predstavljena v Poglavju 12 Numerična taksonomija in Poglavju 13.

Velik izziv za predstavitev odnosov med ekološkimi združbami, s pomočjo klastrske analize predstavlja izbor oz. računanje kriterijev za združevanje skupin. Razvitih je pet osnovnih pristopov: minimalna, maksimalna, povprečna, centroidna in Wardova metoda.

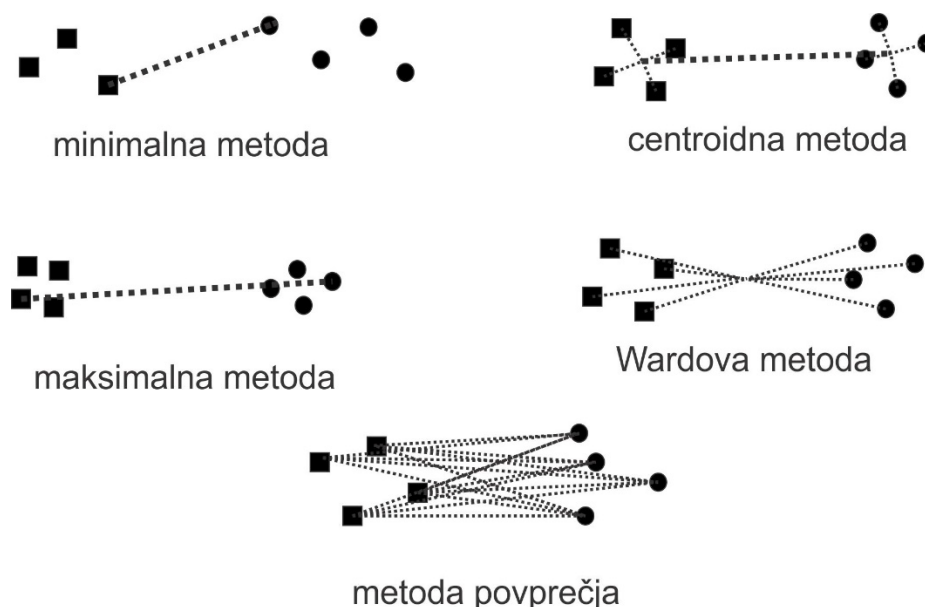
Minimalna metoda ali enojno združevanje v skupine (ang. Minimum or Single Linkage Clustering). Združevanje med dvema skupinama oblikujemo po kriteriju najmanjše razdalje (minimalna vrednost) med elementi skupine 1 in elementi skupine 2 (Slika 7.5). Metoda je učinkovita za dolge razpotegnjene vzorce.

Maksimalna metoda ali polno združevanje v skupine (ang. Maximum or Complete Linkage Clustering). Združevanje med dvema skupinama oblikujemo po kriteriju največje razdalje (maksimalna vrednost) med elementi skupine 1 in elementi skupine 2 (Slika 7.5). Metoda je učinkovita za zbite kompaktne skupine.

Metoda povprečja ali povprečno združevanje v skupine (ang. Mean or Average Linkage Clustering). Združevanje med dvema skupinama oblikujemo po kriteriju povprečne razdalje. Med elementi skupine 1 in elementi skupine 2 izračunamo vse možne razdalje in iz njih izračunamo povprečno razdaljo med elementi skupine 1 in elementi skupine 2 (Slika 7.5). Metoda UPGMA (ang. Unweighted Arithmetic Average) je izpeljanka te skupine analiz razdalj med skupinami. Uporabna je predvsem pri analizi sistematičnega vzorčenja.

Centroidna metoda (ang. Centroid Linkage Clustering). Združevanje med dvema skupinama oblikujemo po kriteriju razdalje med centroidoma. Izračunamo centroid skupine 1 in centroid skupine 2 in nato razdaljo med centroidoma (Slika 7.5).

Wardova metoda (ang. Ward's Minimum Variance Method). Združevanje med dvema skupinama oblikujemo po kriteriju varianc. Izračunamo varianci skupin, skupini z najmanjšo razliko med variancama se združita v skupine (Slika 7.5). Metoda je učinkovita za eliptično oblikovanje skupine.



Slika 7.5 Shematski prikaz kriterija združevanja skupin po različnih metodah: minimalna metoda, maksimalna metoda, metoda povprečja, centroidna metoda in Wardova metoda.

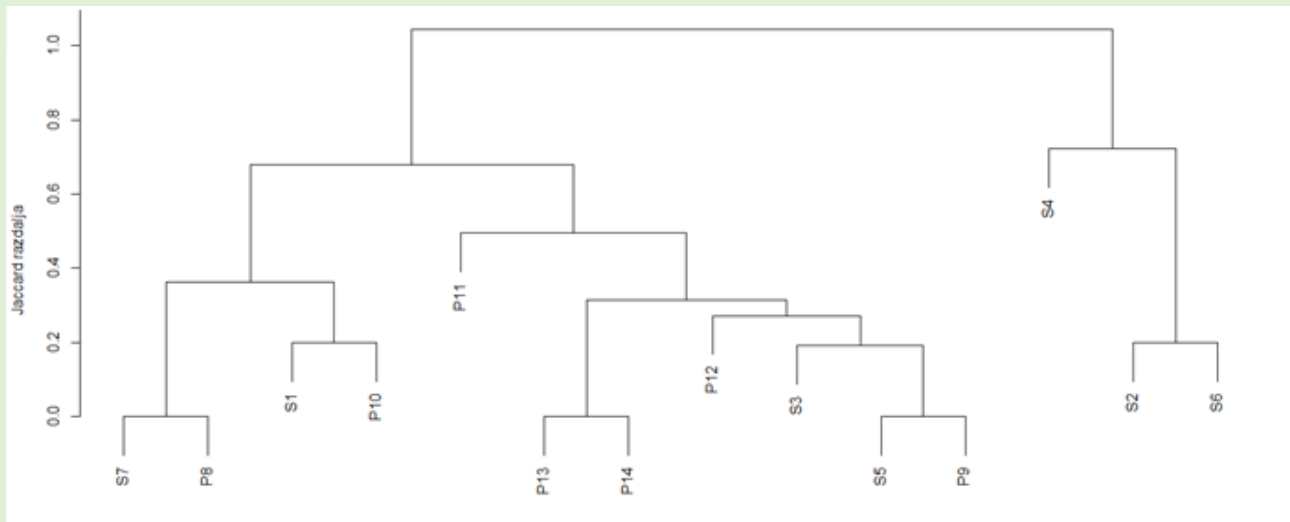
Analizo združevanja v skupine izdelamo v programskem okolju **R** s paketom »stat« oz. »vegan«, in ukazom `hclust()`, kriterijske metode za združevanje: "ward.D", "ward.D2", "single", "complete", "average" (= UPGMA), "mcquitty" (= WPGMA), "median" (= WPGMC) in "centroid" (= UPGMC).



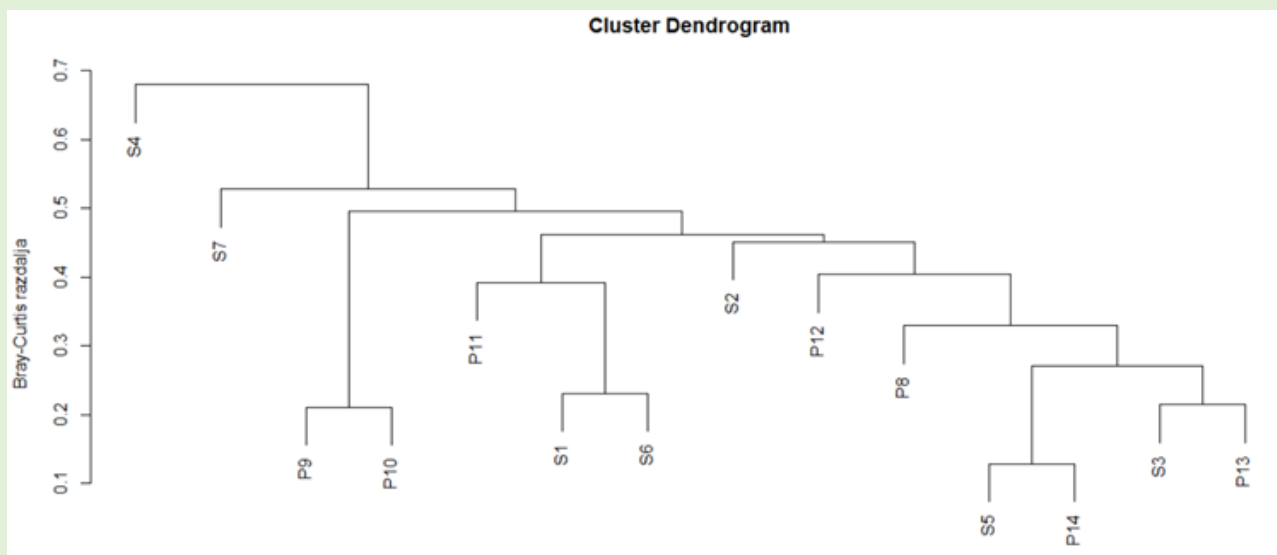
PRIMER 7.10

Klastri prikaz odnosov med združbami

Odnose med 14 združbami malih sesalcev na Smrekovcu in Peci vizualiziramo z drevesom. Vhodni podatki za izgradnjo drevesa so vrednosti simetrične matrike razdalj med lokalitetami. V Poglavju 7.3 je predstavljena plejada pristopov za izračun razdalj. Predstavljamo pristop izrisa drevesa za PA podatke in za abundance. Matriko PA vrednosti vključimo v izračun Jaccardovih koeficientov in prikažemo kot drevo po Ward metodi (Slika 7.6). Abundance vključimo v izračun Bray-Curisove razdalje in jih prikažemo v obliki UPGMA drevesa (Slika 7.7).



Slika 7.6 Wardov dendrogram odnosov med združbami na Jaccardovih razdaljah.



Slika 7.7 UPGMA dendrogram odnosov med združbami na Bray-Curtisovih razdaljah.

```

R > package = »vegan«
# Izračun drevesa, Jacc, Ward
jacc.clust.ward<- hclust(jacc, "ward.D2")
# Izris drevesa v paketu stats
plot(jacc.clust.ward, ylab = "Jaccard razdalja")
# Izračun drevesa, Bray, UPGMA
bray.clust <- hclust(bray, method = "average")
# Izris drevesa v paketu vegan
plot(bray.clust, ylab = "Bray-Curtis razdalja")

```



7.5.2 Multidimenzionalno skaliranje

V analizah ekoloških združb si zastavljamo vprašanja, kako podobne oz. kako različne so si združbe, ter kako se spreminja sestava vrst in relativni odnosi med vrstami. Primerjave med združbami izvajamo na podatkih o vrstnem bogastvu in vrstni pestrosti.

Pri strokovnem in raziskovalnem delu na področju analize ekoloških združb uporabljamo ordinacijske metode za redukcijo statističnega prostora in za študij relativnih odnosov med združbami. Temeljni cilj ordinacijskih metod je redukcija oz. poenostavitev večrazsežnega prostora (ang. Multidimensional Space) in reprezentativna ohranitev relativnih odnosov med objekti v dvodimenzionalnem prostoru.

V družini ordinacijskih metod razlikujemo štiri tipe redukcije prostora: PCA, CA, PCoA in MDS (Preglednica 7.20). Analiza glavnih komponent PCA (ang. Principal Component Analysis) je zelo močna ordinacijska metoda za redukcijo prostora. Metoda je omejena na kvantitativne podatke in je zelo občutljiva na dvojne ničle (glej Poglavje 7.3), deluje na principu Evklidovih razdalj. Korespondenčna analiza CA (ang. Correspondence Analysis) je primerna za analizo PA podatkov in abundanc o vrstah, deluje na principu X^2 razdalje. Analiza glavnih koordinat PCoA (ang. Principal Coordinate Analysis) je primerna za analizo kvantitativnih, kvalitativnih in mešanih podatkov, deluje na vseh tipih razdalj. Multidimenzionalno skaliranje MDS (ang. Multidimensional Scaling) je primerna za analizo kvantitativnih, kvalitativnih in mešanih podatkov, deluje na vseh tipih razdalj.

V tem poglavju je podrobneje predstavljena metoda **multidimenzionalnega skaliranja**. Metoda omogoča dobro interpretacijo primerjav med združbami na podatkih o abundancah. Cilj metode je redukcija multidimenzionalnih informacij s transformacijo večrazsežne matrike, lokalitet krat vrst v samo dve ali tri osi, vzdolž katerih so razporejene združbe. Termin »multidimenzionalnosti« (večrazsežnosti) se nanaša na istočasno obravnavo več lastnosti oz. spremenljivk. V Q analizi združbe so spremenljivke vrste organizmov, v matematičnem kontekstu to pomeni, da ima prostor toliko razsežnosti, kolikor je vrst v matriki. V dvodimenzionalnem prostoru bolj podobne združbe ležijo bližje skupaj, bolj različne združbe pa bolj narazen. Na podlagi razporeditve združb sklepamo na podobnost v sestavi združb oz. prepoznavamo vzorce podobnih združb, ali združb iz podobnih okolij, ali gradacije združb vzdolž gradienta okolja, ali združb iz istih geografskih območij. Prednost metode MDS pred drugimi ordinacijskimi metodami je v izjemni fleksibilnosti obdelave različnih kategorij podatkov.

Premestitev točk iz multidimenzionalnega v dvodimenzionalen prostor in ohranitev informacije o odnosih med združbami merimo z **analizo stresa** (ang. Stress). Rezultate stresa predstavimo grafično in številčno. Grafično s **Shepardovim diagramom**, to je regresijski model med razdaljami v reduciranem prostoru (ang. Distance in Reduced Space,

Ordination Distances) in razdaljami v multidimenzionalnem prostoru (ang. Distance in Multidimensional Space, Observed Disimilarity). Na grafu je prikazan razsevni diagram vseh točk iz matrike lokalitet krat vrst in regresijska krivulja med obema skupinama razdalj. Regresijska krivulja je zgrajena po stopničasti metodi, zato ima obliko stopnic; število stopnic je enako številu primerjav. Uspešnost regresijskega modela izkazuje **učinek velikosti** (ang. Effect Size), ki je podan z **determinacijskim koeficientom R^2** , to je z deležem razložene variance. Za še sprejemljivo analizo mora biti vrednost R^2 nad 0,6. Številčno je uspešnost redukcije informacije podana s koeficientom **stresa**. Nižji kot je koeficient stresa, uspešnejša je analiza. Mejne vrednosti stresa za oceno analize so: < 0,05 odlična redukcija večdimenzionalnosti, < 0,1 zelo dobro, < 0,2 dobro, < 0,3 slaba redukcija informacije. Povezavo med grafično in številčno oceno razberemo iz položaja točk; bližje kot so točke regresijski krivulji, učinkovitejša je metoda; višja je vrednost R^2 in nižji je koeficient stresa.

Glede na tip vhodnih podatkov razlikujemo naslednje kategorije MDS: **klasično multidimenzionalno skaliranje MDS** (ang. Classical Multidimensional Scaling) – nekateri avtorji jo obravnavajo kot samostojno ordinacijsko metodo bolj znano pod imenom **analiza glavnih koordinat PCoA** (ang. Principal Coordinates Analysis; glej Poglavje 5 Funkcionalna pestrost). Vhodni podatki so lahko intervalne, nominalne ali kombinirane vrednosti. **Metrično multidimenzionalno skaliranje mMDS** (ang. Metric Multidimensional Scaling) je optimizirana oblika MDS. **Ne-metrično multidimenzionalno skaliranje nMDS** (ang. Non-metric Multidimensional Scaling) vhodni podatki so kategorične vrednosti.

Postopek računanja MDS poteka v več korakih:

1. Oblikovane hipoteze/cilja. Katere spremenljivke vključimo, koliko jih je in kakšne lastnosti imajo? Definiranje namena, kaj želimo predstaviti.
2. Pridobivanje in organizacija podatkov v matriko dimenzije $m \times n$ združbe x vrste.
3. Izvedba MDS statističnega izračuna. Naprednejši statistični paketi imajo modul za izračun MDS. Več rešitev ponuja programsko okolje R. Osnovna odločitev je med metrično MDS (intervalni ali razmernostni podatki) ali ne-metrično MDS (nominalni, ordinalni ali mešani tip podatkov).
4. Odločitev o številu dimenzij. Osnovni namen je redukcija dimenzionalnosti. Raziskovalčeva odločitev je, koliko dimenzionalen izid želi. Praviloma se odločimo za dvodimenzionalen rezultat, saj je takšen rezultat enostavno predstaviti grafično in enostavno interpretirati.
5. Grafična predstavitev rezultata. Praviloma računalniški programi kot končni rezultat izdelajo razsevni diagram.
6. Preverjanje zanesljivosti in vrednotenje rezultatov. Izračunamo R^2 za določitev razložene variance. Za zanesljivi rezultat MDS ordinacije velja $R^2 > 0,6$. Drugi

pristop je izračun in analiza stresa. Velja: $< 0,05$ odlična, $< 0,1$ zelo dobro, $< 0,2$ dobro, $< 0,3$ slaba redukcija informacije.

7. Celovita predstavitev rezultatov. Med rezultati navedemo in interpretiramo: razsevni diagram, simetrično matriko razdalj ter vrednost determinacijskega koeficienta R^2 in stresa.

Preglednica 7.21 Pregled ordinacijskih metod

| Metoda | Razdalja | Spremenljivke |
|--|--------------------|---|
| Analiza glavnih komponent PCA | Evklidova razdalja | intervalni podatki (grožnja so dvojne ničle) |
| Korespondenčna analiza CA | X^2 razdalja | intervalni podatki, binarni podatki, abundance |
| Analiza glavnih koordinat PCoA (Sin. Klasično multidimenzionalno skaliranje) | različne razdalje | intervalni podatki, nominalni podatki, mešani podatki |
| Multidimenzionalno skaliranje MSD | različne razdalje | intervalni podatki, nominalni podatki, mešani podatki |

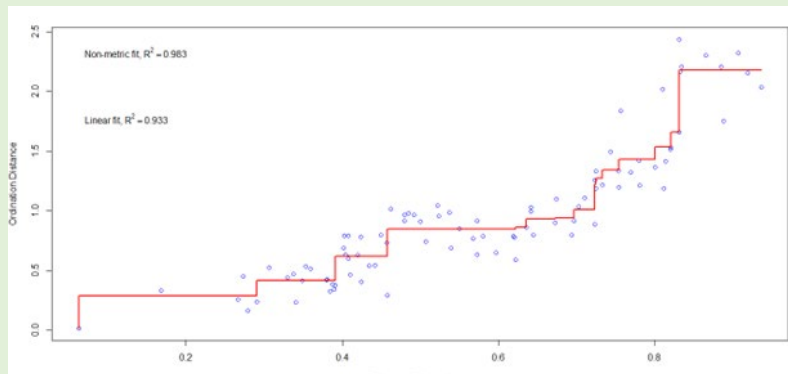


PRIMER 7.11

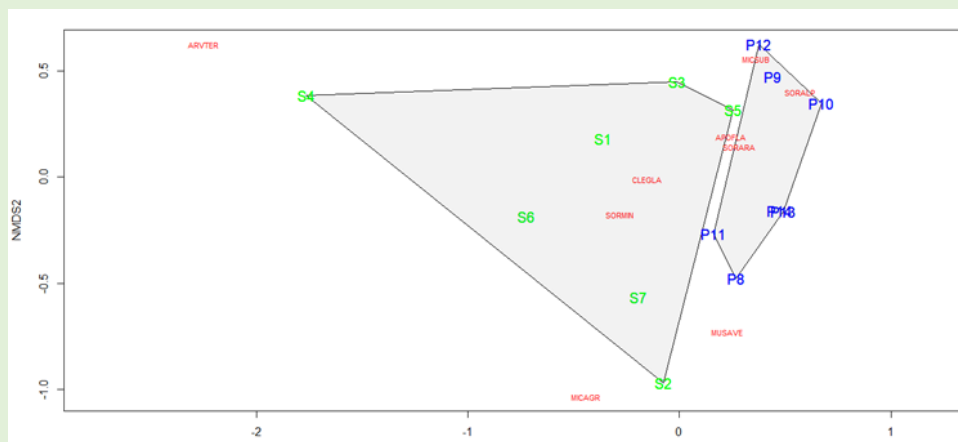
Multidimenzionalno skaliranje

Vhodni podatki v multidimenzionalnem skaliranju so vrednosti abundanc iz osnovne matrike lokalitete x vrste dimenzije 14×9 . Programska koda izvede izračun razdalj med lokalitetami in vrednosti vključi v izris razsevnega diagrama in izračun stresa. Mera podobnosti so Bray-Curtisove razdalje. Vrednost stresa v analizi združb malih sesalcev na Smrekovcu in Peci ima vrednost 0,129, kar je dobra redukcija dimenzionalnosti. Delež razložene variance znaša $R^2 = 0,983$, z drugimi besedami, reducirani model razloži 98,3 % variabilnosti osnovne večrazsežne matrike. Odnos v regresijskem modelu je predstavljen na Shepardovem diagramu (Slika 7.8).

Minimalna konveksna poligona brez prekrivanja povežeta lokalitete obeh gora (Slika 7.9). Geografska vzorca sta ločena vzdolž NMDS1. V območju negativnih vrednosti so razporejeni vzorci iz Smrekovca, močnejše jih opredeljujejo vrednosti vrst SORMIN in CLEGLA. Vrsta ARVTER je edinstvena. Registrirali smo samo en osebek. Močno zaznamuje lokaliteto S4. Združbe iz Pece so bolj homogene in ne izkazujejo tako izrazite variabilnosti v sestavi malih sesalcev, kot je primer na Smrekovcu. To vidimo po bolj kompaktni razporeditvi skupine in velikost njenega poligona je bistveno manjša v primerjavi z velikim poligonom Smrekovca.



Slika 7.8 Shepardov diagram, regresijski model stresa, s katerim ocenimo uspešnost multivariatne metode redukcije večdimenzionalnosti.



Slika 7.9 Razsevni diagram 14 lokalitet oz. združb malih sesalcev na prvi dve osi multidimenzionalnega skaliranja (NMDS). Lokalitete Smrekovca so zeleno označene in Pece modro. Lokalitete iste geografske regije so povezane z minimalnim konveksnim poligonom. Položaj vrst je v rdeči barvi.

```
R > package = »vegan«
```

```
# MDS na surovih abundancah, data=qab, d=Bray-Curtis
```

```
qab.mds <- metaMDS(qab, k=2)
```

```
qab.mds$stress
```

```
stressplot(qab.mds)
```

```
# Biplot
```

```
treat=c(rep("Treatment1", 7), rep("Treatment2", 7))
```

```
ordiplot(qab.mds, type="n")
```

```
# Dodani poligoni za dve skupini
```

```
ordihull(qab.mds, groups=treat, draw="polygon", col="grey90", label=F)
```

```
#Dodajanje vrst
```

```
orditorp(qab.mds, display="species", col="red", air=0.01)
```

```
#Dodajanje lokalitet, barvno označeni dve skupini
```

```
orditorp(qab.mds, display="sites", col=c(rep("green", 7), rep("blue", 7)),  
air=0.01, cex=1.25)
```



Zaključki

- Analizo ekološke združbe izvajamo zaradi opisa združbe in analize variabilnosti združb, primerjav med združbami ali odziva združb vzdolž ekološkega gradienta.
- Združbe med seboj primerjamo na podatkih o prezenci in abundanci vrst. Podatkovni niz tvori pravokotna matrika, združbe (ali lokalitete) so vrstice in vrste organizmov so stolpci (lokalitete \times vrste). Vrednosti v poljih so lahko 0 ali 1 v primeru prezenca ali število osebkov – abundanca vrste ali deleži osebkov ali rangi.
- Združbo analiziramo z vidika lokalitet Q analiza in z vidika vrst R analiza. Primerjamo jih z izračunom podobnosti S ali različnosti D . Razlika med dvema združbama je podana z razdaljo d .
- Dve združbi z enakim naborom vrst in z enako številčnostjo izkazujeta največjo podobnost. Koeficient podobnosti ima vrednost 1, razdalja med njima je 0. Z večanjem razlik v vrstni sestavi in/ali abundanci se koeficient podobnosti približuje vrednosti 0. Razdalja med združbama se večja.
- Mere za izračun razdalj med združbami so vgrajene v programsko okolje **R**, paket `vegan`, ukaz `vegdist()`, metode: `manhattan`, `euclidean`, `canberra`, `clark`, `bray`, `kulczynski`, `jaccard`, `gower`, `altGower`, `morisita`, `horn`, `mountford`, `raup`, `binomial`, `chao`, `cao` in `mahalanobis`.
- Teorija nič predvideva ekološko preferenco za posamezno vrst. Iz tega sklepamo na unimodalno porazdelitev abundance vzdolž okoljskega gradienta. Izven tolerančnega območja preference vrsta ni prisotna, zato matrika podatkov združbe \times vrste vsebuje ničle (0). Delež ničel v podatkovnem nizu narašča z višanjem variabilnosti dejavnikov okolja. Odsotnost vrste na dveh lokalitetah privede do pojava dvojne ničle.
- Dvojne ničle v analizi združbe niso informativne, vplivajo na izračun koeficientov podobnosti in računsko višajo podobnost, zato je pomembna izbira koeficienta glede na strukturo podatkov in namen analize.
- Asimetrični koeficienti ne upoštevajo vrednosti dvojnih ničel. Simetrični koeficienti vrednosti z dvojno ničlo obravnavajo na enak način, kot vse druge vrednosti. Problem dvojnih ničel pripeljejo do paradoksa abundance, kjer so združbe brez skupnih vrst bolj podobne kot združbe s skupnimi vrstami z različno abundanco.
- Porazdelitev abundance vrst ni simetrična. Število osebkov med združbami je lahko zelo različno. Z namenom zmanjšanja asimetrije abundance vrednosti transformiramo s korenjenjem ali logaritmiranjem. Vpliv absolutnih razlik v abundanci odpravimo s tetivno transformacijo ali Hellingerjevo transformacijo, ki ohranita variabilnost v relativni sestavi vrst.
- Odnose med združbami predstavimo s hierarhično klastersko analizo in multidimenzionalnim skaliranjem. Obe metodi temeljita na simetrični matriki razdalj med združbami.

- Rezultat klasične analize je dendrogram, s katerim vizualiziramo odnose med združbami v obliki drevesa.
- Multidimenzionalno skaliranje reducira večdimenzionalni prostor v dve dimenziji. Zruške razporedi v dvodimenzionalnem prostoru.
- Rezultate po klasični analizi in multidimenzionalnem skaliranju interpretiramo, kot da so združbe, ki so na razsevnom diagramu ali na drevesu bližje skupaj tudi bolj podobne oz. imajo podobno sestavo.

Literatura

- Hierarchical Cluster Analysis. 2020. UC Business Analytics R Programming Guide. Pridobljeno dne 5. 5. 2020, iz https://uc-r.github.io/hc_clustering.
- Janžekovič F., Čas M. 2001. Mali sesalci v gozdni krajini in pestrost njihove združbe v alpskem gozdu na Smrekovcu in Peci. *Gozdarski vestnik*, 59 (7–8), 322–327.
- Jesenko J., Jesenko M. 2007. Multivariatne statistične metode. Moderna organizacija. Kranj.
- Kotar M. 2011. Raziskovalne metode v upravljanju z gozdnimi ekosistemi. ZGDS. Ljubljana.
- Legendre P., L. Legendre. 2012. *Numerical Ecology*. Elsevier. Amsterdam.
- Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., Wagner H. 2019. vegan: Community Ecology Package. R package version 2.5-6. <https://CRAN.R-project.org/package=vegan>.



ODNOS MED VRSTNIM BOGASTVOM IN POVRŠINO

Povzetek

8. poglavje: Odnos med vrstnim bogastvom in površino

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

Odnos med vrstnim bogastvom in površino ozemlja SAR temelji na ugotovitvi, da večja območja omogočajo evolucijo in preživetje večjemu številu vrst. Odnos med vrstnim bogastvom in površino ima praviloma potenčno obliko. Primerjave izvajamo na surovih ali logaritmsko transformiranih podatkih. V log-log analizi jih zapišemo v obliki linearne regresijske enačbe $\log S = \log c + z \log A$. Smerni koeficient z opiše strmino krivulje, po kateri narašča število vrst S ob spreminjanju površine A . Pri majhnih površinah (lokalne razmere) je vrednost z visoka, $z \geq 0,25$, ob že relativno majhni spremembi površine lahko pride do velike spremembe v številu vrst. Pri srednje velikih površinah (regionalne razmere) je vrednost z majhna, $z \approx 0,02$ in tudi ob velikih spremembah površine je sprememba v številu vrst majhna. Na površinah velikosti kontinentov (globalne razmere) je vrednost z srednja, $z \approx 0,12$. Pri majhnih površinah se lahko pojavi učinek malega otoka, to je odstopanje od pričakovanega odnosa med vrstnim bogastvom in velikostjo površine. V takem primeru je vrstno bogastvo manjše in neodvisno od površine, dokler površina ne doseže določenega praga velikosti, oz. je stopnja rasti vrstnega bogastva manjša, kot je značilna za večja območja.

Potenčni oz. linearni model ne razloži vedno najboljše odnosa med številom vrst in površino območja. V literaturi je opisanih več kot 20 modelov SAR odnosa. Na primerih je prikazana uporaba R paketa 'sars', ki izračuna vse modele in med njih poišče najbolj priljegajočega. Izračun ob regresijskih parametrih vključuje determinacijski koeficient R^2 , Akaikov informacijski kriterij AIC ter izris regresijske krivulje, informacijo o obliki regresijske krivulje (linearna, konveksna ali sigmoidna) in tudi ali krivulja doseže asimptoto.

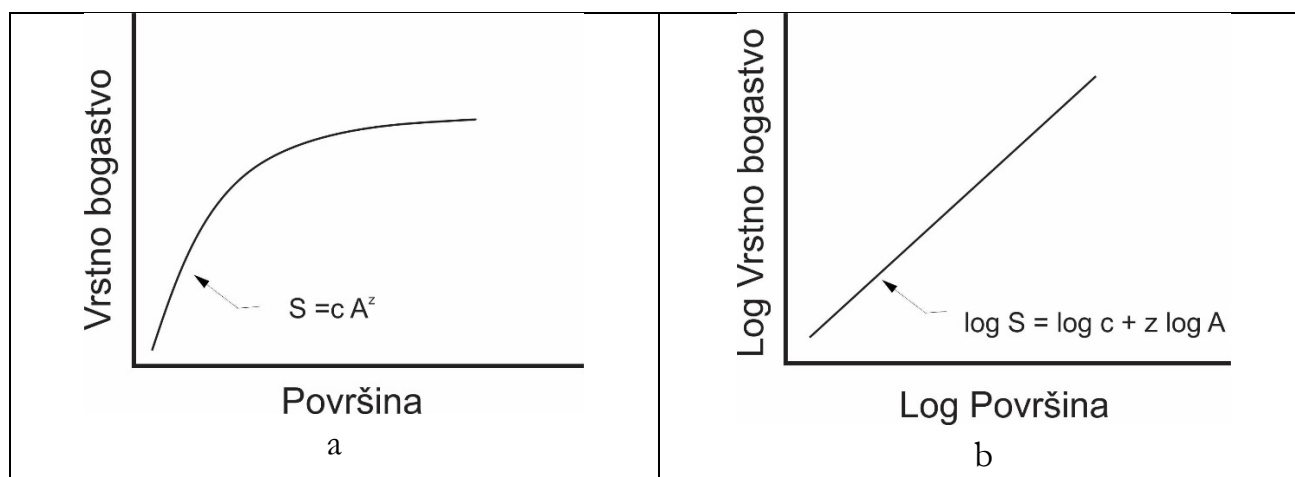
Odnos med vrstnim bogastvom in površino ozemlja, zapisan v obliki regresijskega modela, lahko uporabimo za napovedovanje števila vrst še neraziskanega ozemlja z znano površino. Druga oblika uporabe regresijskega modela je v primerjavi dejansko izmerjenega vrstnega bogastva s pričakovano vrednostjo. Iz tega lahko izpeljemo zaključek, ali je območje bogatejše ali revnejše od pričakovanja. Vrednost z iz regresijske modela je cenilka, s katero primerjamo bogastvo različnih taksonomskih skupin, ali pa z njim primerjamo vrstno bogastvo različnih območij.

Odnos med vrstnim bogastvom in površino ozemlja SAR (ang. Species-area Relationship) opisuje vzorec v katerem narašča število vrst ob večanju površine območja. Odnos temelji na zakonitosti, da večja območja omogočajo preživetje in evolucijo večjemu številu vrst. Najbolj osnoven in univerzalen zapis odnosa med vrstnim bogastvom in površino je definiran z Arrheniusovo enačbo:

$$S = cA^z ,$$

kjer je S število vrst, A površina ozemlja, c presečišče regresijske premice in Y osi ter z smerni koeficient regresijske premice. Koeficienta c in z izračunamo iz podatkov s pomočjo regresijske analize (Slika 8.1). Na logaritemski skali (log-log odnos) sta število vrst in površina v linearnem odnosu (Slika 8.1), pri tem je vrednost c začetna vrednost Y osi, z pa naklon premice:

$$\log S = \log c + z \log A .$$

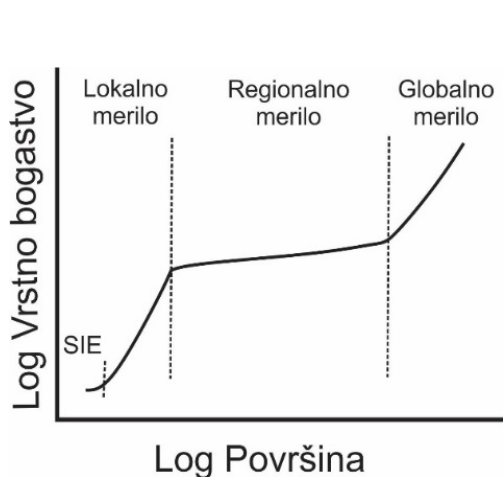


Slika 8.1 Odnos med površino A in vrstnim bogastvom S na surovih (a) in logaritemskih (b) vrednostih.

Pri obravnavi celotnega razpona merilne skale, od majhnih površin (nekaj m^2) do globalnega obsega (površina kontinenta) ima krivulja odnosa SAR sigmoidno obliko črke S (Slika 8.2). Na lokalnem merilu (manjše površine) je krivulja strma, na regionalni skali (srednje velike površinah) je položna in na globalnem merilu (velike površine) spet strma. V obravnavi malih površin je lahko prisoten tudi učinek malega otoka. Kot lokalne površine v tem poglavju obravnavamo jadranske otoke in kot regionalne površine evropske države in kot globalne kontinente.

Vsebina poglavja je narejena po temeljni literaturi Brown (1995), Rosenzweig (1995) in Triantis in sod. (2021).

Učinek malega otoka SIE (ang. Small-island Effect). Pri obravnavi območij z malimi površinami prihaja do odstopanja od pričakovanega odnosa med vrstnim bogastvom in velikostjo površine. To odstopanje se imenuje učinek malega otoka (Slika 8.2). Vrstno bogastvo na malih območjih je lahko manjše in neodvisno od površine, dokler površina ne doseže določenega praga velikosti, oz. je stopnja rasti vrstnega bogastva na manjših površinah nižja, kot je značilna za večja območja (Gao in Wang 2021). Učinek malega otoka razložimo s hipotezama: 1. hipoteza pestrosti habitatov – na majhnih površinah praviloma ni večjih in raznolikih habitatov. 2. otoška hipoteza – v skladu z otoškim učinkom je vrstno bogastvo opredeljeno z ravnovesjem med oddaljenostjo otoka od celine in velikostjo otoka. Prepoznali smo več dejavnikov, ki vplivajo na prag velikosti območja, pod katerim vrstno bogastvo neodvisno variira: 1. tip območja; 2. taksonomska skupina; 3. velikost območja.



Slika 8.2 Odnos med številom vrst in površino ozemlja na majhnem, lokalnem, regionalnem in globalnem merilu; SIE – učinek malega otoka.



Slika 8.3 Diagram odnosov med velikostjo otoka, njegovo oddaljenostjo od celine ter stopnjo priseljevanja in izumiranja. Presečišča krivulj (označena z A, B, C in D) označujejo ravnovesje v številu vrst kot posledico kombinacije dejavnikov.

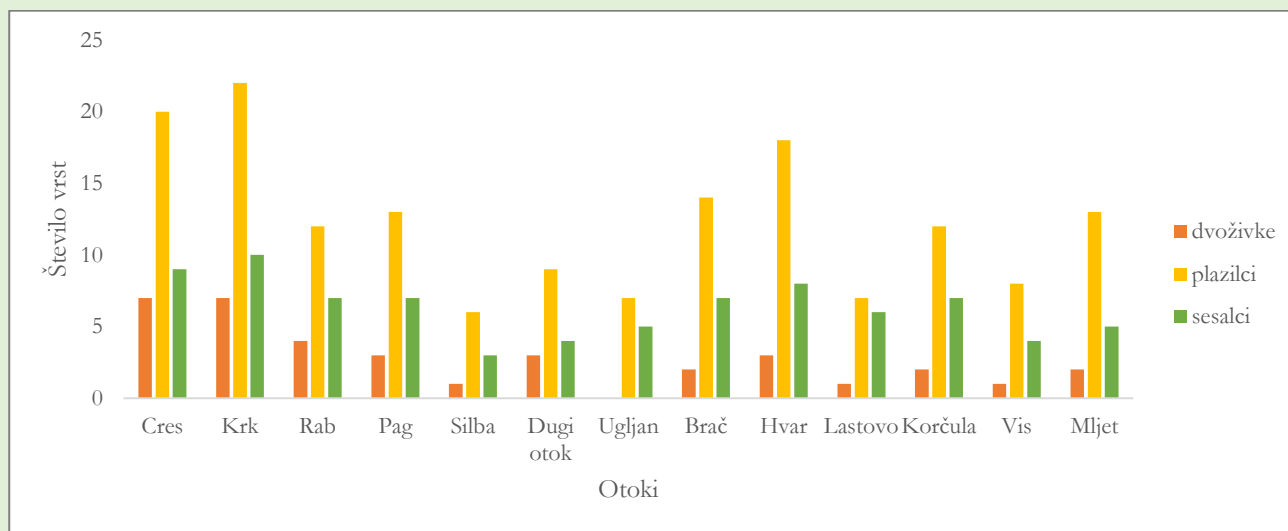
Odnos med številom vrst in površino ozemlja SAR izhaja iz **otoške diverzitete**. Število vrst na otoku je odvisno od ravnovesja (ang. Equilibrium) dejavnikov iz dveh skupin: geografskih in evlucijskih. Geografska dejavnika sta velikost otoka in njegova oddaljenost od celine, evlucijska dejavnika sta stopnja izumiranja in stopnja priseljevanja (sin. imigracije) (Slika 8.3).



PRIMER 8.1

Vrstno bogastvo jadranskih otokov – podatki

Tabelarična in grafična predstavitev števila vrst iz treh razredov kopenskih vretenčarjev: dvoživke, plazilci in sesalci na 14 otokih v Jadranskem morju (Slika 8.4, Preglednica 8.1). Primer raziskave SAR odnosa na lokalnem merilu. Raziskavo sta izvedla Kryštufek in Kletečki (2007).



Slika 8.4 Število vrst sesalcev, plazilcev in dvoživk na otokih v Jadranskem morju.

Preglednica 8.1 Površina otoka in število vrst dvoživk, plazilcev in sesalcev na otoku

| Otok | Površina km ² | Dvoživke | Plazilci | Sesalci |
|-----------|--------------------------|----------|----------|---------|
| Cres | 404 | 7 | 20 | 9 |
| Krk | 410 | 7 | 22 | 10 |
| Rab | 91 | 4 | 12 | 7 |
| Pag | 285 | 3 | 13 | 7 |
| Silba | 15 | 1 | 6 | 3 |
| Dugi otok | 124 | 3 | 9 | 4 |
| Ugljan | 52 | 0 | 7 | 5 |
| Brač | 395 | 2 | 14 | 7 |
| Hvar | 300 | 3 | 18 | 8 |
| Lastovo | 47 | 1 | 7 | 6 |
| Korčula | 276 | 2 | 12 | 7 |
| Vis | 90 | 1 | 8 | 4 |
| Mljet | 100 | 2 | 13 | 5 |
| Šolta | 59 | 2 | 8 | 5 |

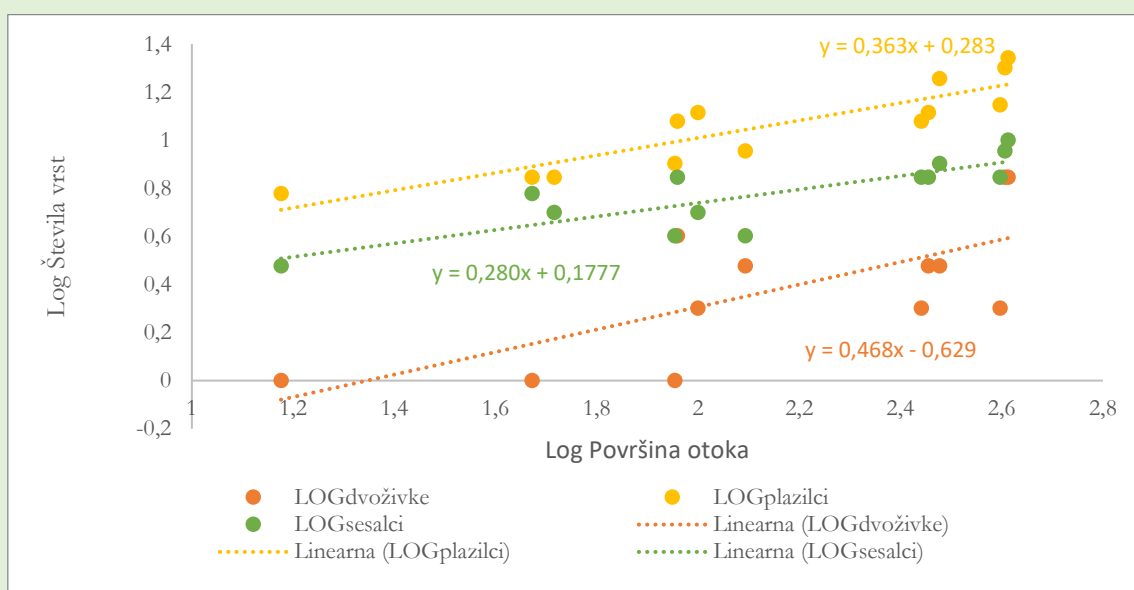




PRIMER 8.2

Vrstno bogastvo jadranskih otokov

Odnos med številom vrst in površino predstavimo na logaritmiranih vrednostih z enostavno linearno regresijo. Rezultat je podan z regresijsko premico in regresijsko enačbo. Najvišje vrstno bogastvo na otokih izkazujejo plazilci v razponu od 6 do 22 vrst s povprečno vrednostjo 12 vrst; sesalcev je v povprečju 6 vrst na otok (razpon od 3 do 10 vrst). Pričakovano je najmanj dvoživk, povprečno so na otoku 3 vrste (razpon od 0 do 7 vrst). Iz smernih koeficientov regresijskih enačb ugotavljamo, da z naraščanjem površine otoka najhitreje narašča vrstno bogastvo dvoživk ($\alpha = 0,478$) in najpočasneje narašča število vrst sesalcev ($\alpha = 0,280$) (Slika 8.5).



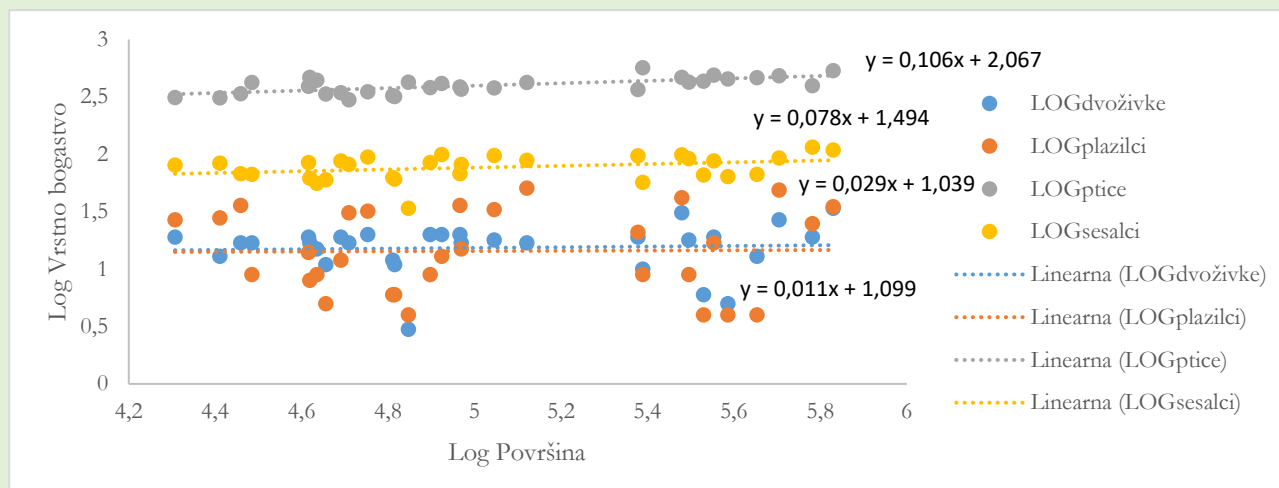
Slika 8.5 Rezultat Log-Log regresijske analize med številom vrst in površino ozemlja.



PRIMER 8.3

Vrstno bogastvo evropskih držav

Primer raziskave na regionalnem merilu in odnos med številom vrst in površino 31 evropskih držav. Analiza je izvedena na logaritmiranih vrednostih z enostavno linearno regresijo. Rezultat je podan z regresijsko premico in regresijsko enačbo. V primeru regionalnega merila so spremembe števila vrst minimalne tudi ob znatnih spremembah površine. Zato so vrednosti smernih koeficientov nizke in komaj večje od nič, od 0,01 do 0,10 (Slika 8.6).



Slika 8.6 Log-Log odnos med površino in vrstnim bogastvom dvoživk, plazilcev, ptic in sesalcev za 31 evropskih držav.

Ko poznamo regresijski odnos med velikostjo površine in številom vrst, lahko podamo dva tipa napovedi:

1. lahko napovemo število vrst za neraziskano površino;
2. za znano površino podamo oceno, ali je ugotovljeno število vrst v skladu s pričakovanji, ali je območje vrstno bogatejše in ali je vrstno revnejše od pričakovanj.

Biodiverzitetu štirih skupin vretenčarjev (Preglednica 8.2) v Sloveniji smo primerjali s pričakovano vrstno pestrostjo iz napovednega modela iz zgornje analize. Zastavili smo raziskovalno vprašanje: Ali je biodiverzitetu tetrapodov v Sloveniji v skladu z modelom SAR. Oblikovali smo ničelno in alternativno hipotezo: H_0 = biodiverzitetu je v skladu z modelom SAR; H_a = biodiverzitetu odstopa od modela SAR. Stopnjo zaupanja smo postavili na 5 % odstopanje.

Iz literature smo povzeli vrstno pestrost in iz regresijskega modela napovedali pričakovano število vrst in izračunali razliko: $\Delta S = S_{\text{opazovana}} - S_{\text{napovedana}}$ (Preglednica 8.2). Glede na sprejete hipoteze vidimo, da biodiverzitetu v vseh štirih primerih odstopa od pričakovane. V treh primerih je bogatejša od pričakovanj (sesalci, plazilci in dvoživke) in v enem primeru (ptiči) revnejša.

Preglednica 8.2 Število opazovanih in napovedanih vrst v Sloveniji

| | $S_{\text{Opazovana}}$ | $S_{\text{Napovedana}}$ | S_{Razlika} | Hipoteza |
|----------|------------------------|-------------------------|----------------------|-----------------|
| Sesalci | 81 | 68 | +13 | H_a bogatejše |
| Dvoživke | 19 | 15 | +4 | H_a bogatejše |
| Plazilci | 27 | 14 | +13 | H_a bogatejše |
| Ptiči | 313 | 334 | -21 | H_a revnejše |

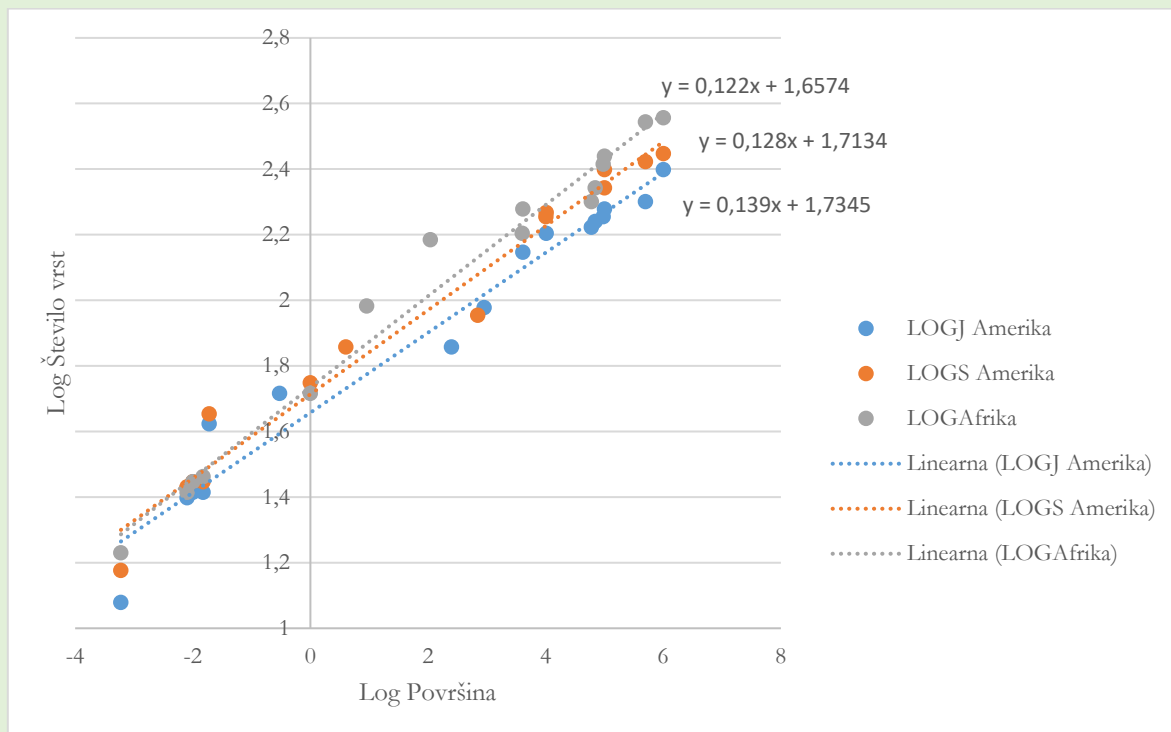




PRIMER 8.4

Vrstno bogastvo kontinentov

Za primer odnosa med številom vrst in površino na globalnem nivoju smo primerjali vrstno bogastvo ptic na treh kontinentih: Južna Amerika, Severna Amerika in Afrika, podatke smo povzeli po Cody (1975). Kontinenti so primerljive velikosti, favne pa so se razvijale bolj ali manj neodvisno ena od druge. Na sliki vidimo, da so regresijske premice skoraj prekrite, to najprej pomeni, da so kontinenti podobne velikosti, in drugič, da imajo podobno vrstno bogastvo. Druga ugotovitev je, da so tudi smerni koeficienti regresijskih enačb primerljive vrednosti (od 0,122 do 0,139) (Slika 8.7). Ta rezultat interpretiramo, da je evolucija ptic na različnih kontinentih potekala z identično dinamiko.



Slika 8.7 Log-Log odnos med površino ozemlja in vrstnim bogastvom ptic za Južno Ameriko, Severno Ameriko in Afriko.



Iskanje prilegajočega se modela. Potenčni oz. linearni model pogosto ne razložita najbolje odnosa med površino območja in številom vrst. Razvitih je več regresijskih modelov za opis odnosa med površino in vrstnim bogastvom. V literaturi je opisanih več kot 20 modelov SAR odnosa, implementirani so v R paket '*sars*' (Matthews in sod. 2019). V nadaljevanju je za vsak model navedeno: ime modela; funkcionalno ime; oblika krivulje

in regresijska enačba, v kateri so: S – število vrst; A – vzorčna površina; d , c , z in f so parametri, ki jih izračunamo.

Asimptotični; sar_asymp(); konveksna; $S = d - cz^A$

Beta-P; sar_betap(); sigmoidna; $S = d(1 - (\frac{A}{c})^z)^{-f}$

Chapman-Richards; sar_chapman(); sigmoidna; $S = d(1 - \exp(-zA))^c$

Logaritmski; sar_loga(); konveksna; $S = c + z \log(A)$

Razširjen potenčni 1; sar_epm1(); konveksna/sigmoidna; $S = c A^{(z A^{-d})}$

Razširjen potenčni 2; sar_epm2(); sigmoidna; $S = c A^{(z - \frac{d}{A})}$

Gompertz; sar_gompertz(); sigmoidna; $S = d \exp(-\exp(-z(A - c)))$

Kobayashi; sar_koba(); konveksna; $S = c \log(1 + \frac{A}{z})$

Linearni; sar_linear(); linearna; $S = c + zA$

Logistični; sar_heleg(); sigmoidna; $S = \frac{c}{f + A^{-z}}$

Monod; sar_monod(); konveksna; $S = \frac{d}{1 + \frac{c}{A}}$

Morgan-Mercer-Flodin; sar_mmf(); sigmoidna; $S = \frac{d}{1 + cA^{-z}}$

Negativni eksponentni; sar_negexpo(); konveksna; $S = d(1 - \exp(-zA))$

Persistenčna funkcija 1; sar_p1(); konveksna; $S = cA^z \exp(-dA)$

Persistenčna funkcija 2; sar_p2(); sigmoidna; $S = cA^z \exp(\frac{-d}{A})$

Potenčni; sar_power(); konveksna; $S = cA^z$

Rosenzweig potenčni; sar_poweR(); konveksna; $S = f + cA^z$

Racioinalni; sar_ratio(); konveksna; $S = \frac{c+zA}{1+dA}$

Weibull-3; sar_weibull3(); sigmoidna; $S = d(1 - \exp(-cA^z))$

Weibull-4; sar_weibull4(); sigmoidna; $S = d(1 - \exp(-cA^z))^f$

Za vsak posamezni model izvedemo izračun prostih parametrov modela in preverimo prileganje modela k dejanskim podatkom. Izračun ob prostih parametrih d , c , z in f , vključuje determinacijski koeficient R^2 in Akaikov informacijski kriterij AIC ter izris regresijske krivulje. Multipli SAR pristop omogoča istočasni izračun vseh modelov in izračun statističnih parametrov. Najbolj prilegajoči se model prepoznamo po najvišji vrednosti R^2 in najnižji vrednosti AIC (za AIC glej razlago v Poglavju 15 Regresija). Izračun dopolnjuje informacija o obliki regresijske krivulje (linearna, konveksna ali sigmoidna) in ali krivulja doseže asimptoto.



PRIMER 8.5

Iskanje prilegajočega se modela

Podatke o vrstnem bogastvu dvoživk, plazilcev in sesalcev na jadranskih otokih (Preglednica 8.1) smo vključili v preverjanje, kateri od modelov najbolje razloži odnos med številom vrst in površino otoka. V analizo smo vključili surove podatke (za razliko od predhodne analize, kjer je predstavljen log-log odnos). Posamezno taksonomsko skupino smo ločeno analizirali, in sicer najnižjo vrednost AIC_c in najvišjo vrednost determinacijskega koeficienta R^2 za dvoživke izkazuje Rosenzweigov potenčni model, ki razloži 64 % variabilnosti v številu vrst dvoživk (Preglednica 8.2.a). V primeru plazilcev in sesalcev je najustreznejši linearni model, ki razloži 78 % in 71 % variabilnosti (Preglednica 8.2. b, c).

Preglednice 8.2 a, b, c Rezultati treh najuspešnejših modelov regresijske analize med površino otokov in številom vrst za a) dvoživke, b) plazilce in c) sesalce. Navedeni so regresijski modeli, utež, korigiran Akaikov informacijski kriterij AIC_c , determinacijski koeficient R^2 in prilagojeni determinacijski koeficient R^2_a , oblika krivulje ter informacija ali krivulja doseže asimptoto

a) dvoživke

| | Model | Utež | AIC_c | R^2 | R^2_a | Oblika | Asimptota |
|---|--------|-------|---------|-------|---------|-------------------|-----------|
| 1 | powerR | 0,234 | 57,436 | 0,641 | 0,534 | konveksna navzdol | ne |
| 2 | linear | 0,149 | 58,330 | 0,490 | 0,397 | linearna | ne |
| 3 | power | 0,124 | 58,695 | 0,476 | 0,381 | konveksna navzgor | ne |

b) plazilci

| | Model | Utež | AICc | R^2 | R^2_a | Oblika | Asimptota |
|---|--------|-------|--------|-------|---------|-------------------|-----------|
| 1 | linear | 0,290 | 71,605 | 0,775 | 0,734 | linearna | ne |
| 2 | power | 0,193 | 72,416 | 0,762 | 0,718 | konveksna navzgor | ne |
| 3 | koba | 0,090 | 73,940 | 0,734 | 0,686 | konveksna navzgor | ne |

c) sesalci

| | Model | Utež | AICc | R^2 | R^2_a | Oblika | Asimptota |
|---|--------|-------|--------|-------|---------|-------------------|-----------|
| 1 | linear | 0,299 | 49,333 | 0,709 | 0,656 | linearna | ne |
| 2 | power | 0,178 | 50,368 | 0,686 | 0,629 | konveksna navzgor | ne |
| 3 | koba | 0,098 | 51,556 | 0,658 | 0,596 | konveksna navzgor | ne |

```
R > package = »sars«
#podatki za dvoživke
mm_otok <- sar_average(data = dvo_otok)
mm_otok
summary(mm_otok)
```



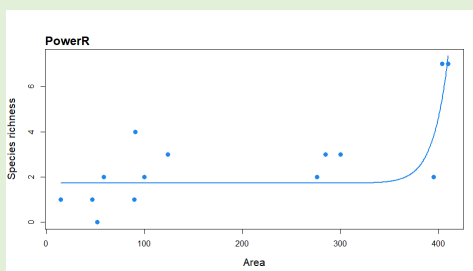
PRIMER 8.6

Prilegajoči se model – dvoživke

V predhodnem preverjanju in iskanju, kateri model najbolj razloži odnos med številom vrst dvoživk in površino otoka, smo ugotovili, da je to Rosenzweigov potenčni model, ki ima obliko $S = f + c * A^z$. Vrednosti parametrov modela so predstavljene v Preglednici 8.3 in grafična predstavitev odnosa na Sliki 8.8.

Preglednica 8.3 Vrednosti parametrov Rosenzweigovega potenčnega modela regresijske analize in vrednosti t -testa med številom vrst dvoživk in površino otoka

| Parameter | Vrednost | Std. napaka | t | p |
|-----------|----------|-------------|--------|--------|
| f | 1,7231 | 0,4105 | 4,1979 | 0,0015 |
| c | 0,00001 | <0,00001 | 0,0095 | 0,9926 |
| z | 26,8780 | 17,4290 | 1,5421 | 0,1513 |



Slika 8.8 Prikaz regresijske krivulje v potenčnem regresijskem modelu med velikostjo otoka in številom vrst dvoživk.

```
R > package = »sars«
#podatki za dvoživke
fit_dvo <- sar_powerR(data = dvo_otok)
fit_dvo
plot(fit_dvo)
summary(fit_dvo)
```



PRIMER 8.7

Prilegajoči se model – plazilci in sesalci

V predhodnem preverjanju modelov smo ugotovili, da linearni model najbolje razloži odnos med številom vrst plazilcev ter sesalcev in površino otoka. Model ima obliko $S = c + m \cdot A$. Vrednosti parametrov modelov so predstavljeni v Preglednicah 8.4.a, b in grafična predstavitev odnosa na Sliki 8.9.

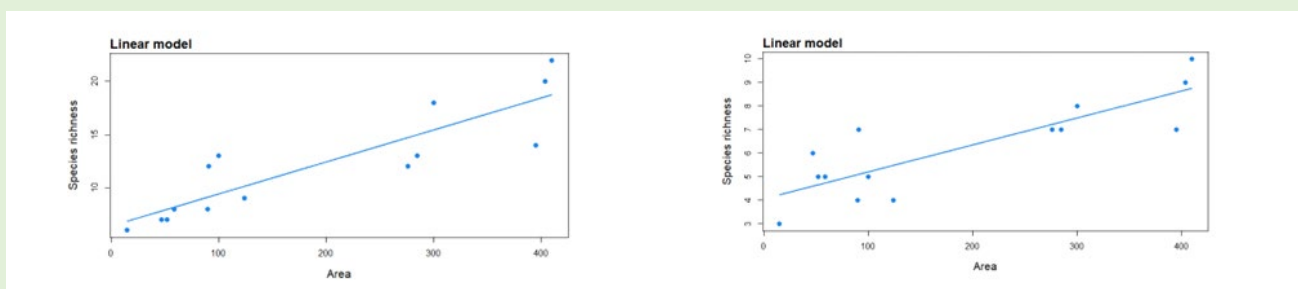
Preglednica 8.4 Vrednosti parametrov linearnega modela regresijske analize in vrednosti t -testa med površino otoka ter številom vrst plazilcev in sesalcev

a) plazilci

| Parameter | Vrednost | Std. napaka | t | p |
|-----------|----------|-------------|--------|--------|
| c | 6,3726 | 1,1095 | 5,7439 | 0,0001 |
| m | 0,0301 | 0,0047 | 6,4311 | 0,0000 |

b) sesalci

| Parameter | Vrednost | Std. napaka | t | p |
|-----------|----------|-------------|--------|--------|
| c | 4,0535 | 0,5008 | 8,0939 | 0,0000 |
| m | 0,0114 | 0,0021 | 5,4022 | 0,0002 |



Slika 8.9 Prikaz regresijske premice v linearnem regresijskem modelu (Linear model) med številom vrst (Species richness) plazilcev (levo) ter sesalcev (desno) in velikostjo otoka (Area).

```
R > package = »sars«  
#plazilci  
fit_plaz <- sar_linear(data = plaz_otok)  
fit_plaz  
plot(fit_plaz)  
summary(fit_plaz)  
#sesalci  
fit_ses <- sar_linear(data = ses_otok)  
fit_ses  
plot(fit_ses)  
summary(fit_ses)summary(fit_dvo)
```



Zaključki

- Z odnosom med površino in vrstnim bogastvom opisujemo spremembo števila vrst ob spremembi površine.
- Večja območja omogočajo evolucijo in preživetje večjemu številu vrst kot manjša območja.
- Odnos med površino in vrstnim bogastvom ima praviloma potenčno obliko.
- Primerjave med vrstnim bogastvom in površino praviloma izvajamo na logaritemsko transformiranih podatkih, zapišemo jih v obliki linearne regresijske enačbe $\log S = \log c + z \log A$.
- Smerni koeficient z opiše strmino krivulje, po kateri narašča število vrst S ob spreminjanju površine A . Pri majhnih površinah (lokalne razmere) je vrednost z visoka $z \geq 0,25$, ob že relativno majhni spremembi površine lahko pride do velike spremembe v številu vrst. Pri srednje velikih površinah (regionalne razmere) je vrednost z majhna $z \approx 0,02$, tudi ob velikih spremembah površine ja spremembah v številu vrst majhna. Na površinah velikosti kontinentov (globalne razmere) je vrednost z srednja $z \approx 0,12$.
- Pri majhnih površinah se lahko pojavi učinek malega otoka, kjer je vrstno bogastvo manjše in neodvisno od površine, dokler površina ne doseže določenega praga velikosti.
- Opisanih je več kot 20 modelov SAR odnosa, implementirani so v R paketa '*sars*', ki izračuna vse modele in med njimi poišče najbolj prilegajočega se. Izračun vsebuje regresijske parametre, determinacijski koeficient R^2 , Akaikov informacijski kriterij AIC, izris regresijske krivulje, informacijo o obliki regresijske krivulje (linearna, konveksna ali sigmoidna) ter ali krivulja doseže asimptoto.

- Zapis v obliki regresijskega modela uporabimo za napovedovanje števila vrst še neraziskanega ozemlja z znano površino.
- Zapis v obliki regresijskega modela uporabimo za primerjanje dejansko izmerjenega vrstnega bogastva s pričakovano vrednostjo, iz česar zaključimo, ali je območje bogatejše ali revnejše od pričakovanja.
- Vrednost $\hat{\zeta}$ iz regresijskega modela uporabimo kot cenilko za primerjavo vrstnega bogastva različnih taksonomskih skupin, ali pa z njo primerjamo vrstno bogastvo različnih območij.

Literatura

- Brown J. 1995. Macroecology. University of Chicago Press. Chicago.
- Cody M.L. 1975. Towards a theory of continental species diversity: Birds distributions over Mediterranean habitat gradients. V: Cody M.L., Diamond J.M. (ur.) Ecology and Evolution Communities. str. 214–257. Harvard University Press. Cambridge.
- Gao D., Wang Y. 2021. A global synthesis of the small-island effect in amphibians and reptiles. *Ecography*, 44: 1–14.
- Kryštufek B., Kletečki E. 2007. Biogeography of small terrestrial vertebrates on the Adriatic landbridge islands. *Folia Zoologica*, 56(3): 225–234.
- Matthews T.J., Triantis K.A., Whittaker R.J., Guilhaumon F. 2019. sars: an R package for fitting, evaluating and comparing species–area relationship models. *Ecography*, 42: 1446–1455 (ver. 0).
- Rosenzweig L.M. 1995. Species Diversity in Space and Time. Cambridge University Press. Cambridge.
- Triantis A.K., Matthews J.T., Whittaker J.R. 2021. The Species-Area Relationship - Theory and Application. Cambridge University Press. Cambridge.

9. poglavje



**VRSTNO BOGASTVO V
GRADIENTU GEOGRAFSKE
ŠIRINE**

Povzetek

9. poglavje: Vrsto bogastvo v gradientu geografske širine

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

Vrstno bogastvo narašča od polov proti tropskim območjem. Je najbolj prepoznaven biodiverzitetni vzorec v makroekologiji. Krivulja porazdelitev vrstnega bogastva dosega vrh okrog ekvatorja. Polovica krivulje, ki opisuje južno poloblo, monotonno upada; druga polovica krivulje, ki opisuje severno poloblo, ima kompleksnejšo obliko, upada od ekvatorja do približno 30°. Sledi prevoj krivulje, ki se v primeru ptic in sesalcev celo obrne navzgor, v primeru dvoživk pa se strmina zmanjša. Pri geografski širini okrog 60° pride do ponovnega prevoja in krivulja strmo upade. Vzorec ohranjanja vrstnega bogastva pri pticah in sesalcih oz. njegovega manjšega upadanja pri dvoživkah v razponu med 30° in 60° je posledica velikih površin kopnega v teh geografskih širinah.

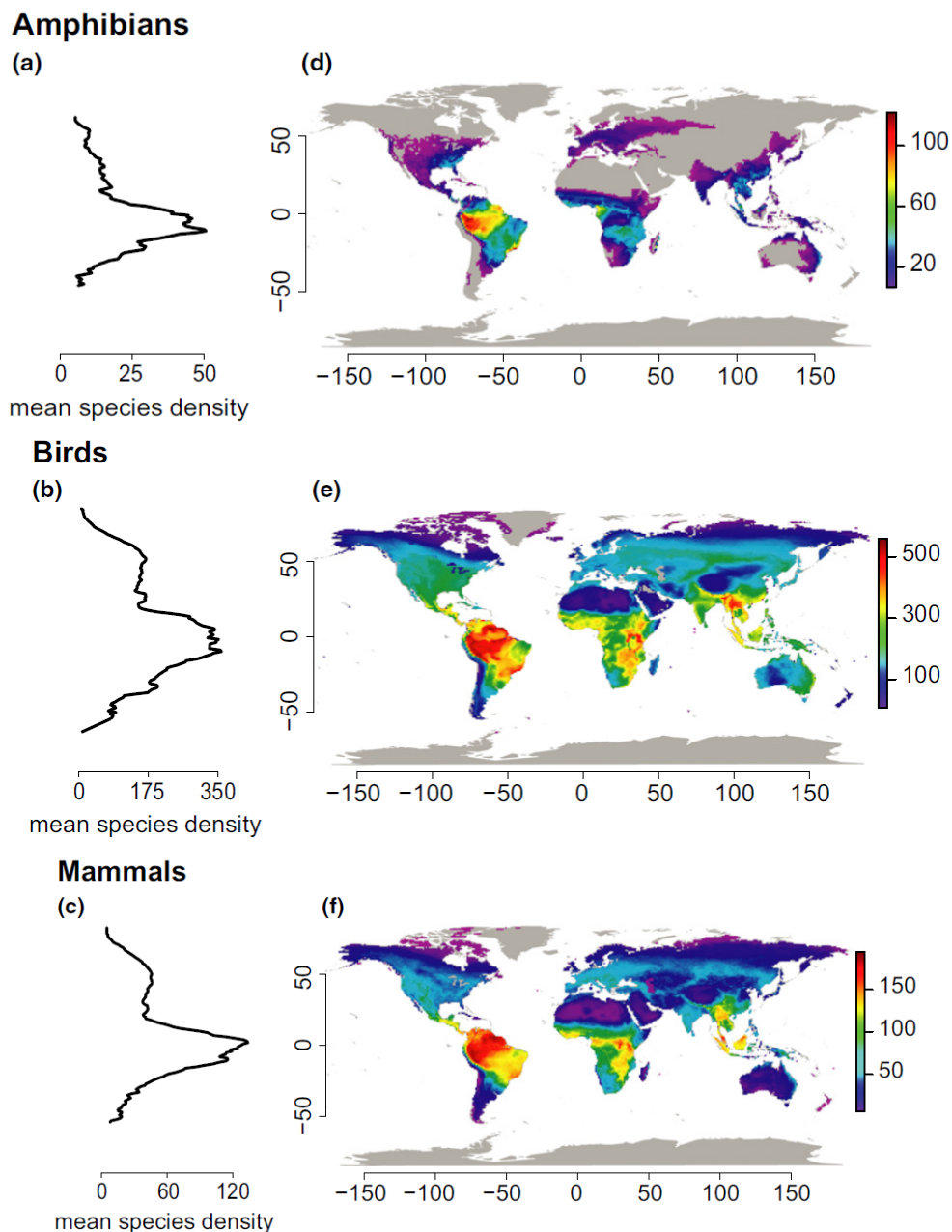
Evolucijski dejavniki, ki oblikujejo vzorec vrstnega bogastva vzdolž gradienta geografske širine, vplivajo na odnos med speciacijo vrst in izumiranjem vrst. Predlagani so trije scenariji za razlago višjega vrstnega bogastva na ekvatorialnem območju: 1. po prvem modelu je izumiranje vrst konstantno po celotni geografski širini, stopnja speciacije je visoka na ekvatorialnem območju in nizka na polih; 2. po drugem modelu je stopnja speciacije konstantna po celotni geografski širini, stopnja izumiranja je višje ob polih in nizka ob ekvatorju; 3. v tretjem modelu je biodiverzitetni vzorec posledica visoke stopnje speciacije ob ekvatorju in visoke stopnje izumiranja ob polih; vrste se premikajo iz tropskih k zunaj tropskim območjem.

Več hipotez razlaga biodiverzitetni vzorec z geofizikalnimi, klimatskimi in ekološkimi vzroki. Biodiverzitetni vzorec vzdolž geografske širine sovпада z več skupinami dejavnikov: 1. velikost površine; 2. gradient klimatskih razmer, npr. od pola proti ekvatorju se večja jakost sončevega sevanja, višajo povprečne temperature, zmanjšuje števila dni s snežno odejo, viša zračna vlaga; 3. zemeljska zgodovina s prisotnostjo ledenih dob in 4. evolucijska dogajanja, stopnja diverzifikacije in starost evolucijskih linij.

Odnos med geografsko širino in vrstnim bogastvom je opisan z enačbo: $S = c + \alpha L$, kjer je S število vrst, L je geografska širina, c in α sta koeficienta, ki ju izračunamo iz podatkov. Na projekciji sta število vrst in geografska širina v linearnem odnosu, pri tem je vrednost c začetna vrednost na Y osi in α je smerni koeficient regresijske premice.

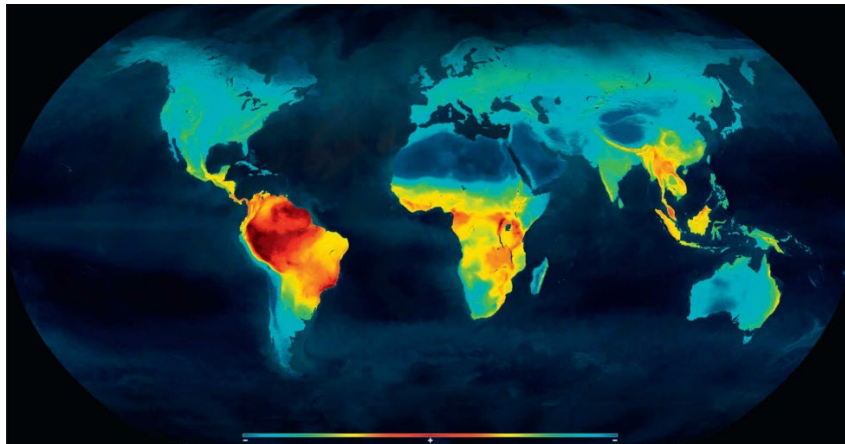
Odnos med vrstnim bogastvom in razlagalnimi spremenljivkami analiziramo s pomočjo korelacijske analize, s katero podamo jakost in smer povezave. Ter linearne in nelinearne regresijske analize, s katero razložimo količino spremembe vrstnega bogastva ob spremembi neodvisne spremenljivke. Vzorec pogosto razlagamo s klimatskimi spremenljivkami; dogovorjenih je 19 bioklimatskih spremenljivk. Za enostavnejšo predstavitev odnosa med klimo in vrstnim bogastvom, 19 bioklimatskih spremenljivk reduciramo v klimatske glavne komponente, s katerimi poenostavljeno razlagamo klimo ter iščemo povezavo med klimo in vrstnim bogastvom.

Biodiverziteteta narašča od polov proti tropskim območjem, to označujemo kot **gradient vrstnega bogastva vzdolž geografske širine** (ang. Latitudinal Diversity Gradient LDG oz. Latitudinal Biodiversity Gradient LBG) (Sliki 9.1 in 9.2). Je najbolj prepoznaven biodiverzitetni vzorec v makroekologiji.



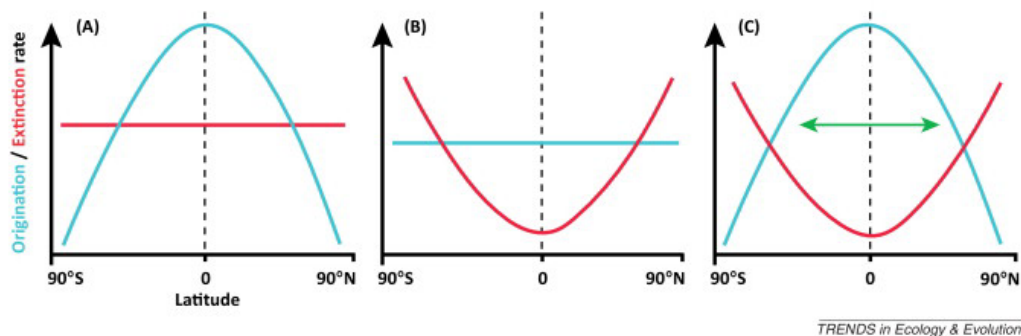
Slika 9.1 Povprečna vrednost vrstnega bogastva glede na geografsko širino za dvoživke, ptiče in sesalce (vzeto iz: Marin in Hedges 2016). Krivulje porazdelitve vrstnega bogastva dosegajo vrh okrog ekvatorja (Slike a, b in c). Polovica krivulje, ki opisuje biodiverzitetno južnega pola, monotono upada; druga polovica krivulje, ki opisuje biodiverzitetno severnega pola, monotono upada od ekvatorja do približno 30°. Sledi prevoj krivulje, ki se v primeru ptic in sesalcev celo obrne navzgor, v primeru dvoživk pa se strmina zmanjša. Pri geografski širini okrog 60° pride do ponovnega prevoja in krivulja strmo upade.

Vzorec ohranjanja vrstnega bogastva pri ptičih in sesalcih oz. njegovega manjšega upadanja pri dvoživkah v razponu med 30° in 60° je posledica velikih površin kopnega v teh geografskih širinah. Na zemljevidu sveta so s toplejšimi barvami označena območja z višjim vrstnim bogastvom in s hladnejšimi barvami območja z manjšim bogastvom vrst (Slike d, e in f).



Slika 9.2 Porazdelitev vrstnega bogastva kopenskih vretenčarjev. Primer biodiverzitetnega vzorca vzdolž gradienta geografske širine. Visoke koncentracije biodiverzitet, predvsem okrog ekvatorialnega pasu, so označene z rdečo barvo. Barvni spekter proti vijolični barvi označuje upadanje biodiverzitetne proti poloma (vzeto iz: Mannion in sod. 2014).

Razumevanje evlucijskih dejavnikov, ki oblikujejo vzorec vrstnega bogastva vzdolž gradienta geografske širine, je predmet znanstvenih debat raziskovalcev biodiverzitetne in makroekologov. Temelj razlage je odnos med speciacijo vrst in izumiranjem vrst. Predlagani so trije scenariji (Slika 9.3), ki bi lahko privedli do višjega vrstnega bogastva na ekvatorialnem območju. Po prvem modelu je izumiranje vrst konstantno po celotni geografski širini. Speciacija je visoka na ekvatorialnem območju in nizka na polih (Slika 9.3.a). Po drugem modelu je speciacija konstantna po celotni geografski širini; izumiranje je višje ob polih in nizko ob ekvatorju (Slika 9.3.b). V tretjem modelu je biodiverzitetni vzorec posledica visoke stopnje speciacije ob ekvatorju in visoke stopnje izumiranja ob polih (Slika 9.3.c).



Slika 9.3 Trije možni modeli zmanjševanja biodiverzitetne proti poloma oz. njenega večanja proti tropskim območjem. A) Model »tropi zibelka biodiverzitet«, po katerem imajo tropska območja najvišjo stopnjo speciacije in polarna najnižjo; stopnja izumiranja je konstantna po celotnem gradientu geografske širine. B) Model »tropi muzej biodiverzitet«, po katerem je stopnja speciacije konstantna in je stopnja izumiranja najnižja v tropih. C) Model »odhod iz tropov«, kjer imajo tropska območja najvišjo stopnjo speciacije; istočasno je stopnja izumiranja najnižja v tropih, vrste se premikajo iz tropskih k zunaj tropskim območjem. Legenda: modra črta – stopnja speciacije, rdeča črta – stopnja izumiranja, črtkana črta – ekvator, zelena črta – disperzija vrst izven tropov (vzeto iz: Mannion in sod. 2014).

Postavljenih je več hipotez, s katerimi skušajo vzorec razložiti z geofizikalnimi, klimatskimi in ekološkimi vzroki. Vse dosedanje hipoteze imajo tudi pomanjkljivosti. Biodiverzitetni vzorec vzdolž gradienta geografske širine sovпада z več skupinami dejavnikov: 1. **velikost površine**; 2. **gradient klimatskih razmer**, npr. od pola proti ekvatorju se večja jakost sončevega sevanja, višajo povprečne temperature, zmanjšuje števila dni s snežno odejo, viša zračna vlaga; 3. **zemeljska zgodovina**, prisotnost ledenih dob in 4. **evolucijska dogajanja**, stopnja diverzifikacije in starost evolucijskih linij. V nadaljevanju so navedene glavne hipoteze s katerimi razlagamo biodiverzitetni vzorec.

Hipoteza geografskega območja (ang. Geographical Area Hypothesis) temelji na dejstvu, da so tropska območja po površini največji biom in velika območja lahko podpirajo več vrst. Večja površina tropskega območja vrstam omogoča večje območje razširjenosti (areal), in posledično večjo številčnost populacij. Prav tako imajo vrste z večjo razširjenostjo nižjo stopnjo izumiranja. In še dodatno, za vrste z velikim območjem razširjenosti je večja verjetnost alopatrične speciacije. Kombinacija nizke stopnje izumiranja in visoke stopnje speciacije je omogočila višjo stopnjo vrstne pestrosti v tropih. Vpliv geografskega območja na vzorec biodiverzitete je zelo odvisen od merilne skale; predvsem zelo vplivajo vrste z majhnim območjem razširjenosti v primerjavi z vrstami z obsežno razširjenostjo.

Energetska hipoteza (ang. Species Energy Hypothesis) temelji na domnevi, da količina razpoložljive energije omejuje vrstno bogastvo. Torej, povečanje sončeve energije (v kombinaciji s pogostostjo vode) pri nizkih geografskih širinah (blizu ekvatorja) povečuje neto primarno produkcijo (fotosintezo). Po predpostavki te hipoteze višja primarna produkcija omogoča večje število osebkov in večje število vrst. Po energetske hipotezi se stopnja izumiranja proti ekvatorju znižuje kot posledica številnejših in stabilnejših populacij na račun večje količine dostopne energije v tropih in to je privedlo k višjemu vrstnemu bogastvu.

Hipoteza ostre klime (ang. Climate Harshness Hypothesis) je še ena hipoteza, povezana s klimo. Hipoteza temelji na domnevi, da je diverzitetni gradient vzdolž geografske širine posledica fiziološke tolerance na ugodnejšo klimo. Klimatske razmere na nižjih geografskih širinah (bližje ekvatorju) so ugodnejše, kot so na višjih geografskih širinah (proti poloma), kjer je hladneje in bolj suho kot v tropih.

Hipoteza stabilne klime (ang. Climate Stability Hypothesis) je podobna hipotezi ostre klime, predvideva pa, da je stabilnost klime pri nižjih geografskih širinah razlog večje pestrosti. Nihajoči okoljski dejavniki po tej hipotezi povečajo stopnjo izumiranja in zavirajo speciacijo. Konstantno okolje omogoča vrstam uspešnejšo specializacijo na predvidljive naravne vire in s tem ožje ekološke niše in speciacijo.

Hipoteza historičnih preobratov (ang. Historical Perturbation Hypothesis) predvideva nižjo biodiverzitetu v višjih geografskih širinah kot posledico poledenitev. Na območjih, ki so bila pod ledenim pokrovom, še ni preteklo dovolj časa, da bi ga vrste ponovno kolonizirale. Po tej hipotezi vrstno bogastvo zmernih območij še ni doseglo optimalnega stanja, zato tam število vrst še narašča.

Stopnja diverzifikacije (ang. Evolutionary Rate Hypothesis) hipoteza predvideva višjo stopnjo speciacije v tropih in s tem višjo biodiverzitetu na nižjih geografskih širinah (bližje ekvatorju). Višja stopnja evolucije v tropih je po tej hipotezi posledica višje temperature okolja, višje stopnje mutacij, krajšega generacijskega časa in hitrejših fizioloških procesov.

Hipoteza efektivnega evolucijskega časa (ang. Hypothesis of Effective Evolutionary Time) predvideva, da je diverzitetu pogojena z dolžino časovnega obdobja z bolj ali manj stabilnimi razmerami v ekosistemu.

Biotska hipoteza (ang. Biotic Hypothesis) predvideva, da so medvrstni odnosi, kot so kompeticija, predacija, mutualizem in parazitizem izraziteje izraženi v tropih. Posledica teh intenzivnejših medvrstnih odnosov je višja stopnja koeksistence in specializacije vrst, kar privede do višje speciacije v tropih.

Učinek osrednjega območja MDE (ang. Mid-domain Effect) temelji na opazovanjih, da se v središču velikega biogeografskega prostora prekrivajo območja razširjenosti večjega števila vrst. To pomeni, da je v središču prisotnih več vrst kot na robnem območju (npr. terestričnih vrst sesalcev je več v središču kontinenta, kot jih je proti robu). Na opisanem spoznanju je narejen teoretični model porazdelitve vrst, v katerem s pomočjo računalniške simulacije predvidimo teoretični razpon porazdelitve vrst vzdolž gradienta. Dejanska opazovanja porazdelitve vrst primerjamo s teoretičnim modelom in ugotavljamo ali so vrste porazdeljene v skladu z napovedjo MDE ali ne.

Odnos med vrstnim bogastvom in geografsko širino podamo s pomočjo formule:

$$S = c + \alpha L,$$

kjer je S število vrst, L je geografska širina, c in α sta koeficienta, ki ju izračunamo iz podatkov. Na projekciji sta število vrst in geografska širina v linearnem odnosu, pri tem je vrednost c začetna vrednost na Y osi (vrednost kjer regresijska premica seka Y os) in α smerni koeficient (naklon premice).

Vsebina poglavja je narejena po temeljni literaturi Brown (1995), Rosenzweig (1995) in Latitudinal gradients in species diversity (2023).



PRIMER 9.1

Vrstno bogastvo dvoživk

Escoriza in Ruhi (2014) sta variabilnost v vrstnem bogastvu dvoživk (Slika 9.4) vzdolž geografske širine v zahodni Evropi razložila s tremi ekološkimi gradienti: povprečno letno temperaturo, indeksom sušnosti in količino biomase (Preglednica 9.1). Korelacijski koeficienti med spremenljivkami okolja in vrstnim bogastvom so visoki in izkazujejo tesno ujemanje števila vrst dvoživk z ekološkimi pogoji. Dvoživke so heterotermni organizmi ter odvisne od temperature okolice in z reprodukcijo vezane na vodne habitate. Zato se z višanjem povprečne letne temperatura njihovo bogastvo povečuje ($r = 0,95$), z večanjem sušnosti pa zmanjšuje ($r = -0,90$) (Preglednica 9.2).

Preglednica 9.1 Okoljski gradienti za pet območij v zahodni Evropi: povprečna geografska širina – (°); povprečna letna temperatura (°C) – pIT, indeks sušnosti – IS, ((povprečne letne padavine/povprečna letna potencialna evapotranspiracija)*10, približek za razpoložljivost vode, vrednosti okrog 0,02 – zelo sušno, >0,65 – vlažno); gostota biomasnega ogljika ($m^3 C ha^{-1}$) – Biomasa (približek za primarno produkcijo); število vrst dvoživk – Št. vrst

| Okoljski gradient /Območje | Geo. širina | pIT | IS | Biomasa | Št. vrst |
|----------------------------|-------------|------|------|---------|----------|
| Iberski polotok | 40 | 13,4 | 6,3 | 33,2 | 28 |
| Francija | 45 | 10,1 | 10 | 18,8 | 19 |
| južna Nemčija | 50 | 8,2 | 9,8 | 27,2 | 17 |
| severna Nemčija in Danska | 55 | 8,6 | 9,8 | 12,2 | 14 |
| skandinavski polotok | 60 | 3,1 | 16,3 | 18,1 | 9 |

Preglednica 9.2 Korelacijski koeficienti med posameznimi ekološkimi spremenljivkami in številom vrst dvoživk

| | Geo. širina | pIT | IS | Biomasa |
|----------|-------------|-------|-------|---------|
| pIT | -0,94 | | | |
| IS | 0,86 | -0,97 | | |
| Biomasa | -0,70 | 0,55 | -0,54 | |
| Št. vrst | -0,97 | 0,95 | -0,90 | 0,76 |



Slika 9.4 Česnovka *Pelobates fuscus*.
Foto. Franc Janžekovič.

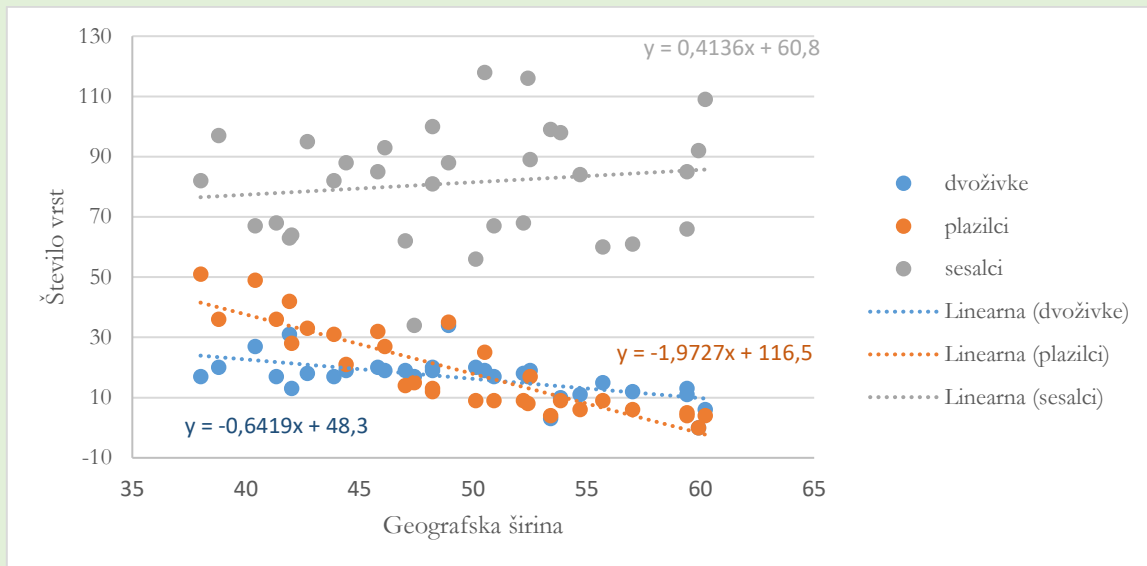




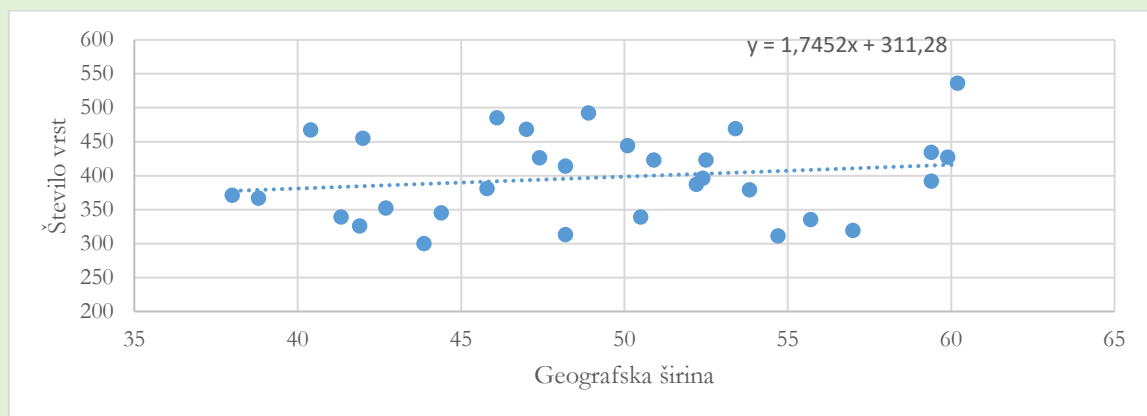
PRIMER 9.2

Vrstno bogastvo tetrapodov

Za 31 evropskih držav smo z enostavno linearno regresijsko analizo prikazali odnos med vrstnim bogastvom tetrapodov in geografsko širino (Sliki 9.5 in 9.6).



Slika 9.5 Vrstno bogastvo dvoživk, plazilcev in sesalcev v evropskih državah vzdolž geografske širine.



Slika 9.6 Vrstno bogastvo ptic v evropskih državah vzdolž geografske širine.

Za proxy geografske širine države smo vzeli geografsko širino glavnega mesta. Regresijski premici obeh homeotermnih organizmov (ptic in sesalcev) proti severu celo rahlo naraščata (smerna koeficienta regresijske enačbe sta pozitivna) in pri obeh heterotermnih skupinah (dvoživke in plazilci) upadata (negativna smerna koeficienta). Naraščajoči trend vrstnega bogastva ptic in sesalcev v geografskih širinah med 30° in 60° je posledica dveh dejavnikov, odnosa SAR in homeotermije. V tem pasu geografskih širin je največ kopnega,

zato je odnos SAR ugoden, večja površina podpira večje vrstno bogastvo. Drugi vzrok je kombinacija fiziologije (stalna telesna temperatura) in zmernih klimatskih razmer na tem območju. To omogoča zadosten evolucijski odziv s stabilnim vrstnim bogastvom. Padajoči trend vrstnega bogastva pri dvoživkah in plazilcih je seveda posledica nižanja temperature proti severu. Kljub temu upad števila vrst ni zelo izrazit, kar kompenzira ugodni SAR odnos.



PRIMER 9.3

Vrstno bogastvo tetrapodov – bioklimatska razlaga

Klimatske razmere zelo neposredno vplivajo na fiziologijo vse biote, in s tem tudi na vrstno bogastvo. Klima vpliva na primarno produkcijo, in s tem na kompleksnost prehranjevalnega spleta, tako se izraža sekundarni vpliv na vrstno bogastvo. V ekoloških analizah operiramo s klimatskimi indeksi oz. bioklimatskimi spremenljivkami (ang. Bioclimatic Predictors). Podatki pokrivajo celotno zemeljsko oblo in so prosto dostopni s pomočjo različnih orodij. Predstavljamo paket »raster« (Hijmans 2020) iz programskega okolja R. Poznamo 19 dogovorjenih bioklimatskih spremenljivk (O'Donnell in Ignizio, 2012), (Preglednica 9.2).

Preglednica 9.2 Pregled 19 bioklimatskih spremenljivk

| Koda | Spremenljivka |
|-------|--|
| BIO1 | Povprečna letna temperatura (*100) |
| BIO2 | Povprečni dnevni razpon (mesečno povprečje (max. tem – min. tem.)) |
| BIO3 | Izotermalnost (BIO2/BIO7) (*100) |
| BIO4 | Sezonskost temperature (standardna deviacija * 100) |
| BIO5 | Maksimalna temperatura vročega meseca |
| BIO6 | Minimalna temperatura hladnega meseca |
| BIO7 | Letni razpon temperature (BIO5 – BIO6) |
| BIO8 | Povprečna temperatura mokrega četrletja |
| BIO9 | Povprečna temperatura suhega četrletja |
| BIO10 | Povprečna temperatura vročega četrletja |
| BIO11 | Povprečna temperatura hladnega četrletja |
| BIO12 | Letne padavine |
| BIO13 | Padavine mokrega meseca |
| BIO14 | Padavine suhega meseca |
| BIO15 | Sezonskost padavin (koeficient variabilnosti) |
| BIO16 | Padavine mokrega četrletja |
| BIO17 | Padavine suhega četrletja |
| BIO18 | Padavine toplega četrletja |
| BIO19 | Padavine hladnega četrletja |

```

R > package = »raster«
r <- getData("worldclim",var="bio",res=10) # Priklic klimatskih podatkov
r <- r[[c(1:19)]] # Definiranje klimatskih spremenljivk
names(r) <- c("bio1", ..., "bio19")
lats <- c(41.33, 48.2, 50.9, ...) # Definiranje lokalitet
lons <- c(19.8, 16.35, 4.53, ...)
coords <- data.frame(x=lons,y=lats)
points <- SpatialPoints(coords, proj4string = r@crs)
values <- extract(r,points)
df <- cbind.data.frame(coordinates(points),values)
df
write.table(df, file = "df.txt", sep="\t") # Izvoz podatkov

```



PRIMER 9.4

Vrstno bogastvo tetrapodov – korelacije

Biodiverzitetni vzorec vrstnega bogastvo tetrapodov vzdolž geografske širine v Evropi lahko deloma razložimo s klimatski dejavniki.

V analizi želimo razložiti variabilnost vrstnega bogastva S (ena odvisna spremenljivka), z bioklimatskimi značilnostmi BIO (19 neodvisnih oz. razlagalnih spremenljivk). Prvi pristop je izračun korelacijskega koeficienta med vsako posamezno BIO in S za posamezno skupino živali, ter interpretacija vpliva posamezne BIO na vrstno bogastvo. Homeotermni skupini, ptice in sesalci, ne izkazujejo niti ene močne povezave, to pomeni, da vrstno bogastvo teh dveh skupin ni povezano z bioklimatskimi razmerami. Obe heterotermni skupini, dvoživke in plazilci, izkazujeta več močnih korelacij predvsem s temperaturnimi spremenljivkami. Korelacija med BIO1 (povprečna letna temperatura) in vrstnim bogastvom plazilcev je $r = 0,837$ in z vrstnim bogastvom dvoživk $r = 0,587$. Obe povezavi sta pozitivni in močni, kar pomeni, da število vrst plazilcev in dvoživk narašča ob naraščanju temperature oz. upada ob nižanju povprečnih temperatur (Preglednica 9.3).

Preglednica 9.3 Korelacijski koeficienti med vrstnim bogastvom tetrapodov (dvoživke, plazilci, ptice, sesalci) ter 19 bioklimatskimi spremenljivkami (BIO1 – 19), in geografsko širino (Y) in geografsko dolžino (X). Močne povezave $|r| > 0,5$ so v krepkem tisku

| Spremenljivka | r Dvoživke | r Plazilci | r Ptice | r Sesalci |
|---------------|-----------------|-----------------|--------------|----------------|
| X | -0,129 | -0,078 | -0,315 | -0,095 |
| Y | -0,601 | -0,880 | 0,183 | 0,140 |
| BIO1 | 0,587 | 0,837 | -0,124 | -0,142 |
| BIO2 | 0,548 | 0,628 | -0,081 | -0,198 |
| BIO3 | 0,400 | 0,511 | 0,035 | 0,018 |
| BIO4 | -0,091 | -0,162 | -0,122 | -0,108 |
| BIO5 | 0,623 | 0,853 | -0,138 | -0,242 |
| BIO6 | 0,398 | 0,587 | 0,009 | -0,028 |
| BIO7 | 0,092 | 0,084 | -0,117 | -0,162 |
| BIO8 | 0,183 | -0,163 | -0,291 | -0,061 |
| BIO9 | 0,352 | 0,729 | 0,017 | -0,158 |
| BIO10 | 0,597 | 0,857 | -0,191 | -0,223 |
| BIO11 | 0,441 | 0,665 | -0,029 | -0,050 |
| BIO12 | -0,117 | 0,034 | -0,055 | 0,198 |
| BIO13 | -0,051 | 0,188 | -0,224 | 0,148 |
| BIO14 | -0,205 | -0,254 | 0,129 | 0,236 |
| BIO15 | 0,109 | 0,349 | -0,216 | -0,095 |
| BIO16 | -0,067 | 0,146 | -0,198 | 0,160 |
| BIO17 | -0,163 | -0,213 | 0,096 | 0,225 |
| BIO18 | -0,288 | -0,553 | 0,054 | 0,171 |
| BIO19 | 0,025 | 0,369 | -0,137 | 0,157 |

R >

#Korelacija med št. vrst tetrapodov in bioklimatskimi spremenljivkami

#s = tabela z vrstnim bogastvom tetrapodov, df = bioklimatske

cor(S\$sesalci, df)

cor(S\$ptiči, df)

cor(S\$plazilci, df)

cor(S\$dvoživke, df)





PRIMER 9.5

Vrstno bogastvo tetrapodov - PCA

Drugi pristop obravnavanega odnosa med vrstnim bogastvom in klimo temelji na redukciji števila klimatskih spremenljivk, z združitvi vseh 19 BIO v manjše število komponent, ki ohranijo variabilnost osnovnega niza podatkov. Analiza glavnih komponent PCA je multivariatna statistična metoda, s katero osnovni niz spremenljivk transformiramo na način, da skupni del variance združimo v nove »sintetične« spremenljivke, ki jih imenujemo glavne komponente PC (ang. Principal Components). Analiza glavnih komponent PCA temelji na korelacijski analizi in bolj, kot so osnovne spremenljivke med seboj povezane, višji kot so korelacijski koeficienti med njimi, večji delež skupne variabilnosti razloži 1. glavna komponenta PC1. V PCA smo vključili vseh 19 BIO klimatskih spremenljivk. Rezultat analize so klimatske glavne komponente KPC, ki so urejene po deležu variance, ki jo razlagajo. KPC1 razloži najvišji delež variance, KPC2 naslednji delež itd. Analizo PCA smo izvedli v okolju R (R Core Team 2022) in s pomočjo paketa »factoextra« (Kassambara in Mundt 2020), grafiko smo izdelali s pomočjo paketa »ggplot2« (Wickham 2016) in Excela.

Absolutne vrednosti surovih podatkov se med spremenljivkami zelo razlikujejo, v multivariatni obravnavi je to moteče, saj variabilnost spremenljivke z absolutno višjimi vrednostmi, zakrije variabilnost spremenljivke z absolutno manjšimi vrednostmi. Za zagotovitev enakovredne obravnave vseh spremenljivk, surove podatke posamezne spremenljivke pred vključitvijo v PCA standardiziramo po formuli:

$$z = \frac{a - \bar{x}}{StD},$$

kjer so z – standardizirana vrednost, a – konkretni podatek, \bar{x} – aritmetična sredina, StD – standardna deviacija. Rezultat standardizacije so vrednosti z , posamezni spremenljivki se priredi vrednost aritmetične sredine 0. Vse vrednosti so praviloma razporejene v treh standardnih deviacijah, to pomeni, da je večina vrednosti v razponu od -3 do 3.

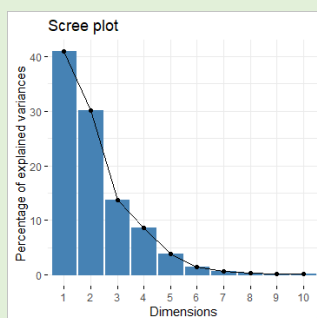
Lastna vrednost v PCA je utež, s katero PC razlaga skupno varianco osnovnega niza podatkov. PC1 ima lastno vrednost 7,78, kar pomeni, da razloži 40,9 % skupne variabilnosti vseh 19 BIO spremenljivk (Preglednica 9.4). Vsota vseh lastnih vrednosti je 19, toliko kot je osnovnih klimatskih spremenljivk.

Na tej stopnji analize moramo sprejeti odločitev, koliko glavnih komponent zadostno razloži variabilnost osnovnega niza podatkov. Odločitev o zadostnem številu PC, ki jih peljemo v nadaljnjo analizo, je arbitrarna. Pomoč pri odločitvi so štiri cenilke: – lastna

vrednost zadnje PC je večja od 1; – zadnja PC razloži vsaj 5 % variabilnosti; – kumulativna vrednost razložene variabilnosti je vsaj 80 %; – metoda komolca, ki poda informacijo, pri kateri PC se delež razložene variance ne poveča več bistveno. Metodo komolca odčitamo iz projekcije deležev razložene variabilnosti na GK. Kjer se krivulja prelomi, postavimo mejo. Na Sliki 9.7 je prevoj na PC3, vendar ni izrazit. Prve tri PC kumulativno razložijo 84,8 % variance, kar pomeni visok in zadosten delež za nadaljnje analize.

Preglednica 9.4 Lastne vrednosti in delež razložene variance za posamezno glavno komponento

| PC | Lastna vrednost | % variance | Kum. % variance |
|------|-----------------|------------|-----------------|
| PC1 | 7,781 | 40,95 | 40,95 |
| PC2 | 5,713 | 30,07 | 71,02 |
| PC3 | 2,618 | 13,78 | 84,80 |
| PC4 | 1,622 | 8,54 | 93,34 |
| PC5 | 0,716 | 3,77 | 97,11 |
| PC6 | 0,285 | 1,50 | 98,61 |
| PC7 | 0,121 | 0,64 | 99,24 |
| PC8 | 0,055 | 0,29 | 99,53 |
| PC9 | 0,046 | 0,24 | 99,77 |
| PC10 | 0,019 | 0,10 | 99,87 |
| PC11 | 0,009 | 0,05 | 99,92 |
| PC12 | 0,005 | 0,03 | 99,95 |
| PC13 | 0,004 | 0,02 | 99,97 |
| PC14 | 0,003 | 0,02 | 99,99 |
| PC15 | 0,001 | 0,01 | 99,99 |
| PC16 | 0,001 | 0,00 | 100,00 |
| PC17 | 0,001 | 0,00 | 100,00 |
| PC18 | 0,000 | 0,00 | 100,00 |
| PC19 | 0,000 | 0,00 | 100,00 |



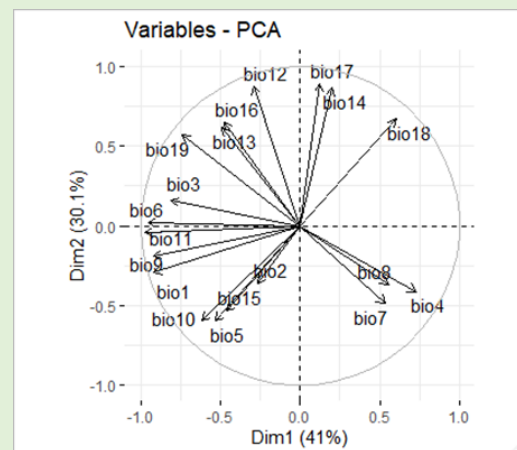
Slika 9.7 Metoda komolca za sprejem ustreznega števila PC.

Prvim trem klimatskim glavnim komponentam KPC1, KPC2 in KPC3, ki jih uporabimo v nadaljnjih analizah, moramo dodati vsebino. To pomeni, da s korelacijsko analizo preverimo, katere deleže variabilnosti osnovnih klimatskih spremenljivk razloži posamezna glavna komponenta. Komponenta KPC1 je v močni negativni korelaciji s spremenljivkami BIO1, 3, 5, 6, 9, 10, 11 in 19 ter v pozitivni korelaciji z BIO4, 7 in 8 (Preglednica 9.5). To pomeni, da z naraščanjem KPC1 $KPC1 > 0$ opisujemo okolje s hladno klimo, mrzlimi zimami in zmernimi poletji ter velikimi razlikami med sezonami; z

nižanjem njene vrednosti $KPC1 < 0$ pa opisujemo okolje s toplo klimo, zmernimi zimami in vročimi poletji ter manjšimi razlikami med sezonami. Posplošimo lahko, da $KPC1$ opisuje temperaturne razmere, v negativno smer opisuje toplejšo klimo v pozitivni smeri pa hladnejšo klimo. Komponenta $KPC2$ je v močni pozitivni korelaciji s spremenljivkami BIO12, 13, 15, 16, 17, 18 in 19 ter v negativni korelaciji z BIO5, 10 in 15 (Preglednica 9.5). To pomeni, da z naraščanjem vrednosti $KPC2$ opisujemo okolje z veliko padavinami, z upadanjem $KPC2$ opisujemo okolje z malo padavin. Posplošimo, komponenta $KPC2$ opisuje padavinski režim, v pozitivni smeri obilje padavin in v negativni smeri sušne razmere. Odnosi med prvima dvema komponentama in BIO spremenljivkami so predstavljeni na krožnem diagramu (Slika 9.8). Krog predstavlja vrednost korelacijskega koeficienta $r = 1$, vektor pa smer in jakost povezave z 1. oz. 2. glavno komponento.

Preglednica 9.5 Korelacijski koeficienti med PC in BIO, močne korelacije $|r| > 0,5$ so poudarjene s krepko pisavo

| | PC1 | PC2 | PC3 |
|-------|---------------|---------------|---------------|
| BIO1 | -0,916 | -0,295 | -0,187 |
| BIO2 | -0,266 | -0,362 | -0,749 |
| BIO3 | -0,810 | 0,156 | -0,030 |
| BIO4 | 0,726 | -0,413 | -0,475 |
| BIO5 | -0,534 | -0,597 | -0,580 |
| BIO6 | -0,952 | 0,019 | 0,205 |
| BIO7 | 0,533 | -0,489 | -0,663 |
| BIO8 | 0,555 | -0,366 | -0,314 |
| BIO9 | -0,918 | -0,189 | 0,006 |
| BIO10 | -0,616 | -0,597 | -0,471 |
| BIO11 | -0,974 | -0,042 | 0,102 |
| BIO12 | -0,292 | 0,874 | -0,335 |
| BIO13 | -0,494 | 0,619 | -0,356 |
| BIO14 | 0,202 | 0,865 | -0,297 |
| BIO15 | -0,458 | -0,533 | 0,097 |
| BIO16 | -0,476 | 0,653 | -0,313 |
| BIO17 | 0,124 | 0,891 | -0,301 |
| BIO18 | 0,603 | 0,671 | -0,323 |
| BIO19 | -0,739 | 0,572 | -0,125 |



Slika 9.8 Smer in jakost posamezne bioklimatske spremenljivke (bio) v dvorazsežnem prostoru prvih dveh klimatskih glavnih komponent Dim1/Dim2.

```
R > package = »factoextra«
head(df) #pregled podatkov
df.active<-df[1:31,3:21] #Izbor spremenljivk
#Izračun PCA
res.pca<-prcomp(df.active, scale. = TRUE) #Rezultati PCA
eig.val<-get_eigenvalue(res.pca) #Izpis lastnih vrednosti
eig.val
fviz_eig(res.pca) #Graf-komolec
res.var <- get_pca_var(res.pca) # Korelacija med PC in BIO
res.var$coord
#Graf - prispevek osnovnih spremenljivk k PC1 in PC2
fviz_pca_var(res.pca, repel = TRUE)
```

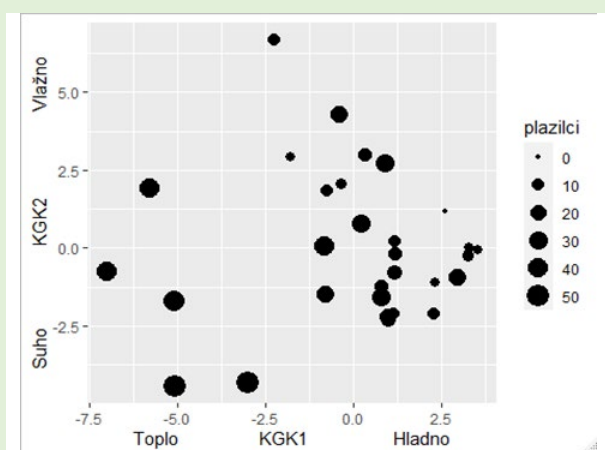




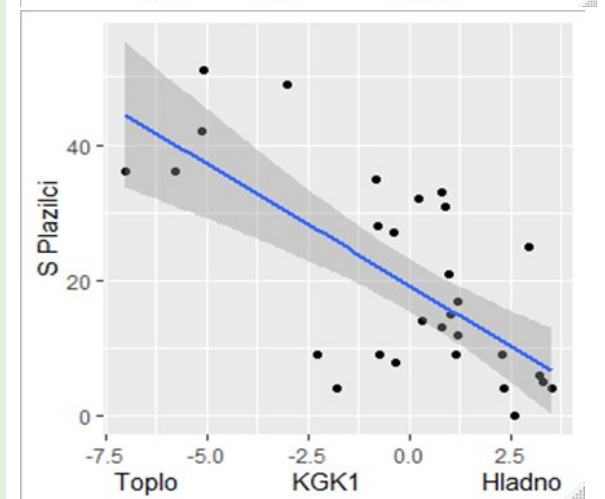
PRIMER 9.6

Vrstno bogastvo tetrapodov – bioklimatska razlaga

Odnos med vrstnim bogastvom in klimatskimi glavnimi komponentami KPC lahko prikažemo na različne načine, predstavljena je samo variabilnost plazilcev. Ena možnost je projekcijo števila vrst plazilcev na prvi dve KPC1 in KPC2, kjer je vrstno bogastvo prikazano z velikostjo pike (Slika 9.9). Velike pike, vrstno bogata območja, so razporejene v levem spodnjem kvartilu, ki opisuje toplo in suho podnebje. Male pike, z vrstno revnim bogastvom, so razporejene v desnem območju grafa, ki opisuje hladno klimo. Med obema skrajnostma so razporejeni preostali vzorci. Druga možnost predstavitve je raztresen graf, kjer je na Y osi vrstna pestrost S plazilcev (odvisna spremenljivka) na X osi pa KPC1 (neodvisna spremenljivka), skozi oblak točk je položena regresijska premica. Število vrst plazilcev je v levem delu grafa visoko (negativne vrednosti KPC1), območje s toplo klimo, v desnem območju grafa je število plazilcev nizko (pozitivne vrednosti KPC1), območje s hladno klimo (Slika 9.10).



Slika 9.9 Projekcija vrstnega bogastva plazilcev (velikost pik) na klimatski komponenti KPC1/KPC2.



Slika 9.10 Projekcija vrstnega bogastva plazilcev na prvo klimatsko komponento KPC1 z regresijsko premico.

```
R -> dplyr
# Združitec KGK in S
kgk_s<-bind_cols(kgk_dr, S)
head(kgk_s)
```

```
R > package = »ggplot2«
#Biplot z velikostjo pik za vrstno pestrost
ggplot(kgk_s, aes(y=PC2, x=PC1))+geom_point(aes(size=plazilci))+
  labs(x="Toplo      KGK1      Hladno",
       y="Suho      KGK2      Vlažno")
# Biplot z regresijsko premico
ggplot(kgk_s, aes(y=plazilci,
x=PC1))+geom_point()+geom_smooth(method=lm)+
  labs(x="Toplo      KGK1      Hladno".    v="S Plazilci")
```



PRIMER 9.7

Časovna razlaga – podatki

Splošni vzorec višjega vrstnega bogastva v tropskih območjih je še vedno predmet raziskav in soočenj različnih hipotez. Po razlagi učinkovitega evlucijskega časa je višja diverziteteta v tropih pogojena z daljšim obdobjem stabilnih okoljskih/ekoloških razmer. Po hipotezi o višji stopnji diverzifikacije (nastajanja novih vrst) je višje vrstno bogastvo v tropih posledica hitrejših evlucijskih procesov, in s tem višje biodiverzitetete. Višja stopnja speciacije v tropih je po tej hipotezi posledica višje temperature okolja, višje stopnje mutacij, krajšega generacijskega časa in hitrejših fizioloških procesov. Podatke za soočenje hipotez smo povzeli po Marin in Hedges (2016) (Preglednica 9.6). Pri treh preučevanih skupinah živali (dvoživke, ptice in sesalci) so imeli tropi v začetku večje število kladov v primerjavi z zmenim pasom. Iz tega lahko sklepamo, da velja teorija učinkovitega evlucijskega časa. Tropi so imeli v oligocenu (približno 30 milijonov let v preteklosti) začetek raziskave višjo biodiverziteteto. Diverzifikacija novih kladov pri nobeni skupini ni konstantna, ampak je v času vse hitrejša. V regresijski analizi na surovih podatkih to vidimo kot potenčno naraščanje krivulj. Po logaritemski transformaciji podatkov sta regresijski premici za trope in zmerno območje vzporedni. To pomeni, da je stopnja diferzifikacije za obe območji izenačena, zaradi začetne prednosti tropov, se razlika v absolutnih vrednosti vrstnega bogastva večja.

Hipotezo o hitrejši diverzifikaciji v tropih lahko zavrnamo, saj smo pokazali, da je stopnja rasti izenačena, do razlike pa je prišlo zaradi začetne razlike v vrstnem bogastvu, ki je bilo naklonjeno tropom. Prednost damo hipotezi o učinkovitem evolucijskem času in bolj stabilnim razmeram v tropih. Zmerni pas je bil izpostavljen večkratnim ledenim dobam, ki so zdesetkale vrstno bogastvo. Tako je nastala začetna razlika med tropskim in zmernim območjem.

Preglednica 9.6 Število kladov v geološki preteklosti za dvoživke, ptice in sesalce

| Skupina | Perioda milijon let v preteklost | Povprečje periode milijon let | Tropi | Zmerno |
|----------|----------------------------------|-------------------------------|-------|--------|
| Dvoživke | 25 - 30 | 28 | 37 | 8 |
| | 20 - 25 | 23 | 59 | 10 |
| | 15 - 20 | 18 | 87 | 18 |
| | 10 - 15 | 13 | 140 | 30 |
| | 5 - 10 | 8 | 190 | 43 |
| | 0 - 5 | 3 | 254 | 52 |
| Ptice | 25 - 30 | 28 | 22 | 4 |
| | 20 - 25 | 23 | 38 | 6 |
| | 15 - 20 | 18 | 78 | 18 |
| | 10 - 15 | 13 | 143 | 23 |
| | 5 - 10 | 8 | 331 | 65 |
| | 0 - 5 | 3 | 421 | 91 |
| Sesalci | 25 - 30 | 28 | 7 | 3 |
| | 20 - 25 | 23 | 15 | 2 |
| | 15 - 20 | 18 | 35 | 7 |
| | 10 - 15 | 13 | 67 | 25 |
| | 5 - 10 | 8 | 172 | 69 |
| | 0 - 5 | 3 | 272 | 93 |

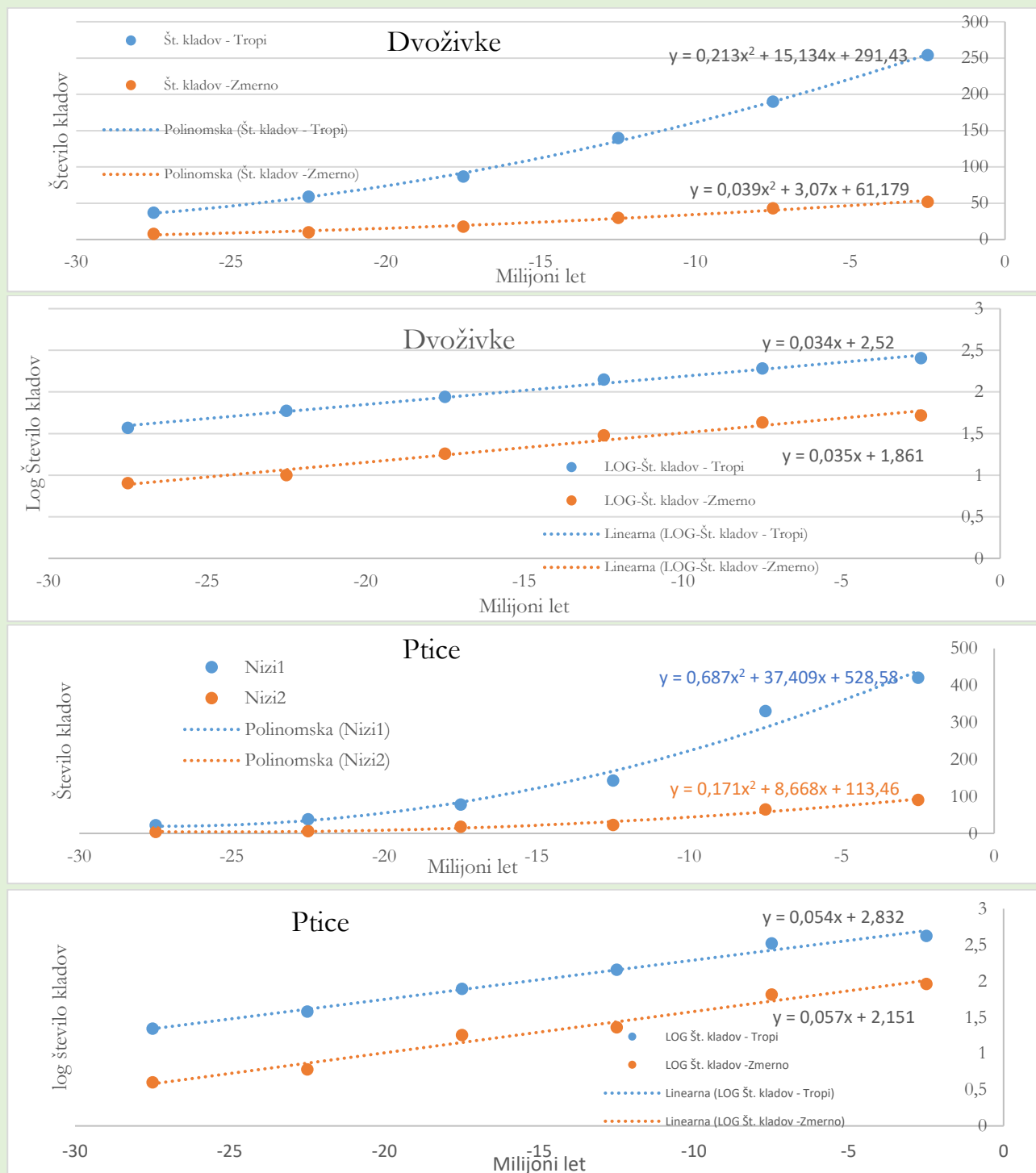




PRIMER 9.8

Časovna razlaga

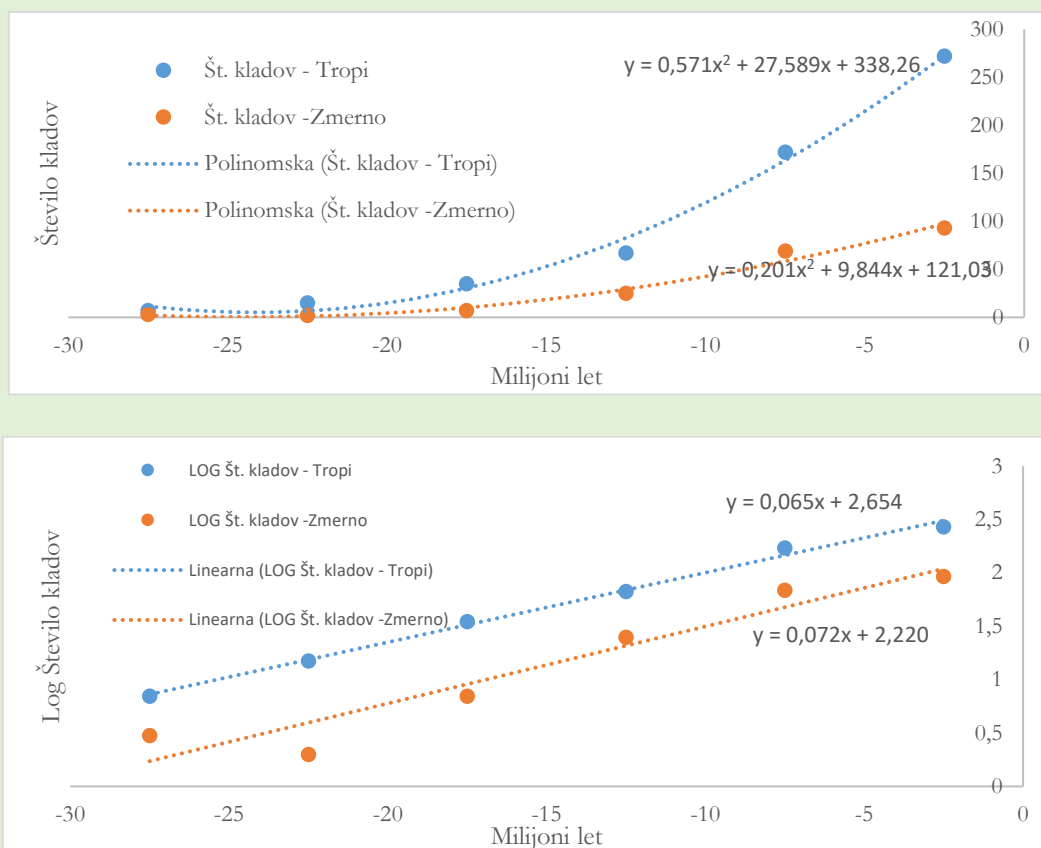
Vrstno bogastvo v času analiziramo z regresijsko analizo. Iščemo model, s katerim najboljše opišemo odnos med številom vrst in časovno vrsto. Porazdelitev surovih podatkov zelo dobro opiše polinomska funkcije druge stopnje, bodisi v tropih bodisi v zmernem območju (Slika 9.11).



Slike 9.11a, b, c, d Regresijska analiza variabilnosti vrstnega bogastva dvoživk in ptic v času. Surovi podatki analizirani s polinomsko regresijo, logaritmirani podatki z enostavno linearno regresijo.

V makroekoloških študijah imamo pogosto opravka z asimetrično porazdeljenimi podatki, zato jih skušamo pretvoriti v bolj simetrično obliko. S tem lahko zagotovimo boljše interpretacijo odnosov med spremenljivkami ali pa zadostimo določenim omejitvam oz. zahtevam v statističnem vrednotenju. Pri logaritemski transformaciji vrednosti transformiramo z logaritmiranjem z osnovo 10, metoda je v analizi ekoloških podatkov zelo uporabna v primerih asimetrične porazdelitve v desno oz. porazdelitve s pozitivno asimetrijo. Omejitev logaritemske transformacije je pri vrednosti 0 in negativni vrednosti, saj logaritmi zanje ne obstajajo.

Število kladov sesalcev v časovni vrsti narašča s potenčno funkcijo. Vrednosti števila vrst za trope in za zmerno območje transformiramo z logaritmiranjem z osnovo 10, s transformacijo jih naredimo primerljive. Na projekciji logaritmiranih vrednosti na časovno vrsto je absolutno večje bogastvo v tropih, odnos med vrstnim bogastvom in časovno vrsto je linearno. Obe regresijski premici, tako za trope kot za zmerno območje, sta blizu vzporednosti. Smerni koeficient za trope znaša $\alpha = 0,065$, za zmerno območje pa $\alpha = 0,072$ (Slika 9.12).



Sliki 12 a, b Regresijska analiza variabilnosti vrstnega bogastva sesalcev v času. Surovi podatki, analizirani s polinomsko regresijo in logaritmirani podatki z enostavno linearno regresijo.



Zaključki

- Vrstno bogastvo narašča od polov proti tropskim območjem.
- Porazdelitev vrstnega bogastva dosega vrh okrog ekvatorja. Polovica krivulje, ki opisuje biodiverzitetno južne poloble, monotono upada; druga polovica krivulje, ki opisuje biodiverzitetno severne poloble, monotono upada od ekvatorja do približno 30° geografske širine. Sledi prevoj in ohranjanja ali nižje zmanjševanje do okrog 60°, kjer se previje in krivulja strmo upade.
- Vzorec vrstnega bogastva vzdolž gradienta geografske širine izhajajo iz odnosa med speciacijo vrst in izumiranjem vrst. Predlagani so trije evlucijski scenariji: 1. izumiranje vrst je konstantno po celotni geografski širini, stopnja speciacije je visoka na ekvatorialnem območju in nizka na polih. 2. stopnja speciacije je konstantna po celotni geografski širini, stopnja izumiranja je višje ob polih in nizka ob ekvatorju. 3. biodiverzitetni vzorec je posledica visoke stopnje speciacije ob ekvatorju in visoke stopnje izumiranja ob polih, vrste pa se premikajo iz tropskih k zmernim območjem.
- Biodiverzitetni vzorec vzdolž geografske širine sovпада z več skupinami dejavnikov: 1. velikost površine; 2. gradient klimatskih razmer, od pola proti ekvatorju se večja jakost sončevega sevanja, viša se povprečna temperatura, zmanjšuje se števila dni s snežno odejo, viša se zračna vlaga; 3. zemeljska zgodovina, prisotnost ledenih dob in 4. evlucijska dogajanja, stopnja diverzifikacije in starost evlucijskih linij.
- Odnos med vrstnim bogastvom in geografsko širino je opisan z enačbo $S = c + z L$, kjer je S število vrst, L je geografska širina in c in z sta koeficienta, ki ju izračunamo iz podatkov. Na projekciji sta število vrst in geografska širina v linearnem odnosu, pri tem je vrednost c začetna vrednost na Y osi in z smerni koeficient regresijske premice.
- Odnos med vrstnim bogastvom in razlagalnimi spremenljivkami analiziramo s pomočjo korelacijske analize ter linearne in nelinearne regresijske analize. Biodiverzitetni vzorec pogosto razlagamo s klimatskimi spremenljivkami, dogovorjenih je 19 bioklimatskih spremenljivk.

Literatura

- Brown J. 1995. *Macroecology*. University of Chicago Press. Chicago.
- Escoriza D., Ruhi A. 2014. Macroecological patterns of amphibian assemblages in the Western Palearctic: Implications for conservation. *Biological Conservation*, 176: 251–261.
- Hijmans R.J. 2020. raster: Geographic Data Analysis and Modeling. R package version 3.3-13. <https://CRAN.R-project.org/package=raster>.
- Kassambara A., Mundt F. 2020. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7. <https://CRAN.R-project.org/package=factoextra>.
- Latitudinal gradients in species diversity. 2023. Wikipedia. Preneseno dne 5. 1. 2023, iz: https://en.wikipedia.org/wiki/Latitudinal_gradients_in_species_diversity.
- Mannion P.D., Upchurch P., Benson R.B.J., Goswami A. 2014. The latitudinal biodiversity gradient through deep time. *Trends in Ecology & Evolution*, 29(1): 42–50.
- Marin J., Hedges S.B. 2016. Time best explains global variation in species richness of amphibians, birds and mammals. *Journal of Biogeography*, 43: 1069–1079.
- O'Donnell M.S., Ignizio D.A. 2012. Bioclimatic predictors for supporting ecological applications in the conterminous United States: U.S. Geological Survey Data Series 691, 10 p.
- R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rosenzweig L.M. 1995. *Species Diversity in Space and Time*. Cambridge University Press. Cambridge.
- Wickham H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. New York.

10. poglavje



**VRSTNO BOGASTVO V
GRADIENTU NADMORSKE
VIŠINE**

Povzetek

10. poglavje: Vrsto bogastvo v gradientu nadmorske višine

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

Vzdolž gradienta nadmorske višine je oblikovan prepoznavni vzorec biodiverzitete. Pogosto je odnos med nadmorsko višino in vrstnim bogastvom negativen, tako da z naraščanjem nadmorske višine vrstno bogastvo upada.

Z višanjem nadmorske višine se hitro spreminjajo ekološki pogoji. Izrazit je gradient sprememb več geofizikalnih in klimatskih dejavnikov: 1. manjša se površina ozemlja; 2. zmanjšuje se zračni tlak vključno s parcialnim tlakom O₂ in CO₂; 3. nižja se temperatura zraka in tal; 4. povečano je sončevo sevanje, zlasti v UV-B delu spektra; 5. krajša se vegetacijska sezona. Ob navedenem se pomembno spreminjajo tudi drugi ekološki dejavniki: relativna zračna vlaga, padavine, hitrost vetra, količina prsti, vsebnost dušikovih spojin v tleh, pH tal, število dni s snežno odejo ipd. Kombinacije vseh dejavnikov praviloma zmanjšujejo primarno produkcijo na višjih nadmorskih višinah, kar spet znižuje biodiverzitetno.

Vzorci biodiverzitete, ki se oblikujejo vzdolž gradientov nadmorske višine, analiziramo s pomočjo korelacijske analize in enostavne linearne regresije. Vpogled v tak tip analize daje primer upadanja števila vrst dvoživk in plazilcev

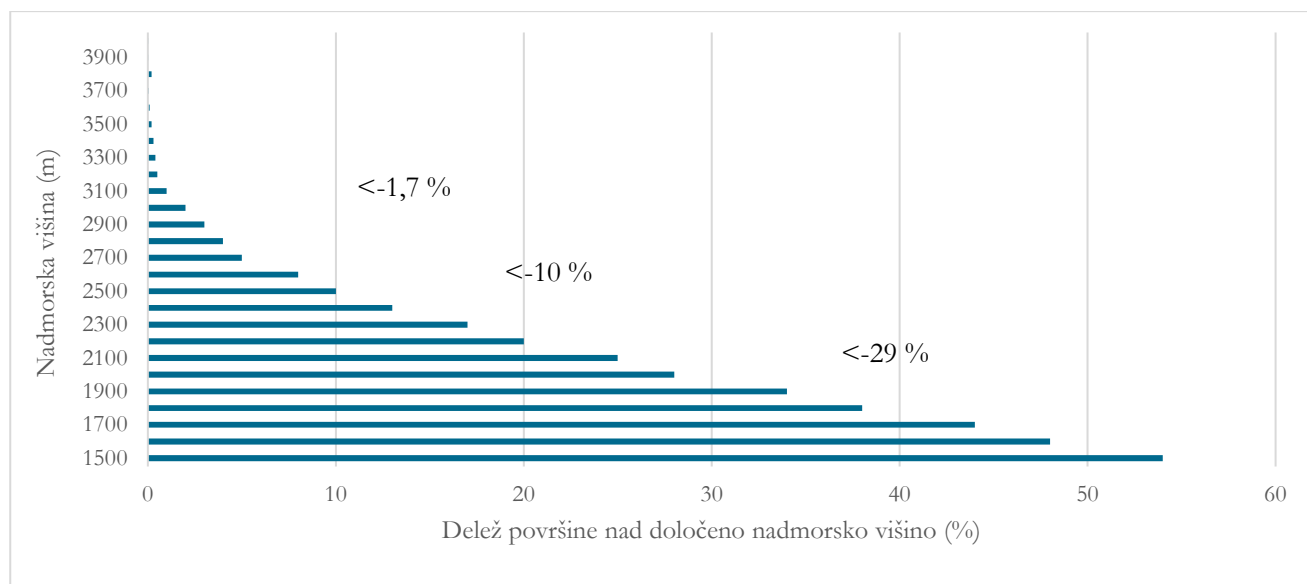
z višanjem nadmorske višine. Obe skupini vretenčarjev sta heterotermni, kar pomeni, da nimata stalne telesne temperature, ampak je ta odvisna od temperature okolice. Nižje povprečne temperature zraka obe skupini živali prizadeneta, zato upadanje števila vrst lahko razložimo s temperaturno hipotezo. Za razlago biodiverzitetnega vzorca s kombinacijo več ekoloških dejavnikov lahko sestavimo kompleksnejši regresijski model.

Druga možnost vrednotenja je primerjanje biodiverzitetnega vzorca z vnaprej izbranim modelom porazdelitve vrstnega bogastva. Huisman-Olff-Fresco tipologija modelov HOF predpostavlja sedem oblik odziva vrst (vrstne pestrosti) na ekološki gradient: I – brez odziva; II – monotoni odziv; III – sigmoidni odziv; IV – optimum; V – asimetrični optimum; VI – dva enakovredna optima; VII – odziv z dvema različnima optimumoma.

Kot modelna skupina pri analizi vzorca višinske razporeditve vrstne diverzitete bodo služili mali sesalci. Domnevamo, da se njihovo vrstno bogastvo vzdolž nadmorskega gradienta sklada z modelom osrednjega območja oz. asimetričnega optima (tip V po Huisman-Olff-Fresco tipologiji).

Vzdolž gradienta nadmorske višine se hitro spreminjajo dejavniki okolja, posledično se spreminja tudi biodiverziteteta. V tem poglavju je predstavljen fenomen odziva vrstnega bogastva v **gradientu nadmorske višine** EDG (ang. Elevational Diversity Gradient). Z višanjem nadmorske višine sta povezana dva pomembna življenjska dejavnika, gradient nižanja temperature zraka in zmanjševanja površine ozemlja. Na splošno se z višanjem nadmorske višine spreminja več geofizikalnih in klimatskih dejavnikov: 1. manjša se površina ozemlja (Slika 10.1); 2. nižja se zračni tlak, prav tako se nižata parcialni tlak O₂ in CO₂; 3. nižja se temperatura zraka; 4. večja se sončevo sevanje in 5. zvišuje se UV-B sevanje; 6. krajša se vegetacijska sezona. Ob navedenih dejavnikih se lahko z nadmorsko višino spreminjajo tudi relativna zračna vlaga, padavine, hitrost vetra, količina prsti, vsebnost dušikovih spojin v tleh, pH tal, število dni s snežno odejo ipd. Kombinacije navedenih dejavnikov vplivajo na zmanjšano primarno produkcijo na višjih nadmorskih višinah, kar spet posledično deluje na porazdelitev živali in vrstno bogastvo. Odnos med biodiverziteteto in nadmorsko višino lahko posplošimo v zakonitost, da z naraščanjem nadmorske višine vrstno bogastvo upada (Korner 2008).

Podoben vzorec upadanja vrstnega bogastva, kot je naveden za nadmorsko višino, lahko opazujemo vzdolž naraščanje zemljepisne širine. Ključna razlika med obema gradientoma je v površini ozemlja, ki se pri nadmorski višini hitro zmanjšuje. Razlikujeta se tudi v prostorski razsežnosti, saj se gradient spreminja z nadmorsko višino na relativno kratkih razdaljah, gradient z zemljepisno širino se spreminja na velikih razdaljah. Pomembne razlike med obema dejavnikoma so tudi v dolžini dneva in v zračnem tlaku.



Slika 10.1 Delež površine nad določeno nadmorsko višino v Alpah. Nekaj več kot 50 % celotne površine Alp (124.035 km² = 100 %) se nahaja nad višino 1.500 m n. v. (Prirejeno iz Korner 2008).

Gradient biodiverzitete vzdolž nadmorske višine razlagamo z več hipotezami, ki pa izhajajo iz gradienta okoljskih spremenljivk. Te hipoteze niso povsem neodvisno ena od druge.

Hipoteza površine (ang. Area Hypothesis) temelji na empiričnem dejstvu, da na večjem območju živi večje število vrst kot na manjšem območju. Z višanjem nadmorske višine se zmanjšuje površina ozemlja, posledično je manjše tudi število vrst. Ta hipoteza ne upošteva vpliva klime in razpoložljivih virov.

Hipoteza padavin (ang. Rainfall Hypothesis) temelji na predpostavki, da diverziteta narašča z naraščanjem padavin, predvsem s količino letnih padavin do 4.000 mm; nad to vrednostjo vrstna pestrost upada. Vzročno-posledična povezava je na relaciji padavine–rodovitnost tal, kar vpliva na primarno produkcijo in diverzitetu.

Hipoteza pestrost virov (ang. Resource Diversity Hypothesis) navaja, da je diverziteta posledica pestrosti virov za preživetje, predvsem prsti in hrane. Viri, ki se nanašajo na hrano, so po navadi odvisni od drugih dejavnikov, npr. padavin in z njimi povezane primarne produkcije. To je že razlaga **hipoteze produktivnosti** (ang. Productivity Hypothesis), po kateri biodiverziteta narašča z naraščanjem produktivnosti.

Po razlagi **temperaturne hipoteze** (ang. Temperature Hypothesis) narašča pestrost vrst z naraščanjem temperature, predvsem zato, ker je s temperaturo povezana tudi produktivnost.

Hipoteza kompeticije (ang. Competition Hypothesis) ima dve razlagi. Po eni od njiju z naraščanjem medvrstne kompeticije narašča število vrst kot posledica specializacije in delitve niš. Po drugi razlagi vodi kompeticija k lokalnemu izumiranju in upadanju vrstnega bogastva. Po tej hipotezi se z naraščanjem nadmorske višine zmanjšuje medvrstna kompeticija.

V biodiverzitetnih raziskavah pogosto analiziramo odziv biodiverzitete vzdolž ekološkega gradienta. Opazovano porazdelitev v naravi primerjamo z vnaprej določenimi statističnimi modeli porazdelitve. Analiza odziva je po eni strani pomembna za napredek ekološke teorije in ima tudi praktični pomen v ekološkem in naravovarstvenem načrtovanju. Informacija o odzivu vrstnega bogastva vzdolž gradienta je osnova za numerično analizo združbe. Oblika krivulje določa lastnosti modela, po katerem je razporejena biodiverziteta, in izpostavi optimalne oz. mejne vrednosti porazdelitve.

V nadaljevanju navajamo tri pristope k analizi vrstnega bogastva v gradientu nadmorske višine: 1. Huisman-Olff-Fresco modeli so statistični oz. regresijski modeli. Z njimi preverjamo, kateri model se najbolje prilega opazovani vrednosti vrstnega bogastva. 2.

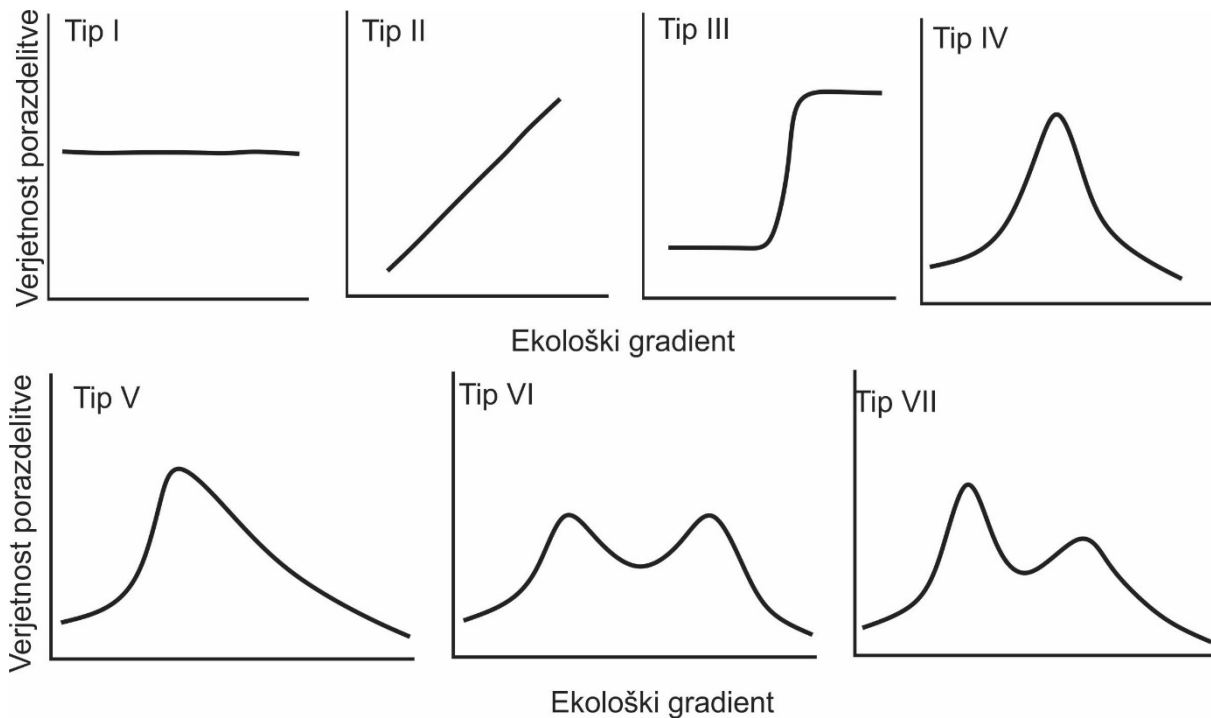
Učinek osrednjega območja je matematični model, ki predvideva največje vrstno bogastvo v osrednjem območju višinskega razpona. 3. Osrednja točka ugodja predvideva unimodalen ekološki gradient, pri katerem je največje vrstno bogastvo v najugodnejši točki gradienta.

Huisman-Olff-Frescojevi modeli HOF. Avtorji, po katerih se imenuje model, so predlagali sedem tipov krivulj (modelov), s katerimi opišemo odziv vrst ali odziv vrstne pestrosti na ekološki gradient: tip I – brez odziva, ravna krivulja, ki je vzporedna abscisi; tip II – monotoni odziv – enakomerno naraščajoča ali enakomerno padajoča krivulja; tip III – sigmoidni odziv, krivulja v obliki črke S, hitra sprememba iz ene stabilne vrednosti v drugo stabilno vrednost; tip IV – optimum, najvišja vrednost odziva je v osrednjem delu gradienta; tip V – asimetrični optimum, najvišja vrednost odziva je v osrednjem delu gradienta, vendar odziv na enem in na drugem kraju ekstrema ni simetričen; tip VI – odziv z dvema enakovrednima optimumoma, krivulja opiše dva enakovredna odziva; tip VII – odziv z dvema različnima optimumoma, krivulja opiše dva odziva, enega večjega in drugega manjšega (Slika 10.2), (Michaelis in Diekmann 2017, Patsiou in sod. 2021).

Učinek osrednjega območja MDE (ang. Mid-Domain Effect). Naraščajočo nadmorsko višino po navadi spremlja monotono upadanja števila vrst. Za nekatere skupine organizmov oz. za nekatera območja je vzorec porazdelitve bogastva vrst drugačen, saj pride do največje zgostitve vrst v srednjem delu območja. Govorimo o učinku osrednjega območja. V središču višinskega gradienta je prisotnih več vrst kot na robnih območjih. Na takšnih opažanjih je izdelan teoretični nevtralni model geometrične porazdelitve vrst. MDE je namreč rezultat naključne razporeditve različnih višinskih razponov večjega števila vrst, ki so na obeh koncih omejeni (s podnožjem in vrhom gore). Oblika krivulje je zato rezultat naključne porazdelitve višinskih razponov, kjer se največ višinskih arealov prekriva v sredini območja. Po tej razlagi drugi dejavniki niso pomembni, vzorec je posledica naključja v porazdelitvi arealov. V modelu MDE s pomočjo simulacije predvidimo teoretični razpon porazdelitve vrst vzdolž prostorskega gradienta. Dejanska opazovanja porazdelitve vrst primerjamo s teoretičnim modelom in ugotavljamo, ali so vrste porazdeljene v skladu z napovedjo modela MDE ali ne (Colwell 2013). Model MDE je identičen tipu IV Huisman-Olff-Frescovih modelov (Slika 10.2).

Osrednja točka ugodja MPA (ang. Midpoint Attractor) je model za razlago vrstnega bogastva, ki združuje unimodalni ekološki gradient in nevtralni model geometrijske porazdelitve (Colwell in sod. 2016). Model MPA predvideva unimodalen ekološki gradient, ki doseže optimum med spodnjo in zgornjo mejo višinskega gradienta. Vrsto bogastvo praviloma sledi ekološkemu gradientu. Izziv v analizi MPA je prepoznavanje unimodalnega ekološkega gradienta in preverjanje odnosa med vrstnim bogastvom in ekološkim gradientom.

Vsebina poglavja je narejena po temeljni literaturi Brown (1995), Rosenzweig (1995) in Elevational diversity gradient (2003).



Slika 10.2 Huisman-Olff-Fresco modeli HOF (tip I–VII) predstavljajo različne oblike odziva vrstnega bogastva ali vrstne pestrosti vzdolž ekološkega gradienta. Tip I – brez odziva; tip II – monotoni odziv (pozitivni ali negativni odnos); tip III – sigmoidni odziv; IV – optimum; V – asimetrični optimum; VI – dva enakovredna optimuma; VII – odziv z dvema različnima optimumoma (prirejeno po Michaelis in Diekmann 2017).

Mantelov test

Mantelov test, poimenovan po biostatistiku Nathanu Mantelu (1919–2002), je statistični test za ugotavljanje korelacije med dvema simetričnima matrikama. Test se pogosto uporablja v ekoloških raziskavah, kjer iščemo povezavo med dvema nizoma podatkov. Npr. prvi niz podatkov je v obliki simetrične matrike podobnosti (indeks beta diverzitete), drugi niz je v obliki simetrične matrike geografskih razdalj med vzorčenimi mesti (razdalje med višinskimi pasovi). Rezultat testa sta korelacijski koeficient in stopnja značilnosti, s čimer je podana jakost povezave med matrikama.



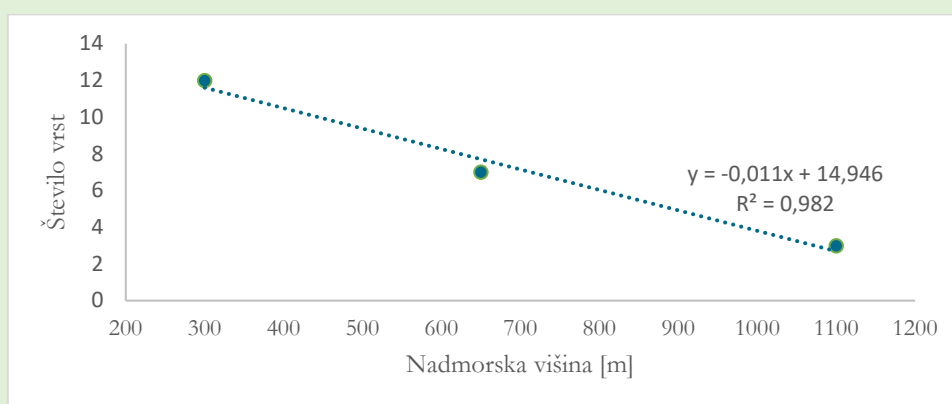
PRIMER 10.1

Dvoživke v gradientu nadmorske višine

Vrstno bogastvo 13 vrst dvoživk vzdolž nadmorskega gradienta na 300, 650 in 1.100 m nadmorske višine na Pohorju je predstavljeno v Preglednici 10.1. S pomočjo enostavne regresijske analize je prikazan odnos med vrstnim bogastvom dvoživk in nadmorsko višino (slika 10.2). Regresijski koeficient je negativen, kar pomeni, da se z višanjem nadmorske višine število vrst dvoživk zmanjšuje. Vrednost regresijskega koeficienta $-0,011$ pomeni, da se na vsakih 100 višinskih metrov število dvoživk zmanjša v povprečju za 1,1 vrsto. Delež razložene variance vrstnega bogastva dvoživk v gradientu nadmorske višine je zelo visok, saj ima determinacijski koeficient vrednost $R^2 = 0,982$, kar pomeni, da model razloži 98,2 % variabilnosti. Torej je ta model uporaben za napovedovanje števila vrst po nadmorskih višinah.

Preglednica 10.1 Seznam vrst dvoživk na Pohorju in njihova prisotnost po višinskih pasovih

| Vrsta | Lokaliteta | Lok A | Lok B | Lok C |
|---------------------|------------------------------|-------|-------|-------|
| Vrsta | Vrsta / NV (m) | 300 | 650 | 1100 |
| Alpski veliki pupek | <i>Triturus carnifex</i> | 1 | 1 | |
| Planinski pupek | <i>Triturus alpestris</i> | | 1 | 1 |
| Navadni pupek | <i>Triturus vulgaris</i> | 1 | 1 | |
| Navadni močerad | <i>Salamandra salamandra</i> | 1 | 1 | |
| Hribski urh | <i>Bombina variegata</i> | 1 | 1 | 1 |
| Česnovka | <i>Pelobates fuscus</i> | 1 | | |
| Navadna krastača | <i>Bufo bufo</i> | 1 | 1 | |
| Zelena krastača | <i>Bufo viridis</i> | 1 | | |
| Zelena rega | <i>Hyla arborea</i> | 1 | | |
| Sekulja | <i>Rana temporaria</i> | 1 | 1 | |
| Rosnica | <i>Rana dalmatina</i> | 1 | | 1 |
| Debeloglavka | <i>Rana ridibunda</i> | 1 | | |
| Zelena žaba | <i>Rana esculenta</i> | 1 | | |



Slika 10.2 Rezultat regresijske analize med številom vrst dvoživk in nadmorsko višino.





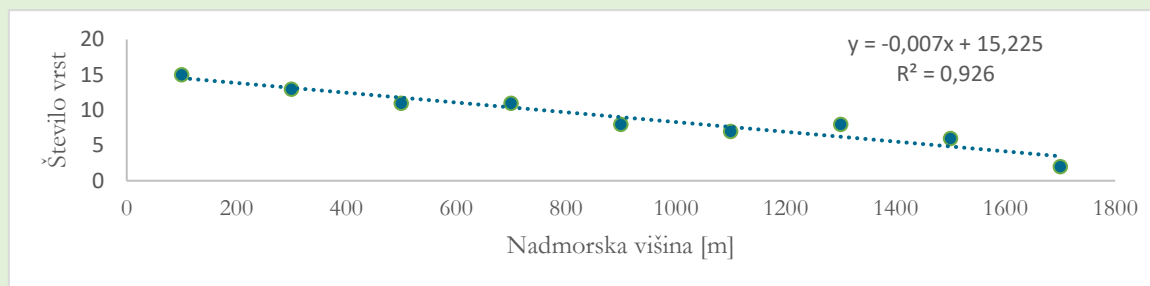
PRIMER 10.2

Plazilci v gradientu nadmorske višine - vrste

Vrstno bogastvo 20 vrst plazilcev vzdolž nadmorskega gradienta od morske obale 0 m do vrha Velebita 1.750 m nadmorske višine je predstavljeno v Preglednici 10.2, podatke smo povzeli po članku Kryštufek in sod. (2008). Fotografiji dveh vrst sta predstavljeni na Slikah 10.4 in 10.5. S pomočjo enostavne regresijske analize je prikazan odnos med vrstnim bogastvom plazilcev in nadmorsko višino (Slika 10.3). Regresijski koeficient je negativen, kar pomeni, da se z višanjem nadmorske višine število vrst plazilcev zmanjšuje. Vrednost koeficienta $-0,007$ pomeni, da se na vsakih 100 višinskih metrov število plazilcev zmanjša v povprečju za 0,7 vrste. Delež razložene variabilnosti vrstnega bogastva plazilcev vzdolž gradienta nadmorske višine je zelo visok, saj ima determinacijski koeficient vrednost $R^2 = 0,925$, kar pomeni, da model razloži 92,5 % variabilnosti. Model je uporaben za napovedovanje števila vrst po nadmorskih višinah.

Preglednica 10.2 Seznam vrst plazilcev na Velebitu in njihova prisotnost po višinskih pasovih

| Vrsta | Razpon višinskega pasu (m) | Povprečje višinskega pasu (m) | | | | | | | | | | | | | | | | | | |
|------------------------|----------------------------------|-------------------------------|---------|---------|---------|----------|-----------|-----------|-----------|-----------|---|---|---|---|--|--|--|--|--|---|
| | | 0-200 | 200-400 | 400-600 | 600-800 | 800-1000 | 1000-1200 | 1200-1400 | 1400-1600 | 1600-1800 | | | | | | | | | | |
| Grška kornjača | <i>Testudo hermanni</i> | 1 | | | | | | | | | | | | | | | | | | |
| Črnopikčasta kuščarica | <i>Algyroides nigropunctatus</i> | 1 | 1 | 1 | 1 | 1 | | | | | | | | | | | | | | |
| Turški gekon | <i>Hemidactylus turcicus</i> | 1 | | | | | | | | | | | | | | | | | | |
| Veliki zelenec | <i>Lacerta trilineata</i> | 1 | 1 | 1 | 1 | | | | | | | | | | | | | | | |
| Zelenec | <i>Lacerta viridis</i> | 1 | | | | | | | | | | | | | | | | | | |
| Velebitska kuščarica | <i>Iberolacerta borvathi</i> | | | | | | | | | | | 1 | 1 | | | | | | | |
| Martinček | <i>Lacerta agilis</i> | | | | | | | | | | | | | 1 | | | | | | |
| Živorodna kuščarica | <i>Zootoca vivipara</i> | | | | | | | | | | | | | 1 | | | | | | |
| Slepec | <i>Anguis fragilis</i> | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | |
| Kraška kuščarica | <i>Podarcis melisellensis</i> | 1 | 1 | 1 | 1 | | | | | | | | | | | | | | | 1 |
| Pozidna kuščarica | <i>Podarcis muralis</i> | | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | | |
| Belica | <i>Coluber gemonensis</i> | 1 | 1 | 1 | | | | | | | | | | | | | | | | |
| Vitka poljarica | <i>Coluber najadum</i> | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | | | | |
| Črnica | <i>Coluber viridiflavus</i> | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | | | | |
| Smokulja | <i>Coronella austriaca</i> | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | | | | |
| Navadni gož | <i>Elaphe longissima</i> | 1 | 1 | | | | | | | | | | | | | | | | | |
| Leopardovka | <i>Elaphe situla</i> | 1 | 1 | 1 | 1 | | | | | | | | | | | | | | | |
| Belouška | <i>Natrix natrix</i> | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | | | | |
| Črnostrel | <i>Telescopus falax</i> | 1 | 1 | | | | | | | | | | | | | | | | | 1 |
| Modras | <i>Vipera ammodytes</i> | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | |



Slika 10.3 Rezultat regresijske analize med številom vrst plazilcev in nadmorsko višino.



PRIMER 10.3

Plazilci v gradientu nadmorske višine – vrstni obrat

Vrstno pestrost plazilcev v gradientu nadmorske višine smo vključili v izračun indeksa beta diverzitete β_w , izračunali smo ga s pomočjo paketa »vegan« (Oksanen in sod. 2019). S katerim smo vrednotili spremembo vrst med višinskimi pasovi (Preglednica 10.3). Sosednji pasovi imajo nizko vrednost indeksa, kar pomeni, da imata sosednja pasova večino skupnih vrst. Vrstni obrat med sosednjima območjema je torej nizek. Z večanjem razdalje med višinskimi pasovi se povečuje vrednost indeksa. To pomeni, da se med oddaljenimi pasovi število skupnih vrst zmanjšuje. Zastavimo raziskovalno vprašanje: Ali je vrstni obrat povezan z razdaljo med pasovi? Ničelna hipoteza H_0 predvideva, da med matrikama ni povezave. Alternativna hipoteza H_a predvideva, da sta matriki povezani. Stopnja zaupanja je pri $\alpha = 0,05$. Povezavo med dvema simetričnima matrikama preverimo z Mantelovim testom, izračunali smo ga s pomočjo paketa »vegan« (Oksanen in sod. 2019).

Matriko vrednosti indeksa beta diverzitete primerjamo z matriko razdalj med višinskimi pasovi (Preglednica 10.4). Mantelov korelacijski koeficient ima vrednost $r = 0,575$, $p = 0,001$. Na podlagi p vrednosti, ki je manjša od stopnje tveganja ($\alpha > p$), zavrnemo ničelno hipotezo in sprejmemo alternativno, po kateri sta matriki statistično značilno povezani. Korelacijski koeficient, ki ima vrednost 0,57, pomeni, da je povezava med matrikama zmerna. Sosednji višinski pasovi imajo bolj podobno vrstno sestavo kot bolj oddaljeni višinski pasovi.

Preglednica 10.3 Vrednosti indeksa beta diverzitete β_w med višinskimi pasovi

| β_w | 100 | 300 | 500 | 700 | 900 | 1100 | 1300 | 1500 | 1700 |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| 100 | 0 | | | | | | | | |
| 300 | 0,143 | 0 | | | | | | | |
| 500 | 0,231 | 0,083 | 0 | | | | | | |
| 700 | 0,308 | 0,167 | 0,091 | 0 | | | | | |
| 900 | 0,478 | 0,333 | 0,263 | 0,158 | 0 | | | | |
| 1100 | 0,545 | 0,400 | 0,333 | 0,222 | 0,067 | 0 | | | |
| 1300 | 0,565 | 0,429 | 0,368 | 0,263 | 0,125 | 0,067 | 0 | | |
| 1500 | 0,810 | 0,789 | 0,765 | 0,647 | 0,571 | 0,538 | 0,429 | 0 | |
| 1700 | 0,765 | 0,733 | 0,846 | 0,846 | 1,000 | 1,000 | 1,000 | 1,000 | 0 |

Preglednica 10.4 Razdalje (m) med višinskimi pasovi

| m | 100 | 300 | 500 | 700 | 900 | 1100 | 1300 | 1500 | 1700 |
|------|------|------|------|------|-----|------|------|------|------|
| 100 | 0 | | | | | | | | |
| 300 | 200 | 0 | | | | | | | |
| 500 | 400 | 200 | 0 | | | | | | |
| 700 | 600 | 400 | 200 | 0 | | | | | |
| 900 | 800 | 600 | 400 | 200 | 0 | | | | |
| 1100 | 1000 | 800 | 600 | 400 | 200 | 0 | | | |
| 1300 | 1200 | 1000 | 800 | 600 | 400 | 200 | 0 | | |
| 1500 | 1400 | 1200 | 1000 | 800 | 600 | 400 | 200 | 0 | |
| 1700 | 1600 | 1400 | 1200 | 1000 | 800 | 600 | 400 | 200 | 0 |

R > package = »vegan«

Izračun β_w indeksa beta diverzitete med višinskimi pasovi

beta <- betadiver(data, "w")

beta

mantel <- mantel(data1, data2) # Mantel test

mantel labs(x="Toplo KGK1 Hladno", y="S Plazilci")



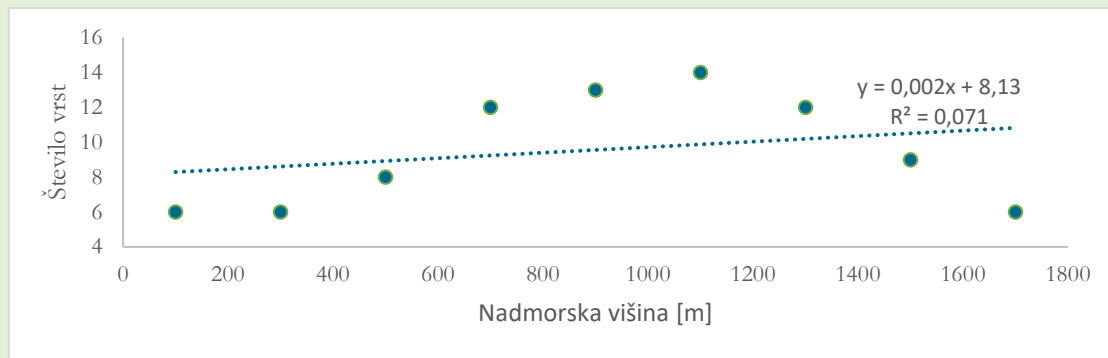
Slika 10.4 Slepec *Anguilla fragilis*.

Foto. Franc Janžekovič.



Slika 10.5 Pozidna kuščarica *Podarcis muralis*.

Foto. Franc Janžekovič.



Slika 10.6 Rezultat linearne regresijske analize med številom vrst in nadmorsko višino.

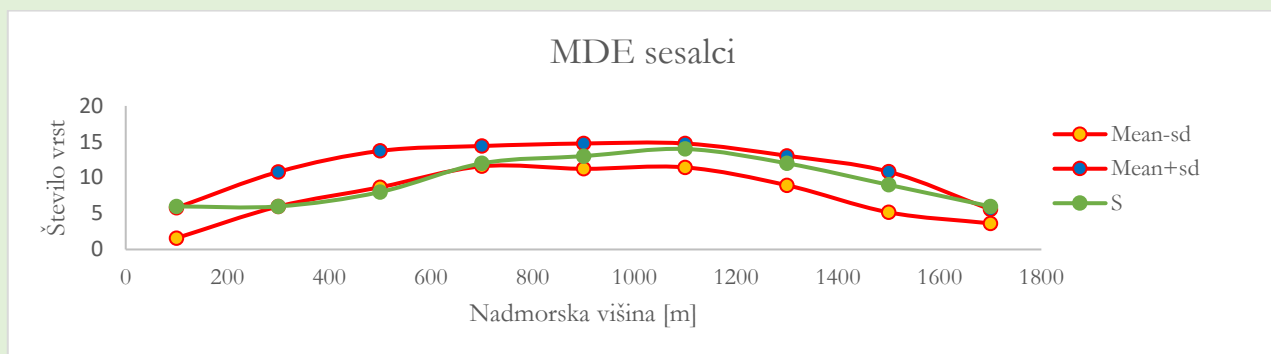


PRIMER 10.5

Mali sesalci v gradientu nadmorske višine – MDE

Vrstno bogastvo malih sesalcev vzdolž nadmorske višine (Preglednica 10.6) na območju Dinaridov poskušajmo razložiti s teoretičnim modelom MDE. Model predvideva, da je najvišje vrstno bogastvo v osrednjem delu gradienta pri nadmorski višini okrog 1.000 m in proti njegovima robovoma upada. MDE model je enakovreden tipu IV HFO modelov. S Slike 10.5 je razvidno, da se dejanska porazdelitev vrst dobro ujema z modelom. Na več mestih krivulja števila vrst dosega mejne vrednosti enega standardnega odklona, vendar jih ne presega bistveno. Zaključimo, da je porazdelitev vrstnega bogastva malih sesalcev v skladu z modelom MDE.

Po primerjavi s HOF tipologijo modelov opazovano krivuljo uvrstimo v tip V (asimetrični optimum). Na Sliki 10.7 je krivulja vrstnega bogastva zvonaste oblike z negativno asimetrijo. S tem smo opisali vzorec biodiverzitetne variabilnosti malih sesalcev vzdolž gradienta nadmorske višine in še vedno nismo razložili dejanskih ekoloških vzrokov za takšno porazdelitev.



Slika 10.7 Dejansko število vrst (zeleno) in teoretični MDE model porazdelitve (rdeča) malih sesalcev po nadmorskih višinah.

Preglednica 10.6 Priprava podatkov za analizo modela MDE s paketom »rangemodelR« v okolju R (Marathe 2019). Min – spodnja meja razširjenosti (povprečje razreda), Max – zgornja meja razširjenosti (povprečje razreda), Range – razpon razširjenosti, Mid – center razširjenosti, N – število razredov

| | Vrsta | Min | Max | Range | Mid | N |
|----------------------|--------------------------------|------|------|-------|------|---|
| Gozdna rovka | <i>Sorex araneus</i> | 500 | 1500 | 1000 | 1000 | 6 |
| Gorska rovka | <i>Sorex alpinus</i> | 700 | 1700 | 1000 | 1200 | 6 |
| Mala rovka | <i>Sorex minutus</i> | 1100 | 1500 | 400 | 1300 | 3 |
| Povodna rovka | <i>Neomys fodiens</i> | 700 | 1100 | 400 | 900 | 3 |
| Močvirska rovka | <i>Neomys anomalus</i> | 700 | 1300 | 600 | 1000 | 4 |
| Vrtna rovka | <i>Crocidura suaveolens</i> | 100 | 900 | 800 | 500 | 5 |
| Poljska rovka | <i>Crocidura leucodon</i> | 100 | 700 | 600 | 400 | 4 |
| Etruščanska rovka | <i>Suncus etruscus</i> | 100 | 100 | 0 | 100 | 1 |
| Gozdna voluharica | <i>Clethrionomys glareolus</i> | 700 | 1700 | 1000 | 1200 | 6 |
| Dinarska voluharica | <i>Dinaromys bogdanovi</i> | 300 | 1700 | 1400 | 1000 | 8 |
| Poljska voluharica | <i>Microtus arvalis</i> | 500 | 1100 | 600 | 800 | 4 |
| Ilirska voluharica | <i>Microtus liechtensteini</i> | 1100 | 1100 | 0 | 1100 | 1 |
| Snežna voluharica | <i>Chionomys nivalis</i> | 1300 | 1700 | 400 | 1500 | 3 |
| Slepo kuže | <i>Nannospalax leucodon</i> | 900 | 1300 | 400 | 1100 | 3 |
| Rumenogrla miš | <i>Apodemus flavicollis</i> | 900 | 1700 | 800 | 1300 | 5 |
| Navadna belonoga miš | <i>Apodemus sylvaticus</i> | 100 | 1300 | 1200 | 700 | 7 |
| Kraška miš | <i>Apodemus epimelas</i> | 100 | 700 | 600 | 400 | 4 |
| Vrtni polh | <i>Eliomys quercinus</i> | 100 | 1300 | 1200 | 700 | 7 |
| Drevesni polh | <i>Dryomys nitedula</i> | 900 | 1500 | 600 | 1200 | 4 |
| Stepska miš | <i>Mus spicilegus</i> | 1500 | 1700 | 200 | 1600 | 2 |



Slika 10.8 Poljska voluharica *Microtus arvalis*.

Foto. Franc Janžekovič

```
R > package = »rangemodelR«
# MDE izračun teoretične porazdelitve
mde <- range_shuffle(x=data, boundary
= 'hh', interval = 200, var = NULL,
sites = seq(100, 1700, 200),
reps =10, degen = FALSE)
mde
```



Zaključki

- Vzdolž gradienta nadmorske višine se oblikuje prepoznaven biodiverzitetni vzorec. Odnos med nadmorsko višino in vrstnim bogastvom je praviloma negativen, kar pomeni, da z naraščanjem nadmorske višine vrstno bogastvo upada.
- Z višanjem nadmorske višine se spreminjajo ekološki pogoji. Gradient geofizikalnih in klimatskih dejavnikov je izrazit: manjša površina ozemlja; nižji zračni tlak; nižji parcialni tlak O_2 in CO_2 ; nižja temperatura zraka; višje sončevo in UV-B sevanje; krajša vegetacijska sezona; viša relativna zračna vlaga; manj padavin, večja hitrost vetra, manj prsti, manj dušikovih spojin v tleh, večje število dni s snežno odejo ipd.
- Neugodne klimatske razmere in kombinacije drugih ekoloških dejavnikov vplivajo na zmanjšano primarno produkcijo na višjih nadmorskih višinah, kar dodatno znižuje biodiverzitetno.
- Biodiverzitetne vzorce, ki se oblikujejo vzdolž gradientov nadmorske višine, analiziramo s korelacijsko analizo in enostavno linearno regresijsko analizo. Navedena pristopa sta primerna za skupine organizmov, pri katerih vrstno bogastvo monotono upada z višjo nadmorsko višino.
- Primera monotonega upadanja števila vrst sta heterotermni skupini dvoživk in plazilcev. Nižje povprečne temperature zraka prizadeneta obe skupini, zato upadanje števila vrst razložimo s temperaturno hipotezo.
- Biodiverzitetni vzorec vrstnega bogastva je lahko skladen s katerim od teoretičnih modelov. Huisman-Olff-Frescova tipologija modelov HOF predpostavlja sedem oblik odziva vrst ali vrstne pestrosti vzdolž ekološkega gradienta: brez odziva, monotoni odziv, sigmoidni odziv, optimum, asimetrični optimum, dva enakovredna optima, odziv z dvema različnima optimumoma. Enakovreden tipu IV po HOF je modelom osrednjega območja MDE.
- Model osrednjega območja MDE predvideva, da je najvišje vrstno bogastvo v osrednjem delu višinskega gradienta in se proti robovoma gradienta zmanjšuje.
- Primer skupine, ki je razporejena po MDE modelu, so homeotermni mali sesalci. Vrstno bogastvo je najvišje v nadmorski višini okrog 1000 m, v nižjih in na višjih nadmorskih višinah število vrst malih sesalcev upada.
- Osrednja točka ugodja MPA je model za razlago vrstnega bogastva, ki predvideva obstoj unimodalnega ekološkega gradienta, ki ima optimum nekje med spodnjo in zgornjo mejo višinskega gradienta. Vrstno bogastvo praviloma sledi ekološkemu gradientu.

Literatura

- Brown J. 1995. *Macroecology*. University of Chicago Press. Chicago.
- Colwell R.K. 2013. EstimateS, Version 9.1: Statistical Estimation of Species Richness and Shared Species from Samples (Software and User's Guide).
- Colwell R.K., Gotelli N.J., Ashton L.A., Beck J., Brehm G., Fayle T.M., Fiedler K., Forister M.L., Kessler M., Kitching R.L., Klimes P., Kluge J., Longino J.T., Maunsell S.C., McCain C.M., Moses J., Noben S., Sam K., Sam L., Shapiro A.M., Wang X., Novotny V. 2016. Midpoint attractors and species richness: Modelling the interaction between environmental drivers and geometric constraints. *Ecology Letters*, 19: 1009–1022.
- Elevational diversity gradient. 2003. Wikipedia. Pridobljeno iz: https://en.wikipedia.org/wiki/Elevational_diversity_gradient, dne 5. 1. 2023.
- Korner C. 2007. The use of 'altitude' in ecological research. *Trends in Ecology and Evolution*, 22(11): 569–575.
- Kryštufek B., Janžekovič F., Režek Donev N. 2008. Elevational diversity of reptiles on two Dinaric mountains. *Journal of natural history*, 42 (5/8): 399–408.
- Marathe A. 2019. rangemodelR: Mid-Domain Effect and Species Richness. R package version 1.0.4. <https://rdrr.io/cran/rangemodelR/>.
- Michaelis J., Diekmann M.R. 2017. Biased niches \pm Species response curves and niche attributes from Huisman-Olff-Fresco models change with differing species prevalence and frequency. *PLoS ONE* 12(8): e0183152. <https://doi.org/10.1371/journal.pone.0183152>.
- Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., Wagner H. 2019. vegan: Community Ecology Package. R package version 2.5-6. <https://CRAN.R-project.org/package=vegan>.
- Patsiou T.-S., Walden N., Willi Y. 2021. What drives species' distributions along elevational gradients? Macroecological and evolutionary insights from Brassicaceae of the central Alps. *Global Ecol Biogeogr*, 00:1–13. <https://doi.org/10.1111/geb.13280>.
- Rosenzweig L.M. 1995. *Species Diversity in Space and Time*. Cambridge University Press. Cambridge.

11. poglavje



FENOTIPSKA PLASTIČINOST

Povzetek

11. poglavje: Fenotipska plastičnost

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

Morfometrija je veda o kvantitativni analizi oblike in velikosti organizmov, kar raziskujemo s statističnimi orodji. Odnose analiziramo z univariatnimi in bivariatnimi (npr. ANOVA, korelacija, regresija) ter multivariatnimi statističnimi metodami (npr. analiza glavnih komponent, diskriminantna analiza). Rezultate številčnih izračunov kombiniramo z grafičnimi predstavitvami odnosov med skupinami (npr. razsevni diagram, okvir z ročaji).

Morfotip opišemo s parametri deskriptivne statistike: aritmetična sredina (\bar{x}), standardna deviacija (StD), minimalna vrednost (Min), maksimalna vrednost (Max) in velikost vzorca (n); v krajši obliki: $\bar{x} \pm StD, Min - Max, n$.

V analizi morfotipov s statističnim sklepanjem preverjamo ali se vzorci v izbranih lastnostih razlikujejo med skupinami (npr. med spoloma, starostnimi razredi, geografskimi vzorci ali taksoni). S *t*-testom preverjamo razlike v izbrani spremenljivki med dvema vzorcema; z ANOVA oz. *F*-testom preverjamo razlike v izbrani spremenljivki med tremi ali več skupinami; z MANOVA preverjamo razlike med dvema ali več skupinami, pri čemer istočasno vključimo dve ali več odvisnih spremenljivk.

Analiza glavnih komponent PCA je multivariatna metoda, namenjena zmanjšanju razsežnosti osnovnih podatkov, kar dosežemo z izračunom glavnih komponent PC. Prva glavna komponenta PC1 poteka skozi multivariatni prostor v smeri, ki razloži največji delež skupne variabilnosti osnovnih spremenljivk. Rezultate PCA prikažemo kot razsevni diagram s projekcijo osebkov na prvi dve komponenti ali pa vrednosti glavnih komponent vključimo v nadaljnje analize. Enakovreden prispevek vseh spremenljivk v multivariatni obravnavi dosežemo s standardizacijo.

Organizem se v procesu rasti večja. Če vsi telesni segmenti rastejo z enako stopnjo, se velikost organizma večja, njegova oblika ostaja nespremenjena. To je izometrična rast, ki je redek

pojav, znan pri malih plazilcih in repatih dvoživkah. Pri večini živali posamezni telesni segmenti rastejo z različno hitrostjo, kar privede do spremembe oblike. Takšna rast je alometrična. Z rastjo se večja volumen organizma s tretjo potenco v^3 , površina se večja s kvadratom dimenzije p^2 , kar neposredno vpliva na fiziološke in mehanske lastnosti organizma. Razlikujemo tri vrste alometrije: statična, ontogenetska in evolucijska alometrija.

Fenotip je celotna kombinacija lastnosti organizma, ki zajema vse ravni organizacije ob anatomskih in morfoloških tudi fiziološke lastnosti, njegovo vedenje in rezultate njegovih aktivnosti. Fenotip je izraz genskega zapisa, vpliva okolja in medsebojnega vpliva med genetskimi dejavniki in dejavniki okolja.

Fenotipska variabilnost je pestrost v obliki in velikosti med posamezniki populacije in je skupaj z genetsko variabilnostjo temelj za evolucijo s selekcijo.

Klina je postopno spreminjanja izbrane lastnosti vzdolž geografskega transekt. Nastane, kjer nasprotno delujeta selekcija in genski pretok. Klina se razvije bodisi s primarno diferenciacijo zaradi prostorske heterogenosti okolja in prilagajanja morfotipa ekološkemu gradientu bodisi s sekundarno diferenciacijo kot posledica ponovnega stika predhodno izoliranih populacij, ki sta diferencirali v alopatriji.

Odzivi organizmov na okoljske gradiente so praviloma podobni. Na tej osnovi so se oblikovala pravila, ki jih s skupnim imenom imenujemo biološka pravila (ekogeografska pravila, makroekološka pravila), ki so posplošitve in poenostavljene razlage pojavov pri odzivu organizmov na ekološke in evolucijske dejavnike, npr. Bergmannovo pravilo.

Morfometrija (ang. Morphometrics, Morphometry) je metodološki pristop kvantitativne analize oblike in velikosti organizmov. Morfometrično analizo lahko izvajamo na celih organizmih, posameznih segmentih telesa, organih, tkivih ali subceličnih strukturah. Vidiki obravnave so raznoliki biološki in ekološki fenomeni, od spremljanja velikosti in oblike telesa v procesu ontogenetskega razvoja, morfometričnega odziva organizma na okoljske dejavnike, kovariabilnosti z genetskimi dejavniki, do vrednotenja morfometričnih sprememb v obliki in velikosti v evolucijskem procesu. Glavni namen morfometričnih raziskav je s statističnimi orodji preverjati hipoteze o dejavnikih, ki vplivajo na obliko in velikost organizma.

Na splošno razlikujemo tri metodološke pristope: tradicionalna morfometrija, geometrijska morfometrija in orisna morfometrija. Tradicionalna morfometrija (ang. »Traditional« Morphometrics) temelji na analiziranju enorazsežnih meritev, predvsem dolžin, mase in površine organizmov. Vrednost meritve je podana z merilno skalo, npr. telesna dolžina je 250 mm. Geometrijska morfometrija (ang. Landmark-Based Geometric Morphometrics) zajema dvo dimenzionalne 2D ali tri dimenzionalne 3D meritve, meritev je položaj izbrane anatomske strukture v prostoru. Vrednost 2D meritve je koordinata točke (X, Y) , npr. $x = 3,4$, $y = 25,5$. Orisna morfometrija (ang. Outline Analysis) zajema 2D zaporedje meritev oboda anatomske strukture, pri kateri je meritev zaporedje koordinat, npr. (x_1, y_1) , (x_2, y_2) , ..., (x_n, y_n) .

Za preverjanje hipotez o dejavnikih, ki vplivajo na obliko in velikost organizmov, so najprimernejše multivariatne statistične metode. Veliko informativno vrednost ima slikovno gradivo (npr. razsevni diagram, poligon, okvir z ročaji, itd.), za kar moramo večdimenzionalne podatke reducirati v samo dve ali eno dimenzijo. Analiza glavnih komponent je najpogosteje uporabljena ordinacijska metoda za zmanjšanje števila dimenzij v samo dve ali tri glavne komponente. Rezultate predstavimo bodisi z dvodimenzionalnim razsevnim diagramom, bodisi vrednosti glavnih komponent vpeljemo v nadaljnje statistične analize.

V tem poglavju so predstavljene vsebine:

- Opis morfotipa in opisna statistika.
- Analiza morfotipa.
 - Preverjanje razlik med skupinami.
 - Redukcija razsežnosti.
- Alometrija: statična, ontogenetska, evolucijska.
- Fenotipska plastičnost.
 - Fenotip
 - Fenotipska variabilnost
 - Klina
- Ekogeografska pravila.

11.1 Opis morfortipa in opisna statistika

Opis morfortipa se nanaša na podajanje dimenzij in lastnosti preučevanega objekta. Podamo jih z navajanjem povprečij, variabilnosti in razpona. Navedene parametre uvrščamo na področje opisne statistike. V opisna statistiki (sin. deskriptivna statistika; ang. Descriptive Statistics) kvantitativno opišemo lastnosti zbranih meritev, vrednosti povzemamo in organiziramo na način, da so pregledne, razumljive in informativne.

Parametri opisne statistike. **Velikost vzorca n** je število statističnih enot, ki jih zajamemo v raziskavi. **Aritmetična sredina \bar{x}** je definirana kot kvocient med vsoto vseh vrednosti in skupnim številom enot v vzorcu in je najpogosteje uporabljena mera srednje vrednosti. **Minimum Min** je najnižja izmerjena vrednost. **Maksimum Max** je najvišja izmerjena vrednost. **Standardna deviacija StD** je mera razpršenosti porazdelitve vrednosti. **Varianca Var** je povprečje vsote kvadriranih odklonov od aritmetične sredine. **Sploščenost porazdelitve** (ang. Kurtosis) je mera sploščenosti. S tem koeficientom preverjamo ali je frekvenčna porazdelitev spremenljivke koničasta (če je vrednost koeficienta večja od 0), normalno sploščena (če je vrednost koeficienta enaka 0) ali sploščena (če je vrednost koeficienta manjša od 0). **Asimetričnost porazdelitve** (ang. Skewness) je mera asimetrije. S tem koeficientom preverimo porazdelitev spremenljivke, ki je lahko asimetrična v desno oz. ima pozitivno asimetrijo (vrednost koeficienta večja od 0), simetrična (če je vrednost koeficienta enaka 0) ali asimetrična v levo, kadar ima negativno asimetrijo (če je vrednost koeficienta manjša od 0).

V strokovni in znanstveni literaturi je praksa, da se vrednosti opisne statistike podajo z navajanjem: aritmetične sredine, standardne deviacije, minimuma, maksimuma in velikost vzorca; seveda dodamo tudi mersko enoto: $\bar{x} \pm StD, Min - Max, n$.

11.2 Analiza morfortipa

Pri analiziranju morfortipov preverjamo razlike med lastnostmi organizmov bodisi glede na skupine, iz katerih le-ti izhajajo, ali glede na ekološke gradiente, v katerih živijo. Skupine, med katerimi preverjamo razlike, oblikujemo v skladu z izhodiščem in vidikom raziskave, npr. glede na spol, starostni razred, geografski vzorec ali taksonomsko uvrstitev. Podatke praviloma sestavljata dva sklopa podatkov, prvi sklop podatkov so intervalne spremenljivke $X_1, X_2, X_3, \dots, X_p$, s katerimi so podane morfometrične lastnosti objekta oz. meritve proučevanega organizma. Drugi sklop podatkov sestavljajo nominalne spremenljivke K_1, K_2, \dots, K_p , s katerimi so podane kategorije za oblikovanje skupin (npr. sistematska uvrstitev, geografska lokaliteta, spol ipd.). Urejeni podatki imajo obliko pravokotne matrike velikosti $m \times n$, v kateri so vrstice posamezni objekti oz. osebki organizmov na katerih so izvedene meritve, stolpci so spremenljivke:

| Indikator | Nominalne spr. | | Intervalne spremenljivke | | | | |
|-----------|----------------|----------|--------------------------|----------|----------|-----|----------|
| | K_1 | K_2 | X_1 | X_2 | X_3 | ... | X_p |
| Id_1 | k_{1j} | k_{2b} | x_{11} | x_{12} | x_{13} | ... | x_{1p} |
| Id_2 | k_{1j} | k_{2b} | x_{21} | x_{22} | x_{23} | ... | x_{2p} |
| Id_3 | k_{1j} | k_{2b} | x_{31} | x_{32} | x_{33} | ... | x_{3p} |
| Id_4 | k_{1g} | k_{2b} | x_{41} | x_{42} | x_{43} | ... | x_{4p} |
| Id_5 | k_{1g} | k_{2b} | x_{51} | x_{52} | x_{53} | ... | x_{5p} |
| Id_6 | k_{1g} | k_{2c} | x_{61} | x_{62} | x_{63} | ... | x_{6p} |
| Id_7 | k_{1g} | k_{2c} | x_{71} | x_{72} | x_{73} | ... | x_{7p} |
| ... | ... | ... | ... | ... | ... | ... | ... |
| Id_p | k_{1p} | k_{2p} | x_{p1} | x_{p2} | x_{p3} | ... | x_{pp} |

Prvi stolpec je indikator ID z inventarno številko osebk, na katerem so izvedene meritve, po navadi gre za muzejsko inventarno številko. Indikator je pomembna informacija za sledljivost in možnost naknadnega preverjanja meritev in ni namenjen statističnemu vrednotenju.

11.2.1 Preverjanje razlik med skupinami – testna statistika

Razlike med aritmetičnima sredinama dveh neodvisnih vzorcev preverimo s ***t*-testom**. Test je ustrezen tudi za preverjanje razlik pri majhnih vzorcih ($n < 30$). Testna statistika sledi Studentovi porazdelitvi oz. *t* porazdelitvi ob predpostavkah, da so vrednosti normalno porazdeljene, da imata vzorca enako varianco in da so podatki zbrani neodvisno in naključno.

ANOVA (ang. One-way Analysis of Variance) oz. ***F*-test** je metoda za preverjanje razlike med aritmetičnimi sredinami več neodvisnih vzorcev. Testna statistika sledi *F* porazdelitvi ob predpostavkah, da so vrednosti normalno porazdeljene, da imata vzorca enako varianco in da so podatki zbrani neodvisno in naključno.

MANOVA (ang. Multivariate Analysis of Variance) je metoda za preverjanje razlike med aritmetičnimi sredinami več neodvisnih vzorcev s hkratno obravnavo več odvisnih spremenljivk. Testna statistika sledi *F* porazdelitvi ob predpostavkah, da so vrednosti normalno porazdeljene, da imajo vzorci podobno varianco in da so podatki zbrani neodvisno in naključno.

11.2.2 Redukcija razsežnosti – analiza glavnih komponent

Analiza glavnih komponent PCA (ang. Principal Component Analysis) omogoča zmanjšanje več razsežne podatkovne matrike. S pomočjo PCA pretvorimo osnovne spremenljivke v nove spremenljivke imenovane glavne komponente PC (ang. Principal Components), njihovo število je enako številu osnovnih spremenljivk. Značilnost glavnih komponent je orientiranost v multivariatnem prostoru, prva glavna komponenta PC1 poteka skozi multivariatni prostor v smeri, ki razloži največji delež skupne variance

osnovnih spremenljivk. Druga glavna komponenta PC2 je pravokotna na PC1 in razlaga naslednji skupni delež variance, sledijo preostale PC. Rezultate PCA po navadi prikažemo kot razsevni diagram s projekcijo osebkov (objektov) na prvi dve ali prve tri komponente. Vrednosti glavnih komponent (ang. Scores) lahko vključimo v nadaljnje analize oz. testiranja.

Glavne komponente povzemajo informacijo o lastnostih posamezne osnovne spremenljivke, med seboj pa so neodvisne. Za razumevanje rezultatov je pomembna interpretacija vsebine posamezne glavne komponente, ki jo preverimo s korelacijsko analizo med osnovnimi spremenljivkami in komponentami. Močna korelacija med izbrano glavno komponento in osnovno spremenljivko pomeni, da je večina informacije osnovne spremenljivke povzeta v obravnavani komponenti. Če je korelacija pozitivna, interpretiramo, da z naraščanjem vrednosti komponente opisujemo tudi naraščanje vrednosti lastnosti izbrane osnovne spremenljivke, in obratno, če je korelacija negativna, potem z naraščanjem vrednosti komponente opišemo nižanje vrednosti spremenljivke (dodatna razlaga analize glavnih komponent je podana v 12. Poglavju).

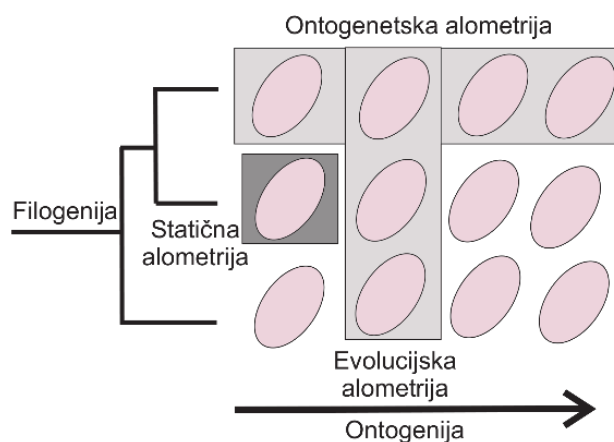
11.3 Alometrija: statična, ontogenetska, evolucijska

Organizem se v procesu rasti večja. Enakomerna (proporcionalna) rast vseh telesnih segmentov, privede do izometrične rasti. Z drugimi besedami, če vsi telesni segmenti rastejo z enako stopnjo, se organizem večja, oblika organizma pa ostaja nespremenjena. Takšen tip rasti – izometrična rast, je redek pojav, prisoten pri nekaterih vrstah malih plazilcev (nekateri kuščarji) in dvoživk (pupki). Pri večini živali posamezni telesni segmenti rastejo z različno hitrostjo. Ne proporcionalna (alometrična) rast telesnih segmentov privede do razlik v obliki, to je do alometrije.

Z rastjo celotnega telesa, oz. rastjo posameznih telesnih segmentov, se večja volumen organizma s tretjo potenco linearne dimenzije v^3 , medtem se površina večja s kvadratom dimenzije p^2 . V procesu rasti prihaja do spremembe razmerja med površino in volumnom, kar neposredno vpliva na fiziološke in mehanske lastnosti organizma. Npr. v prehrani in absorpciji snovi na površinah endodermalnega epitela, pri dihanju in izmenjavi dihalnih plinov na respiratornih površinah, oskrbi celic, in s tem preseku krvnih žil, skeletni opori s presekom mišic in debelino kosti, izločanju s povečanjem ledvične skorje (npr. brazdaste ledvice goveda).

Koncept alometrije se nanaša na kovariabilnost telesnih segmentov. V tradicionalni in geometrijsko morfometriji obliko določamo kot razmerja med telesnimi segmenti ob predhodni izločitvi vpliva velikosti.

Razlikujemo tri vrste alometrije: statična, ontogenetska in evolucijska (Slika 11.1). **Statična alometrija** se nanaša na variabilnost oblike organizma na izbrani stopnji ontogenetskega razvoja. Podatki predstavljajo »povprečne preseke« izbrane starostne kategorije (ang. Cross-sectional Data). **Ontogenetska alometrija** se nanaša na spremembe velikosti in oblike organizma tekom osebnega razvoja. Rast je vir variabilnosti, zato meritve osebkov izvajajo daljše časovno obdobje, ki lahko traja več let (ang. Longitudinal Data). **Evolucijska alometrija** se nanaša na alometrijo evolucijskih skupin oz. filogenetskih linij.



Slika 11.1 Razmerje med statično, ontogenetsko in evolucijsko alometrijo (prirejeno po Ivanović in Kalezić, 2009).

11.4 Fenotipska plastičnost

Fenotip (ang. Phenotype) je celotna kombinacija vidnih znakov in značilnosti posameznega organizma. Termin zajema vse ravni organizacije, zgradbo in morfološki izgled osebkov, njegove razvojne procese, njegove biokemijske in fiziološke lastnosti, njegovo vedenje in rezultate njegovih aktivnosti.

Fenotip je rezultat: 1. izražanja genskega zapisa; 2. vpliva okolja, tako na stopnjo morfogeneze v procesu ontogenetskega razvoja, kot na metabolno-fiziološke, imuno-hormonalne in etološke procese tekom odrasle stopnje življenja; 3. medsebojnega vpliva med genetskimi dejavniki in dejavniki okolja; 4. nenapovedljivih naključnih dogodkov. Sestavina fenotipa so tudi rezultati aktivnosti, ki jih izvajajo osebkovi, npr. ptičje gnezdo, bobrov jez, krtina krta ali satovje čebele. Nekateri avtorji govorijo o razširjenem fenotipu (ang. Extended Phenotype). Vrsta je **polimorfna** (sin. polifena), kadar v populaciji obstajata dva ali več diskretnih fenotipov.

Fenotipska variabilnost (ang. Phenotypic Variation) je pestrost v obliki in velikosti (in vseh drugih znakih) predstavnikov populacij izbrane vrste. Fenotipska variabilnost je skupaj z genetsko variabilnostjo temelj za evolucijo z naravnim izborom oz. selekcijo. Fenotip v procesu razmnoževanja prenese dednino na naslednjo generacijo (ali pa ne). To

pomeni, da uspešnost fenotipa v naravnem izboru posredno vpliva na genetsko strukturo populacije v naslednji filetski liniji. Z drugimi besedami, naravni izbor neposredno deluje na fenotip in prek njega posredno na genotip. Uspeh fenotipa, in s tem uspeh genotipa, je izražen v deležu dedne informacije v naslednji generaciji. Posplošimo lahko, da naravni izbor deluje na nivoju fenotipa, zato brez fenotipske variabilnosti ni naravnega izbora oz. selekcije. Genotipi so po navadi prožni pri nastajanju modifikacij in realizaciji fenotipa. Pri večini organizmih se fenotipi lahko zelo razlikujejo neodvisno od genetskih dejavnikov. Z drugimi besedami, predstavniki iste genetske linije se lahko zelo razlikujejo v fenotipih. V teh primerih govorimo o ekofenotipski variabilnosti ali fenotipski plastičnosti.

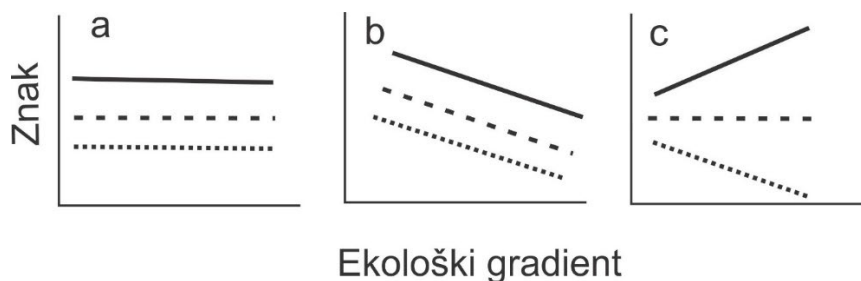
Fenotipska plastičnost (ang. Phenotypic Plasticity) je lastnost populacije in vrste, da se osebki ali populacije med seboj razlikujejo v fenotipu. Fenotipska plastičnost zajema različne vidike prilagoditev, s katerimi se organizmi odzivajo na dejavnike. Prilagoditve so: morfološke, fiziološke, vedenjske (etološke) in fenološke (prisotnost oz. odsotnost organizma glede na periodičnost življenjskega cikla in kako nanj vpliva sezona oz. menjava letnih časov).

Genetska variabilnost in okoljski dejavniki sta glavna vira morfološke variabilnosti. Ob njiju na razvoj osebkov in njegovo velikost ter obliko delujejo tudi nenapovedljivi in naključni dogodki, ki so:

- Osebna variabilnost osebkov, kot posledica intrističnih organizmovih dejavnikov v procesu ontogenije od spočetja, preko embrionalnega razvoja, spolnega dozorevanja in staranja (npr. usklajeno delovanje hormonalnega, imunskega in živčnega sistema).
- Razvojna nestabilnost, kot posledica stohastičnih nihanj v procesu razvoja. Vzrok zanjo so lahko abiotski dejavniki okolja (npr. polutanti), biotski dejavniki (npr. okužbe, paraziti, socialni stres), genetske nepravilnosti.
- Razvojni šum, ki je komponenta variabilnosti fenotipa in je ne moremo pojasniti z nobenih od prej navedenih dejavnikov.

Za celoviti vpogled v morfološko variabilnost in za razumevanje koncepta fenotipske plastičnosti moramo vpeljati dodatne elemente. Ne odraža se vsak vpliv (genetski, okoljski, naključni) na fenotip. Odsotnost fenotipske spremembe ob delovanju relevantnega dejavnika na organizem, je pokazatelj, da ima razvojni sistem posameznika (ali skupine) prožnost in zmožnost nevtralizacije zunanjega ali notranjega vira vplivanja. Ta biološka značilnost je privedla do razvoja koncepta, imenovanega razvojna stabilnost (ang. Development Stability), robustnost (ang. Robustness), elastičnost (ang. Buffering) ali kanalizacija (ang. Canalization). Kljub deljenim mnenjem strokovnjakov, o točnih definicijah navedenih pojmov jih lahko strnemo, da gre za lastnost organizma, ki ima sposobnost nevtralizirati vplive in ohraniti nespremenjeni fenotip. Vpeljimo še pojem

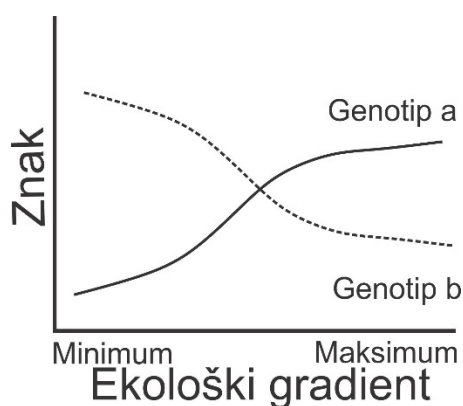
ciljni fenotip (ang. Target Phenotyp), ki je konstrukt in napoveduje obliko in velikost osebkov istega genotipa, v enakem okolju in brez naključnih vplivov. V kontekstu ciljnega fenotipa je fenotipska plastičnost članov istega genotipa posledica razvoja v različnem okolju (Klingenberg 2019).



Slika 11.2 Fenotipska plastičnost je definirana kot sposobnost organizmov, da znotraj istega genotipa realizirajo več kot en fenotip, kot odziv na izpostavljenost ekološkemu gradientu ali naključnim vplivom. Posamezna oblika črte predstavlja reakcijsko normo različnih znakov v odvisnosti od ekološkega gradienta: a znaki so brez fenotipske plastičnosti in se ne odzivajo na ekološki gradient. b znaki izkazujejo fenotipsko plastičnost in se vsi na enak način odzivajo na ekološki gradient. c fenotipska plastičnost znaki se na različne načine odzivajo na ekološki gradient (prirejeno po Phenotypic plasticity 2019).

Usmerjenost in jakost fenotipskega odziv organizmov (iz enotnega genotipa) vzdolž okoljskega gradienta, imenujemo **reakcijska norma** (Slika 11.2). Različni genotipi lahko v ekološkem gradientu izkazujejo podobno ali različno smer in jakost odziva (Sliki 11.2 in 11.3).

Morfoprostor (ang. Morphospace) tudi morfometrični prostor je matematični način vizualizacije lastnosti fenotipa glede na eno ali več morfometričnih spremenljivk. Morfoprostor definirajo spremenljivke, s katerimi so opisane izbrane lastnosti morfotipa. Variabilnost skupine (vrste, geografskega vzorca ipd.) je grafično predstavljena s poligonom, npr. minimalni konveksni poligon (ang. Convex Hull) in posamezne točke na grafu predstavljajo razporeditev osebkov ali povprečje skupine (npr. Slika 12.3).



Slika 11.3 Fenotipski odziv genotipa *a* in genotipa *b* vzdolž ekološkega gradienta (povzeto po Phenotypic plasticity 2019).

Klina (ang. Cline) je merljivi gradient postopnega spreminjanja izbrane lastnosti v populaciji oz. biološkega znaka vzdolž geografskega območja. Klina se nanaša na variabilnost izbrane lastnosti ali več njih v prostorskem gradientu. Večkratne kline so lahko enako usmerjene, vendar ni nujno, da je odziv večkratnih klin usklajen. Dodatno kompleksnost vnaša ugotovitev, da so kline lahko nasprotno usmerjene ali tudi neodvisne ena od druge (Slika 11.3).

Klina je rezultat dveh nasprotujočih si dejavnikov: selekcije in genskega pretoka (migracij). Posledica selekcije je prilagajanje na lokalno okolje, izraženo v različnih fenotipih (oz. genotipih), ki so favorizirani v gradientu ekoloških razmer. Nasproti razhajanju zaradi prilagajanja lokalnim razmeram deluje genski pretok, ki ima homogenizacijski učinek in preprečuje speciacijo. Izmenjava dednine med populacijami preprečuje vzpostavitev reproduktivne pregrade.

Do razvoja klina lahko pride po dveh poteh, pri čemer so mehanizmi nastanka različni: primarna diferenciacija (primarni stik) in sekundarna diferenciacija (sekundarna introgresija). Primarna diferenciacija je posledica prostorske heterogenosti okolja. Areal vrste praviloma obsega več okoljskih gradientov (npr. temperaturni, padavinski, vegetacijske tipe ipd.), kar privede do različnih selekcijskih pritiskov, ki favorizirajo različne morfotipe. V primarni diferenciaciji, okoljski gradient povzroči postopno spreminjanje fenotipske lastnosti v prostoru. Sekundarna diferenciacija nastane kot posledica ponovnega stika predhodno izoliranih populacij, ki sta se diferencirali v alopatriji (geografsko ločeno). Z vzpostavitvijo prehodnega območja se v območju stika obeh populacij razvije postopna sprememba fenotipske lastnosti. Mehanizem klina v sekundarnem scenariju je posledica ponovnega združevanja genomov dveh populacij, ki se kaže v vmesnih fenotipih.

Organizacija in analiza podatkov. Za ugotavljanje povezanosti med morfotipom in ekološkim gradientom podatke organiziramo na način, ki omogoča izračun povprečja morfometričnih spremenljivk glede na skupine k_{ij} , k_{ig} , ..., k_{ip} , ki smo jih določili v prvem koraku. Če smo morfometrične spremenljivke predhodno vključili v analizo glavnih komponent, potem izračunamo povprečja vrednosti glavnih komponent za posamezno skupino. Pravokotna matrika dimenzije $m \times n$ ima toliko vrstic, kot je skupin k_{ip} (i je skupina, p je število skupin) in toliko stolpcev, kot je morfometričnih spremenljivk oz. glavnih komponent PC1 in PC2 (po navadi vključimo prvo ali prvi dve), ter ekoloških spremenljivk E_1, E_2, \dots, E_p , kot sledi:

| Skupina | Morfometrične spremenljivke | | | Ekološke spremenljivke | | | |
|----------|-----------------------------|----------------|----------------|------------------------|----------|-----|----------|
| | PC1 | PC2 | PCn | E_1 | E_2 | ... | E_p |
| k_{1j} | \bar{x}_{j1} | \bar{x}_{j2} | \bar{x}_{jn} | y_{j1} | y_{j2} | ... | y_{jp} |
| k_{1g} | \bar{x}_{g1} | \bar{x}_{g2} | \bar{x}_{gn} | y_{g1} | y_{g2} | ... | y_{gp} |
| ... | ... | ... | ... | ... | ... | ... | ... |
| k_{1p} | \bar{x}_{p1} | \bar{x}_{p2} | \bar{x}_{pn} | y_{p1} | y_{p2} | ... | y_{pp} |

Analizo fenotipske plastičnosti izvajamo na dveh nivojih. Prvi nivo analiz vključuje izračun jakosti fenotipske plastičnosti, torej oceno variabilnosti znaka in primerjanje variabilnosti med skupinami (npr. med populacijami iz različnega okolja). Primerjave izvajamo zgolj z razvrščanjem po zaporedju ali z uporabo statističnih testov (npr. t -test pri primerjanju dveh skupin, F -test pri primerjanju treh ali več skupin). Drugi nivo obsega analizo smeri in stopnjo odziva reakcijske norme v ekološkem gradientu. Slednji metodološki pristop vključuje korelacijsko analizo in različne tipe regresijskih analiz, kjer smer in jakost odziva v gradientu ekoloških dejavnikov, podamo z regresijskim koeficientom.

Za izračun jakosti fenotipske variabilnosti je v uporabi več cenilk, nekaj osnovnih povzemamo po Valladares in sod. (2006).

Koeficient variacije CV (ang. Coefficient of Variation) je najpogostejša cenilka za predstavitev morfometrične variabilnosti, računamo ga po formuli:

$$CV = \frac{StD}{\bar{x}} \quad \text{oz.} \quad CV = 100 \frac{StD}{\bar{x}}$$

Koeficient variacije, ki temelji na povprečjih **CV_m** (ang. Coefficient of Variation, Based on Means) je omejen na primerjave med vrstami. Posamezni skupini izračunamo povprečje, iz povprečij izračunamo standardno deviacijo $StD_{\bar{x}}$ in skupno povprečje $\bar{\bar{x}}$. Koeficient izračunamo po formuli:

$$CV_m = \frac{StD_{\bar{x}}}{\bar{\bar{x}}}$$

Cenilka fenotipske plastičnosti PIV, ki temelji na maksimalnem $Max_{\bar{x}}$ in minimalnem povprečju $Min_{\bar{x}}$ in jo računamo po formuli:

$$PIV = \frac{Max \bar{x} - Min \bar{x}}{Max \bar{x}}$$

Cenilka relativne razdalje v plastičnosti RDPI (ang. Relative Distances Plasticity Indeks) računamo kot povprečje relativnih razdalj med pari naključnih osebkov. Cenilka je primerna za primerjanje morfološke variabilnosti med vrstami, ki so sestavljene iz več skupin in izpostavljene različnim ekološkim gradientom. Primer izračuna za en znak za eno vrsto, s tremi skupinami S_1 , S_2 in S_3 iz različnega okolja izvedemo v naslednjem postopku. Izračunamo razdalje d med naključnimi osebki iz prve in druge skupine, iz prve in tretje skupine, ter iz druge in tretje skupine:

$$d_{x1 \rightarrow 2} = |x_{i1} - x_{i2}|, d_{x1 \rightarrow 3} = |x_{i1} - x_{i3}|, d_{x2 \rightarrow 3} = |x_{i2} - x_{i3}|.$$

Nato razdaljo relativiziramo z vsoto:

$$\frac{d_{x1 \rightarrow 2}}{x_{i1} + x_{i2}}, \frac{d_{x1 \rightarrow 3}}{x_{i1} + x_{i3}}, \frac{d_{x2 \rightarrow 3}}{x_{i2} + x_{i3}}.$$

Vse relativne razdalje seštejemo in delimo s skupnim številom razdalj n :

$$RDPI = \frac{1}{n} \sum \frac{d_{in \rightarrow m}}{x_{in} + x_{im}}$$

Cenilka RDPI ima nabor vrednosti od 0 do 1, višja vrednost pomeni večjo variabilnost.

Cenilka poenostavljene relativne razdalje v plastičnosti RDPIs (ang. Simplified Relative Distances Plasticity Indeks) je poenostavitev cenilke RDPI. V primeru z velikim številom skupin in več vrstami se računanje cenilke zelo komplicira. V izogib zamudnemu računanju lahko izračun naredimo samo na povprečjih skupin. To pomeni, da izračunamo povprečja skupin, nato razdaljo med njimi in jo relativiziramo, po formuli:

$$RDPI_s = \frac{1}{n} \sum \frac{d_{\bar{x}n \rightarrow m}}{\bar{x}_{in} + \bar{x}_{im}}$$

Cenilka okoljsko standardizirane plastičnosti ESPI (ang. Environmentally Standardized Plasticity Indeks) izračunamo kot razliko med maksimalno in minimalno vrednostjo znaka X in jo nato relativiziramo z razdaljo okoljskega gradienta E , po formuli:

$$ESPI = \frac{Max_x - Min_x}{|Max_E - Min_E|}$$

Vsebina poglavja je narejena po temeljni literaturi Cline (2020), Klingenberg (2019), Legendre in Legendre (2012), Phenotypic plasticity (2019), Valladares in sod. (2006) in Zelditch in sod (2012).



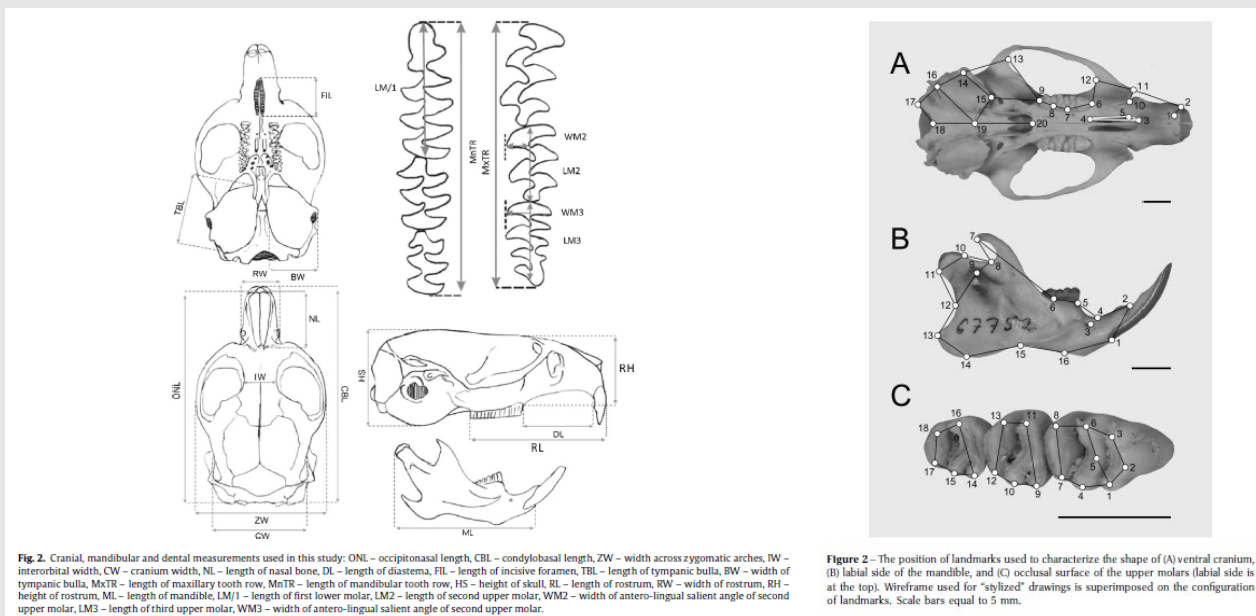
PROJEKT 11.1

Morfometrija malih sesalcev

Mali sesalci so priljubljena modelna skupina organizmov za študij fenotipske plastičnosti. Večina vrst ima široki areal razširjenosti. Praviloma so r ekološki strategji z velikimi populacijskimi nihanjem, kar občasno privede do izjemnih gostot v naravnem ali umetnem okolju. Imajo kratko življenjsko dobo in praviloma več razmnoževalnih ciklov na leto. Praviloma ne uživajo naravovarstvene zaščite. V naravoslovnih muzejih obstajajo bogate zbirke lobanj in mehov, ki so jih zbrali v okviru različnih kartiranj ali monitoringov.

Kombinacija navedenih dejstev omogoča edinstveni dostop do materiala ter analiziranje in vpogled v odziv fenotipa na dejavnike, ki so jim bile te živali izpostavljene. Morfometrična analiza lobanj malih sesalcev je dobro znanstveno utemeljen pristop za študij odnosov med morfometrijo in okoljskimi oz. ekološkimi spremenljivkami. Za prikaz morfometričnih lastnosti v glavnem uporabljajo dva metodološka pristopa: tradicionalna morfometrija in geometrijska morfometrija; glavna razlika med njima je v zajemanju meritev. Pri tradicionalni morfometriji so meritve in vhodni podatki dolžine oz. razdalje med anatomskimi strukturami. Z drugimi besedami, variabilnost lobanje opišemo z dolžinami dobro definiranih anatomskih lastnosti, npr. kondilobazalna dolžina lobanje. Meritve izvajamo s kljunastim merilom, kot so to predstavili Mahmoudi in sod. (2017) (Slika 11.4). V geometrijski morfometriji so meritve koordinate točk izbranih anatomskih struktur, posamezna meritev je zapisana z vrednostjo koordinat X in Y . Oslonilne točke določamo na fotografijah lobanj, kot so to predstavili Kryštufek in sod. (2017) (Slika 11.4).

Statistična obravnava je namenjena predvsem preverjanju razlik med skupinami (med vrstami ali med geografskimi vzorci), za kar se uporabljata univariatna t -test in F -test oz. multivariatna MANOVA. Analiza glavnih komponent je sploh najpogostejši pristop za redukcijo števila morfometričnih spremenljivk. Odnose med spremenljivkami preverjamo s korelacijsko analizo. Za razlaganje variabilnosti morfometričnih spremenljivke z ekološkimi dejavniki uporabljamo regresijsko analizo, z množico različnih pristopov.



Slika 11.4 Morfometrične spremenljivke meritev dolžin v pristopu s tradicionalno morfometrijo (levo) (Mahmoudi in sod., 2017) in v pristopu z geometrijsko morfometrijo (desno), meritve oslonilnih točk (Kryštufek in sod. 2017).





PRIMER 11.1

Fenotipska plastičnost velikosti gozdne voluharice – spremenljivke

Gozdna voluharica *Clethrionomys (Myodes) glareolus* naj bo modelni organizem za preverjanje reakcijske norme morfotipa (Slika 11.5). V prostorskem gradientu ekoloških dejavnikov bomo analizirali odziv telesne velikosti voluharic na klimatske razmere. Telesna velikost je odvisna spremenljivka, ki jo razlagamo s klimo kot neodvisno spremenljivko, gradienti okolja imajo prostorsko razsežnost, vzorci voluharic izvirajo iz 7 lokalitet v Sloveniji. Namen analize je preverjanje fenotipske plastičnosti, kot odgovor na različne klimatske razmere, v katerih živi posamezna populacija. Projekt je primer analize morfološke variabilnosti v prostoru kot odziv fenotipa na okoljske dejavnike, in s tem študij reakcijske norme.

Oblikujemo tri nize podatkov. Prvi niz tvorijo biometrične meritve lobanje in jih imenujemo morfološki znaki, drugi niz so prostorski podatki za posamezno lokaliteto, tretji niz podatkov so klimatske spremenljivke in jih imenujemo klimatski znaki.

Morfološki znaki obsegajo meritve 204 lobanj gozdne voluharice, podatkovni niz je sestavljen iz ene kategorične spremenljivke: lokaliteta oz. nominalna spremenljivka; in 8 morfometričnih znakov oz. intervalnih spremenljivk: dolžina lobanje CL, širina lobanje BB, višina lobanje VL, višina lobanje z bobničnim mehurjem VB, zigomatična širina ZB, višina rostruma RH, dolžina rostruma RL in dolžina diasteme DI (Preglednica 11.1 in Slika 11.4). Lobanje smo izmerili s kljunastim merilom z natančnostjo 0,1 mm. Material je del zbirke Prirodoslovnega muzeja Slovenije. Prostorska razporeditev lokalitet je podana z vrednostmi geografske dolžine, geografske širine in nadmorske višine (Preglednica 11.5). Klimatske spremenljivke so podane z 19 bioklimatskimi – BIO spremenljivkami (Preglednica 11.6).

Analiza morfoloških znakov – podatkovni niz je v obliki matrike velikosti 204 vrstic z osebki in 9 stolpcev z morfometričnimi spremenljivkami, od tega 8 intervalnih in ena nominalna, ki oblikuje 7 geografskih vzorcev. Telesna velikost organizma je kompleksna vrednost, v njen izračun smo vključili 8 linearnih meritev (Preglednica 11.1). Redukcijo razsežnosti, z ohranitvijo njihove informativne vrednosti, omogoča multivariatna statistična metoda analiza glavnih komponent PCA. Rezultat PCA so nove spremenljivke, imenovane glavne komponente PC, njihova osnovna lastnost je, da so urejene po deležu variabilnosti, ki jo razlaga. To pomeni, da prva glavna komponenta PC1 razlaga najvišji delež skupne variabilnosti, druga glavna komponenta PC2 naslednji najvišji delež variabilnosti, vendar ne tistega, ki ga je razložila že PC1 itd. Največja variabilnost živali je v njihovi telesni velikosti in ker PC1 razlaga najvišjo variabilnost, je splošno sprejeta in potrjena lastnost, da je PC1 komponenta telesne velikosti.

V obravnavanem primeru prvo glavno komponento PC1 interpretiramo kot telesno velikost gozdne voluharice. Preverimo ali vzorci izkazujejo fenotipsko variabilnost, sledi izračun povprečne vrednosti PC1 za posamezni geografski vzorec.

Preglednica 11.1 Začetek tabele z meritvami lobanjskih spremenljivk, celotna matrika ima dimenzije 204 x 9

| Nominalna spr. | Intervalne spremenljivke | | | | | | | |
|----------------|--------------------------|-----|------|------|-----|-----|------|-----|
| Lokaliteta | CBL | RoH | SL | ZG | VB | VL | RoL | DiL |
| Črni log | 23,4 | 6,2 | 10,8 | 13,1 | 9,2 | 7,3 | 13,3 | 6,5 |
| Črni log | 22,2 | 6,3 | 11,1 | 13,6 | 9,0 | 7,0 | 12,8 | 6,1 |
| Črni log | 23,9 | 6,5 | 11,5 | 13,5 | 9,3 | 7,2 | 14,1 | 7,1 |
| Črni log | 22,0 | 6,1 | 11,0 | 12,6 | 8,8 | 7,2 | 12,6 | 6,2 |
| Črni log | 22,5 | 5,9 | 10,9 | 13,0 | 9,0 | 6,7 | 12,8 | 6,3 |



Slika 11.5 Gozdna voluharica *Clethrionomys glareolus*.

Foto. Franc Janžekovič.



PRIMER 11.2

Fenotipska plastičnost velikosti gozdne voluharice – PCA

Z namenom redukcije števila morfometričnih spremenljivk smo 8 intervalnih spremenljivk transformirali v eno samo spremenljivko z analizo glavnih komponent PCA. Surove vrednosti meritev se med spremenljivkami močno razlikujejo npr. kondilobazalna dolžina lobanje CL ima vrednosti okrog 23 mm, višina rostruma RH pa okrog 6 mm. V multivariatni analizi variabilnost absolutno večje spremenljivke lahko zakrije variabilnost absolutno manjše spremenljivke. Zato surove vrednosti pred vključitvijo v PCA standardiziramo, in s tem zagotovimo enakovreden prispevek vsake spremenljivke. Rezultat PCA so glavne komponente, za analizo velikosti voluharic je informativna le prva glavna komponenta PC1, ki jo interpretiramo kot komponento velikosti in jo imenujemo morfološka PC1 – MPC1. Morfološka prva glavna komponenta MPC1 razloži 71,97 % celotne variabilnosti osmih intervalnih spremenljivk (Preglednica 11.2), z drugimi besedami, od celotne variabilnosti lobanj je 72 % vsebovane v komponenti velikosti.

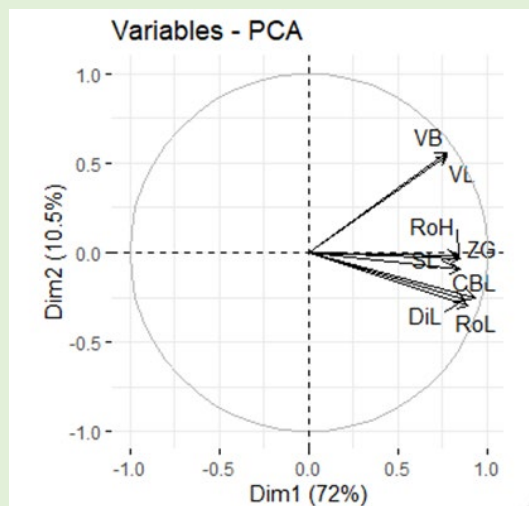
Vseh 8 lobanjskih spremenljivk je v pozitivni in močni korelaciji z MPC1 (Slika 11.6), kar pomeni, da z naraščanjem MPC1 naraščajo vrednosti vseh osnovnih spremenljivk, oz. da se z naraščanjem MPC1 večja velikost lobanje gozdnih voluharic, in obratno z manjšanjem vrednosti MPC1 se zmanjšuje velikost gozdnih voluharic.

Preglednica 11.2 Lastne vrednosti in deleži razložene varianc glavnih komponent

| PC | Lastna vrednost | Razložena varianca % |
|----|-----------------|----------------------|
| 1 | 5,76 | 71,97 |
| 2 | 0,84 | 10,52 |
| 3 | 0,48 | 5,96 |
| 4 | 0,35 | 4,37 |
| 5 | 0,24 | 3,01 |
| 6 | 0,17 | 2,10 |
| 7 | 0,11 | 1,35 |
| 8 | 0,06 | 0,72 |

```
R >
##Uvoz podatkov gozdna voluharica
#Izbor podatkov za vpeljavo v PCA
lob.active<-lob[1:204, 2:9]
#Pregled začetka in konca tabele
head(lob.active)
tail(lob.active)
# z standardizacija
z.lob<- scale(lob.active)
```

```
R > package = »factoextra«
#aktivacija paketa za izračun PCA
library(factoextra)
#PCA analiza
res.pca<-prcomp(z.lob)
#Lastne vrednosti
eig.val<-get_eigenvalue(res.pca)
eig.val
#Korelacija med PC in lob
```



Slika 11.6 Korelacija med osnovnimi spremenljivkami in prvima dvema glavnima komponentama. Smer vektorjev izkazuje povezavo med spremenljivkami in komponentami, dolžina vektorjev izkazuje jakost povezave. Krog predstavlja korelacijo $r = 1$.



**PRIMER 11.3****Fenotipska plastičnost velikosti gozdne voluharice – fenotipska variabilnost**

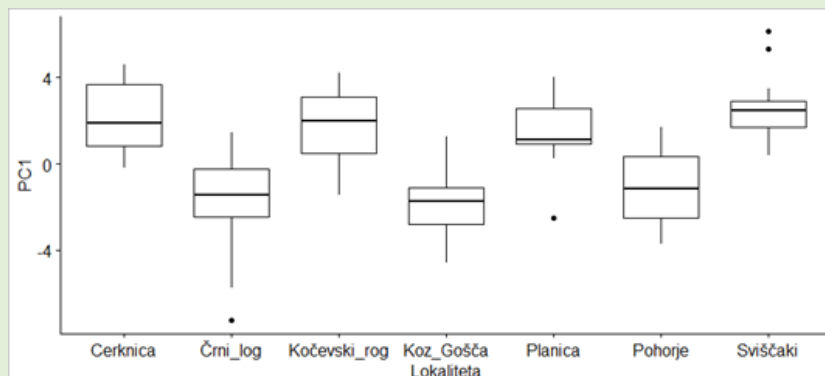
Variabilnost telesne velikosti gozdnih voluharic (na osnovi prve morfološke glavne komponente MPC1) med lokalitetami preverimo z ANOVA. V ničelni hipotezi H_0 se povprečne vrednosti MPC1 med nahajališči ne razlikuje; alternativna hipoteza H_a , napoveduje, da se v velikost voluharice med lokalitetami razlikujejo. Stopnjo tveganja postavimo na vrednost $\alpha = 0,05$. F -test vrne visoko značilno vrednost $F = 38,7$ $p < 0,0001$ (Preglednica 11.3). Zato sprejmemo H_a in trdimo, da se vrednosti MPC1 med skupinami razlikujejo. S tem smo potrdili fenotipsko variabilnost. S posteriornim testiranjem primerjamo posamezne lokalitete med seboj in preverjamo statistično značilnost razlik med vsemi pari (Preglednica 11.4). Lokalitete se delijo v dva velikostna razreda, v telesno manjšo in večjo skupino (Slika 11.7).

Preglednica 11.3 Rezultati ANOVA F -testa

| | df | Vsota kvadratov | Povprečje kvadratov | F | p |
|------------|-----|-----------------|---------------------|------|---------|
| Lokaliteta | 6 | 632,4 | 105,39 | 38,7 | <0,0001 |
| Ostanki | 197 | 536,4 | 2,72 | | |

Preglednica 11.4 Rezultati Tukey-evega HSD posteriornega testiranja med pari lokalitet. Značilne razlike so v krepkem tisku

| Lokalitete | diff | lwr | upr | p |
|-------------------------------|--------------|--------------|--------------|---------------|
| Črni_log-Cerknica | -3,78 | -5,22 | -2,34 | 0,0000 |
| Kočevski_rog-Cerknica | -0,44 | -1,88 | 1,00 | 0,9695 |
| Koz_Gošča-Cerknica | -4,02 | -5,54 | -2,50 | 0,0000 |
| Planica-Cerknica | -0,78 | -2,85 | 1,30 | 0,9221 |
| Pohorje-Cerknica | -3,27 | -4,86 | -1,67 | 0,0000 |
| Sviščaki-Cerknica | 0,41 | -1,36 | 2,18 | 0,9929 |
| Kočevski_Rog-Črni_log | 3,34 | 2,37 | 4,30 | 0,0000 |
| Koz_Gošča-Črni_log | -0,24 | -1,33 | 0,85 | 0,9946 |
| Planica-Črni_log | 3,00 | 1,22 | 4,78 | 0,0000 |
| Pohorje-Črni_log | 0,51 | -0,67 | 1,70 | 0,8570 |
| Sviščaki-Črni_log | 4,19 | 2,78 | 5,60 | 0,0000 |
| Koz_Gošča-Kočevski_Rog | -3,58 | -4,66 | -2,50 | 0,0000 |
| Planica-Kočevski_Rog | -0,33 | -2,11 | 1,44 | 0,9977 |
| Pohorje-Kočevski_Rog | -2,82 | -4,00 | -1,65 | 0,0000 |
| Sviščaki-Kočevski_Rog | 0,85 | -0,55 | 2,26 | 0,5405 |
| Planica-Koz_Gošča | 3,24 | 1,40 | 5,08 | 0,0000 |
| Pohorje-Koz_Gošča | 0,75 | -0,53 | 2,03 | 0,5834 |
| Sviščaki-Koz_Gošča | 4,43 | 2,94 | 5,92 | 0,0000 |
| Pohorje-Planica | -2,49 | -4,39 | -0,59 | 0,0025 |
| Sviščaki-Planica | 1,19 | -0,86 | 3,24 | 0,5977 |
| Sviščaki-Pohorje | 3,68 | 2,12 | 5,24 | 0,0000 |



Slika 11.7 Okvirji z ročaji za MPC1 – telesna velikost gozdne voluharice, za 7 lokalitet. S pikami so označeni osamelci.

```
R >
#ANOVA
res.ftest<-aov(PC1 ~ Lokaliteta, data=loc_prva)
summary(res.ftest)
TukeyHSD(res.ftest)
```

```
R > package = »ggpubr«
#Boxplot
library(ggpubr)
ggboxplot(loc_prva, x="Lokaliteta", y="PC1")
```



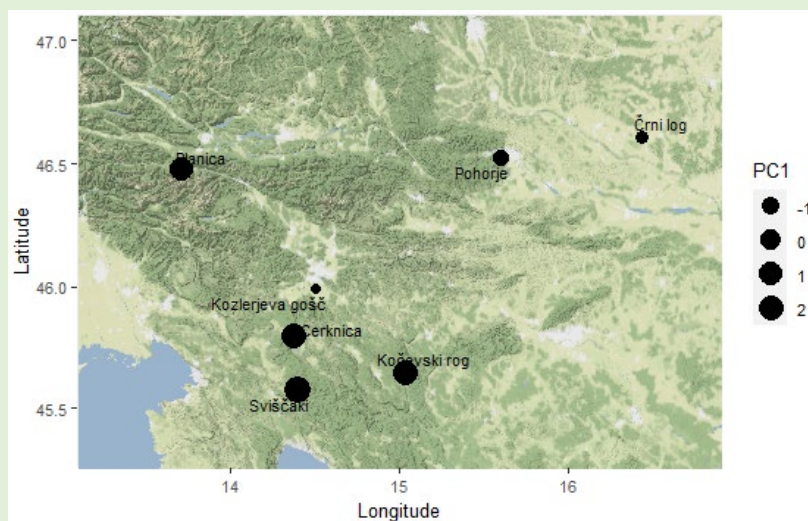
PRIMER 11.4

Fenotipska plastičnost velikosti gozdne voluharice – prostorska komponenta

Posamezni lokaliteti izračunamo aritmetično sredino prve glavne komponente MPC1 (Preglednica 11.5). Glede na telesno velikost razdelimo vzorce v dve skupini. V prvi so telesno manjše živali, $MPC1 < 0$, v kateri so vzorci iz severovzhodne in osrednje Slovenije: Črni log, Pohorje in Kozlarjeva gošča. V skupini telesno večjih živali $MPC1 > 0$ so uvrščeni vzorci iz jugo- in severozahodne Slovenije: Sviščaki, Cerknica, Kočevski Rog in Planica. Variabilnost v velikosti sovpada s prostorsko razporeditvijo vzorcev. Prostorska razporeditev je podana s koordinatami in navedbo nadmorske višine (Preglednica 11.5) in prikazana na zemljevidu (Slika 11.8), na kateri je telesna velikost vzorcev prikazana z velikostjo kroga. Zemljevid je izdelan v okolju R (R Core Team 2022) s paketi: »maps« (Becker in Wilks 2018), »ggplot2« (Wickham 2016), »ggmap« (Kahle in Wickham 2013), »rlang« (Henry in Wickham 2020), »ggrepel« (Slowikowski 2020), »dplyr« (Wickham in sod. 2020).

Preglednica 11.5 Geografske koordinate: Y – geografska širina, X – geografska dolžina, A – nadmorska višina in povprečna vrednost MPC1 za 7 lokalitet iz katerih izvirajo vzorci gozdnih voluharic

| Lokaliteta | Y Latitude | X Longitude | A Altitude (m) | MPC1 |
|------------------|------------|-------------|----------------|--------|
| Črni log | 46,60832 | 16,42672 | 160 | -1,631 |
| Pohorje | 46,52623 | 15,59868 | 300 | -1,119 |
| Kozlerjeva Gošča | 45,99366 | 14,50385 | 290 | -1,871 |
| Planica | 46,47579 | 13,71620 | 1000 | 1,371 |
| Kočevski Rog | 45,64882 | 15,03029 | 930 | 1,705 |
| Cerknica | 45,79666 | 14,37225 | 550 | 2,148 |
| Sviščaki | 45,57501 | 14,40023 | 1240 | 2,559 |



Slika 11.8 Razporeditev lokalitet z vzorci gozdne voluharice, telesna velikost je predstavljena z velikostjo krogca. Veliki krogi velike živali, manjši krogi male živali.

```
R >
#ANOVA
#Povprečje PC1 po lokalitetah
pc.mean<-tapply(loc_prva$PC1, loc_prva$Lokaliteta, mean)
pc.mean
```

```
R > package = »ggplot2«, »maps«, »ggmap«, »rlang«, »ggrepel«, »dplyr«
#Izris zemljevida
glgmap <- get_stamenmap(bbox = c(left = 13.10, bottom = 45.25, right = 16.90, top = 47.10),
  maptype=c("terrain-background"), zoom = 10, crop=TRUE)
p1 <- ggmap(glgmap)
p2 <- p1 + theme(text = element_text(size=10))
p3 <- p2 + geom_point(data = koord, aes(y=Y, x=X, size=PC1)) +
  scale_size_continuous(range=c(2,6)) +
  geom_text_repel(data = koord, aes(x=X, y = Y, label = Lokaliteta), size=3)
p4 <- p3 + xlab("Longitude") + ylab("Latitude")
p4
```





PRIMER 11.5

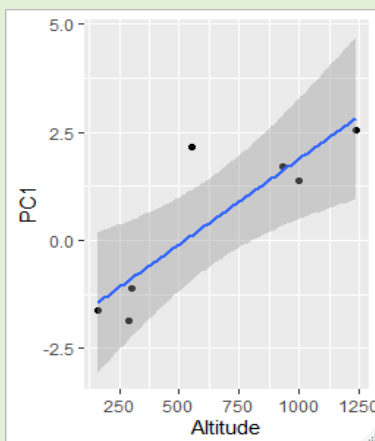
Fenotipska plastičnost velikosti gozdne voluharice – nadmorska višina

Za gozdno voluharic smo dokazali fenotipsko variabilnost telesne velikosti v prostoru, zato lahko govorimo o geografski variabilnosti. V nadaljevanju želimo poiskati okoljske spremenljivke, s katerimi lahko razložimo čim večji delež fenotipske plastičnosti.

Najprej preverimo odnos med telesno velikostjo, MPC1, in prostorskimi koordinatami X, Y s pomočjo korelacijske analize. Korelacijski koeficienti so v vseh treh primerih močni $|r| > 0,6$ (Preglednica 11.6). To pomeni, da velikost izkazuje prostorski vzorec. Velikost živali upada v smeri od juga proti severu in v smeri od zahoda proti vzhodu. Z višanjem nadmorske višine velikost živali narašča. Na nižjih nadmorskih višinah so voluharice manjše; odnos je predstavljen z razsevnim diagramom in regresijsko premico (Slika 11.9).

Preglednica 11.6 Korelacijski koeficienti med telesno velikostjo MPC1 gozdnih voluharic in prostorskimi koordinatami.

| Prostorska koordinata | MPC1 |
|------------------------|-------|
| Y – geografska širina | -0,66 |
| X – geografska dolžina | -0,60 |
| A – nadmorska višina | 0,87 |



```
R >
#Korelacija, data=koord
cor(koord$PC1, koord$Altitude)
cor(koord$PC1, koord$Y)
cor(koord$PC1, koord$X)
```

```
R >
#Biplot, data=koord
biplot<-ggplot(koord, aes(y=koord$PC1,
x=koord$Altitude))+geom_point()+geom_smooth(method='lm')
bp<-biplot + xlab("Altitude") + ylab("PC1")
bp
```

Slika 11.9 Z regresijsko premico je prikazan odnos med nadmorsko višino (Altitude – m) in telesno velikostjo MPC1 (PC1) gozde voluharice.





PRIMER 11.6

Fenotipska plastičnost velikosti gozdne voluharice – klimatski podatki

Klima je pomemben okoljski in ekološki dejavnik, ki na organizme vpliva neposredno, npr. homeotermne živali vlagajo veliko energije v vzdrževanje stalne telesne temperature in večja, kot je razlika med temperaturo telesa in temperaturo okolice, večjim izgubam toplote so izpostavljeni. Vpliv klime na bioto je tudi posreden s primarno produkcijo in s tem na kompleksnost prehranjevalnega spleta. V ekološke analize vključujemo klimatske indekse oz. bioklimatske spremenljivke (ang. Bioclimatic Predictors).

Klimatski podatki so prosto dostopni in pokrivajo celotno zemeljsko oblo, dosegljivi so s pomočjo različnih orodij, v tej analizi klimatske podatke pridobimo s paketom »raster« iz programskega okolja R (Hijmans 2020). Poznamo 19 dogovorjenih bioklimatskih spremenljivk BIO1-19 (O'Donnell in Ignizio, 2012) (Preglednica 11.6).

Preglednica 11.6 Vrednosti 19 bioklimatskih spremenljivk na 7 lokalitetah. Razlaga spremenljivk je v tabeli 11.7

| Lokalitete | BIO1 | BIO2 | BIO3 | BIO4 | BIO5 | BIO6 | BIO7 | BIO8 | BIO9 | BIO10 | BIO11 | BIO12 | BIO13 | BIO14 | BIO15 | BIO16 | BIO17 | BIO18 | BIO19 |
|------------|------|------|------|------|------|------|------|------|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 | 100 | 98 | 31 | 7399 | 261 | -46 | 307 | 192 | 17 | 192 | 2 | 808 | 98 | 38 | 30 | 282 | 122 | 282 | 130 |
| 2 | 85 | 98 | 31 | 7330 | 246 | -62 | 308 | 177 | 1 | 177 | -12 | 1009 | 122 | 47 | 30 | 354 | 149 | 354 | 161 |
| 3 | 96 | 88 | 30 | 6969 | 252 | -34 | 286 | 102 | 16 | 185 | 6 | 1322 | 140 | 81 | 17 | 403 | 252 | 344 | 287 |
| 4 | 44 | 86 | 31 | 6436 | 186 | -83 | 269 | 125 | -30 | 125 | -39 | 1305 | 139 | 65 | 22 | 411 | 223 | 411 | 229 |
| 5 | 90 | 85 | 30 | 6852 | 242 | -37 | 279 | 51 | 10 | 177 | 1 | 1309 | 143 | 86 | 17 | 407 | 262 | 320 | 300 |
| 6 | 88 | 80 | 29 | 6647 | 235 | -34 | 269 | 95 | 11 | 173 | 3 | 1349 | 142 | 87 | 16 | 413 | 264 | 341 | 294 |
| 7 | 67 | 74 | 29 | 6417 | 208 | -47 | 255 | 76 | -7 | 150 | -14 | 1347 | 142 | 85 | 16 | 407 | 264 | 343 | 290 |

R > package = »raster«

```
#BIOKLIM, priklic klimatskih podatkov
```

```
r <- getData("worldclim", var="bio", res=10)
```

```
r <- r[[c(1:19)]] # Definiranje klimatskih spremenljivk
```

```
names(r) <- c("bio1", "bio2", "bio3", "bio4", "bio5", "bio6", "bio7", "bio8", "bio9", "bio10", "bio11", "bio12", "bio13", "bio14", "bio15", "bio16", "bio17", "bio18", "bio19")
```

```
# Definiranje lokalitet
```

```
lats <- c(46.60832, 46.52623, 45.99366, 46.47579, 45.64882, 45.79666, 45.57501)
```

```
lons <- c(16.42672, 15.59868, 14.50385, 13.7162, 15.03029, 14.37225, 14.40023)
```

```
coords <- data.frame(x=lons,y=lats)
```

```
points <- SpatialPoints(coords, proj4string = r@crs)
```

```
values <- extract(r,points)
```

```
df <- cbind.data.frame(coordinates(points),values)
```

```
df # Prikaz podatkov
```

```
write.table(df, file = "df.txt", sep="\t") # Izvoz podatkov
```





PRIMER 11.7

Fenotipska plastičnost velikosti gozdne voluharice – klima

Povezava med 19 bioklimatskimi spremenljivkami in velikostjo živali (MPC1) je predstavljena s korelacijskimi koeficienti (Preglednica 11.7). Na splošno je velikost gozdnih voluharic v negativnem odnosu s temperaturami in v pozitivnem odnosu s padavinami. Lahko posplošimo, da so živali v klimatih z nižjimi temperaturami in več padavinami telesno večje, in obratno, telesno manjše živali živijo v okolju z višjimi temperaturami in manj padavinami. Za oblikovanje preglednejšega dvodimenzionalnega klimatskega modela vključimo 19 bioklimatskih spremenljivk v analizo glavnih komponent PCA, z namenom redukcije razsežnosti klimatskih spremenljivk in ekstrakcije glavnih komponent. Analizo smo izvedli v R okolju (R Core Team 2022) in s pomočjo paketa »factoextra« (Kassambara in Mundt 2020).

Preglednica 11.7 Pregled 19 bioklimatskih spremenljivk – BIO, in korelacijski koeficient med velikostjo gozdnih voluharic in bioklimatsko spremenljivko, močne korelacije $|r| > 0,5$ so v krepkem tisku

| Koda | Opis | MPC1 |
|-------|---|---------------|
| BIO1 | Povprečna letna temperatura (*100) | -0,539 |
| BIO2 | Povprečni dnevni razpon (mesečno povprečje (max temp. – min temp.)) | -0,843 |
| BIO3 | Izotermalnost (BIO2/BIO7) (*100) | -0,633 |
| BIO4 | Sezonskost temperature (standardna deviacija * 100) | -0,837 |
| BIO5 | Maksimalna temperatura vročega meseca | -0,670 |
| BIO6 | Minimalna temperatura hladnega meseca | -0,035 |
| BIO7 | Letni razpon temperature (BIO5-BIO6) | -0,859 |
| BIO8 | Povprečna temperatura mokrega četrletja | -0,699 |
| BIO9 | Povprečna temperatura suhega četrletja | -0,446 |
| BIO10 | Povprečna temperatura vročega četrletja | -0,607 |
| BIO11 | Povprečna temperatura hladnega četrletja | -0,320 |
| BIO12 | Letne padavine | 0,668 |
| BIO13 | Padavine mokrega meseca | 0,656 |
| BIO14 | Padavine suhega meseca | 0,658 |
| BIO15 | Sezonskost padavin (Koefficient variabilnosti) | -0,637 |
| BIO16 | Padavine mokrega četrletja | 0,655 |
| BIO17 | Padavine suhega četrletja | 0,679 |
| BIO18 | Padavine toplega četrletja | 0,310 |
| BIO19 | Padavine hladnega četrletja | 0,627 |

```
R > package = »factoextra«
#PCA, data=df
df.active<-df[1:7, 3:21] #Izbor spremenljivk
head(df.active)
res.pca<-prcomp(df.active, scale. = TRUE) #Izračun PCA
eig.val<-get_eigenvalue(res.pca) #Eigenvalue
eig.val
gk<-as.data.frame(res.pca$x) #izvoz PCjev
```

```
R >
#Korelacije
cor(koord$PC1, df.active)
```





PRIMER 11.8

Fenotipska plastičnost velikosti gozdne voluharice – PCA klime

Prvi dve klimatski komponenti KPC1 in KPC2, ki ju dobimo po vključitvi 19 klimatskih spremenljivk v PCA, skupaj razložita 94 % variabilnosti klime (Preglednica 11.8). K prvi klimatski glavni komponenti KPC1 največ prispevajo spremenljivke BIO2, 4, 7, 8, 12, 13, 14, 15, 16, 17 in 19. BIO2, 4 in 7 so v negativnem odnosu s KPC1 (Preglednica 11.9 in Slika 11.9) in razlagajo predvsem sezonsko in dnevno variabilnost v temperaturi in nižjo povprečno temperaturo. BIO12, 13, 16, 17, ki so prav tako v pozitivnem razmerju s KPC1, razlagajo padavine. Iz navedenega lahko izpeljemo, da z naraščanjem KPC1 naraščajo padavine, manjša pa se variabilnost v temperaturi. Nižja je tudi povprečna temperatura. Nizke vrednosti KPC1 opisujejo klimo z malo padavinami, večjimi dnevnimi in sezonskimi amplitudami v temperaturi in višjimi povprečnimi temperaturami (Preglednica 11.9 in Slika 11.10).

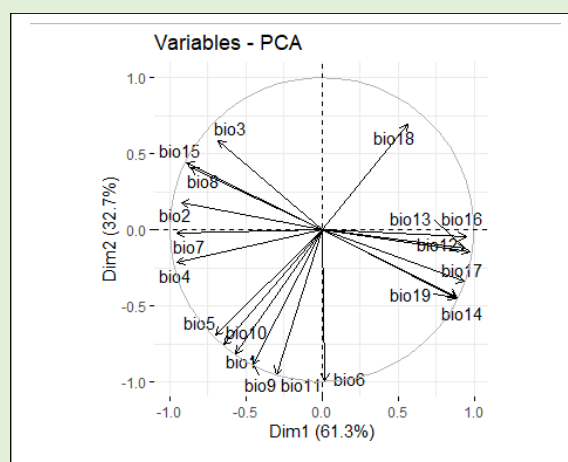
Preglednica 11.8 Lastne vrednosti in delež razložene variance

| PC | Lastna vrednost | Varianca % | Kumulativna varianca % |
|-----|-----------------|------------|------------------------|
| PC1 | 11,64 | 61,25 | 61,25 |
| PC2 | 6,22 | 32,75 | 94,00 |
| PC3 | 0,83 | 4,35 | 98,35 |
| PC4 | 0,23 | 1,21 | 99,55 |
| PC5 | 0,08 | 0,41 | 99,96 |
| PC6 | 0,01 | 0,04 | 100,00 |
| PC7 | 0,00 | 0,00 | 100,00 |

Preglednica 11.9 Korelacijski koeficienti med bioklimatskimi spremenljivkami BIO1–19 in klimatskima glavnima komponentama KPC1 in KPC2. Močne korelacije $|r| > 0,5$ so v krepkem tisku

| BIO | KPC1 | KPC2 |
|-------|---------------|---------------|
| BIO1 | -0,568 | -0,817 |
| BIO2 | -0,921 | 0,181 |
| BIO3 | -0,679 | 0,588 |
| BIO4 | -0,949 | -0,214 |
| BIO5 | -0,701 | -0,692 |
| BIO6 | 0,021 | -0,995 |
| BIO7 | -0,951 | -0,026 |
| BIO8 | -0,859 | 0,405 |
| BIO9 | -0,455 | -0,887 |
| BIO10 | -0,643 | -0,756 |

| BIO | KPC1 | KPC2 |
|-------|---------------|---------------|
| BIO11 | -0,301 | -0,951 |
| BIO12 | 0,968 | -0,143 |
| BIO13 | 0,941 | -0,125 |
| BIO14 | 0,886 | -0,456 |
| BIO15 | -0,888 | 0,441 |
| BIO16 | 0,945 | -0,046 |
| BIO17 | 0,933 | -0,341 |
| BIO18 | 0,562 | 0,694 |
| BIO19 | 0,886 | -0,444 |



Slika 11.10 Korelacijski grafikon med BIO klimatskimi spremenljivkami in prvima dvema klimatskima glavnima komponentama (Dim1 in Dim2).

```
R > package = »factoextra«
# korelacija med PCji in BIO
res.var <- get_pca_var(res.pca)
res.var$coord
# graf korelacij med PC in BIO
fviz_pca_var(res.pca, repel = TRUE)
```





PRIMER 11.9

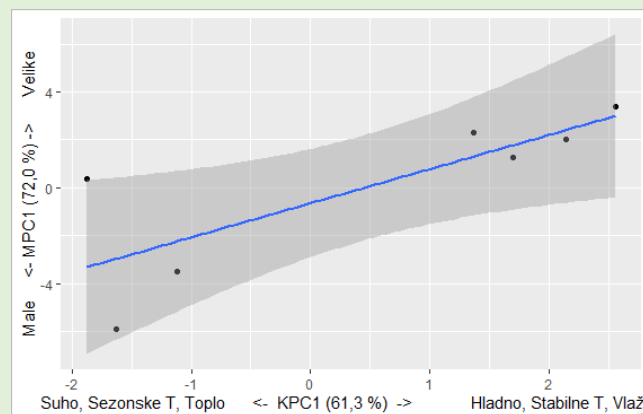
Fenotipska plastičnost velikosti gozdne voluharice – MPC1 in KPC1

Po dosednji analizi je morfološka variabilnost gozdne voluharice zajeta v MPC1 – komponenta velikost. Variabilnost klimatskih razmer na lokalitetah, s katerih živali izvirajo, je zajeta v KPC1 in KPC2 – komponenta klime.

Povezavo med MPC1 ter KPC1 in KPC2 preverimo s korelacijsko analizo, ki vrne močno pozitivno povezavo samo med prvima dvema glavnima komponentama in korelacija s KPC2 ni pomembna (Preglednica 11.10). Ker naraščanje KPC1 opisuje klimo z več padavin in manjšo variabilnost v temperaturi, tako na dnevnem kot letnem nivoju in tudi nižje povprečne temperature, zaključimo, da telesno velike živali živijo v klimatskih razmerah z vlažno in hladnejšo klimo z manjšim temperaturnim razponom. Telesno manjše živali pa živijo v klimi z manj padavin in v povprečju toplejšem podnebjju ter z večjimi sezonskimi razlikami v temperaturi (Slika 11.11).

V kontekstu fenotipske plastičnosti vlažna in hladnejša klima z manjšim temperaturnim razponom, kar velja za višje nadmorske višine (nad 600 m), favorizira telesno večji fenotip. V okolju z manj padavin, višjimi temperaturami in večjim temperaturnim razponom, predvsem v nižjih nadmorskih višinah (pod 500 m), je favoriziran telesno manjši fenotip. Takšni zaključki se v glavnem skladajo z Bergmannovim pravilom, po katerem je v hladnejših razmerah favoriziran telesno večji fenotip.

Možna razlaga za telesno večje živali v bolj vlažni in hladnejši klimi je v energetski bilanci. Z vidika izgube toplote preko površine je pomembno razmerje med volumnom in površino živali, toplota nastaja v telesu (v volumnu) oddaja se preko površine. Zato je v hladnejši klimi favoriziran telesno večji fenotip, saj ima ugodnejše razmerje med volumnom in površino telesa. Lahko pa tudi nutrienti v enem tipu klime omogočajo živalim, da zrastejo večje.



Slika 11. Regresijska premica med klimatsko KPC1in morfološko MPC1glavno komponento.

Preglednica 11.11 Korelacijska koeficienta med velikostjo živali – MPC1 in klimo KPC1 in KP2

| PC | KPC1 | KPC2 |
|------|-------|-------|
| MPC1 | 0,797 | 0,065 |

```
R >
#Korelacija med MPC1 in KPC
cor(koord$PC1, gk)
```

```
R > package = »ggplot2«
#Biplot, MPC1 KPC1
biplot<-ggplot(koord.klima, aes(y=koord.klima$PC11,
x=koord.klima$PC1))+geom_point()+geom_smooth(method='lm')
bp<-biplot + xlab("Suho, Sezonske T, Toplo   <- KPC1 (61,3 %) ->   Hladno, Stabilne T, Vlažno") +
  ylab("Male   <- MPC1 (72,0 %) ->   Velike")
bp
```

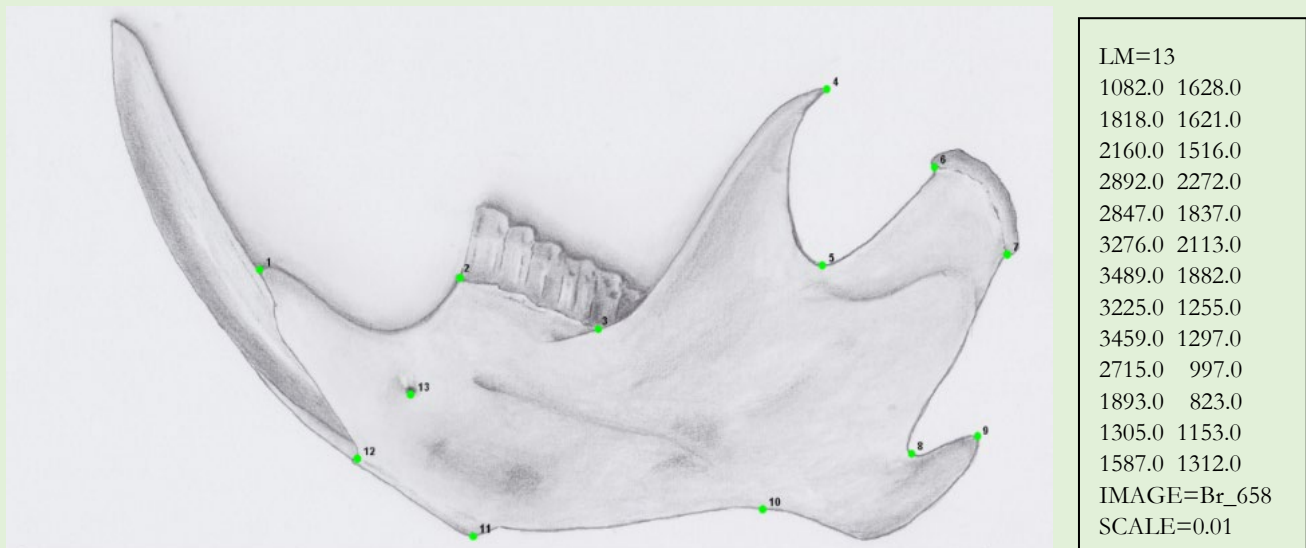
**PRIMER 11.10**

Morfometrična variabilnost mandibule velikega voluharja GM1 – ideja

Veliki voluhar *Arvicola terrestris* je vrsta z velikim arealom razširjenosti. Živi v zmernem pasu od Atlantske obale do sredine Azije. Vrsta živi v različnih odprtih travniških do močvirnih okoljih in predvsem obdelanih habitatih in izkazuje dobro prilagodljivost na široki spekter ekoloških gradientov. Dobra prilagodljivost pomeni razvoj različnih morfotipov, vključno s klinami. S tega vidika je zelo privlačen modelni organizem za študij morfometrije. V tem poglavju je predstavljen metodološki pristop in interpretacija rezultatov v variabilnosti velikosti in obliki mandibule velikega voluharja. V analizi obravnavamo dva vzorca, izvirata iz populacij, ki živijo v dveh različnih habitatih. En vzorec voluharjevih mandibul pripada populaciji iz kopenskega habitata (terestrični morfotip) in drugi iz močvirnega habitata (akvatični morfotip). Morfometrična variabilnost je ovrednotena z merjenjem oslonilnih točk. Analiza metodološko sodi na področje geometrijske morfometrije. Z metodami geometrijske morfometrije informacijo o morfometrični variabilnosti mandibule razstavimo na centroidno velikost CS (ang. Centroid Size) in obliko (ang. Shape), ki jo povzemajo Prokrustove koordinate. Oba vidika variabilnosti vzpostavimo v relacijo s tipom habitata (kopenski, močvirni). Analiza morfotipa je narejena s programskim orodjem MorphoJ (Klingenberg 2011).

Projekt je primer analiziranja morfometrične variabilnosti, kot odziv fenotipa na okoljske dejavnike, in s tem študij reakcijske norme. Fenotipsko plastičnost mandibule velikega voluharja interpretiramo kot posledico selekcijskih pritiskov okolja, kopensko proti vodno okolje.

Izpis meritev 13 oslonilnih točk LM – koordinate X in Y za mandibulo z inventarno številko Br_658 (IMAGE). Umeritveni koeficient (SCALE) ima vrednost 0,01 in poda razmerje med dejanskim merilom v mm in povečano fotografijo mandibule.



Slika 11.12 Mandibula velikega voluharja z označenimi oslonilnimi točkami: 1. Zgornji rob zobne jamice sekalca, 2. Sprednji rob zobne jamice prvega meljaka, 3. Presečišče zobnega grebena s sprednjim robom kavljastega izrastka, 4. Vrh kavljastega izrastka, 5. Najnižja točka na krivini med kavljastim in sklepnim izrastkom, 6. Sprednji rob sklepnega izrastka, 7. Zadnji rob sklepnega izrastka, 8. Najnižja točka na krivini med sklepnim in kotnim izrastkom, 9. Vrh kotnega izrastka, 10. Najvišja točka na spodnjem robu telesa mandibule, 11. Najnižja točka mandibularne zrasti, 12. Spodnji rob zobne jamice sekalca, 13. Odprtina v mandibuli.

Avtorica risbe Polona Bitenc.



PRIMER 11.11

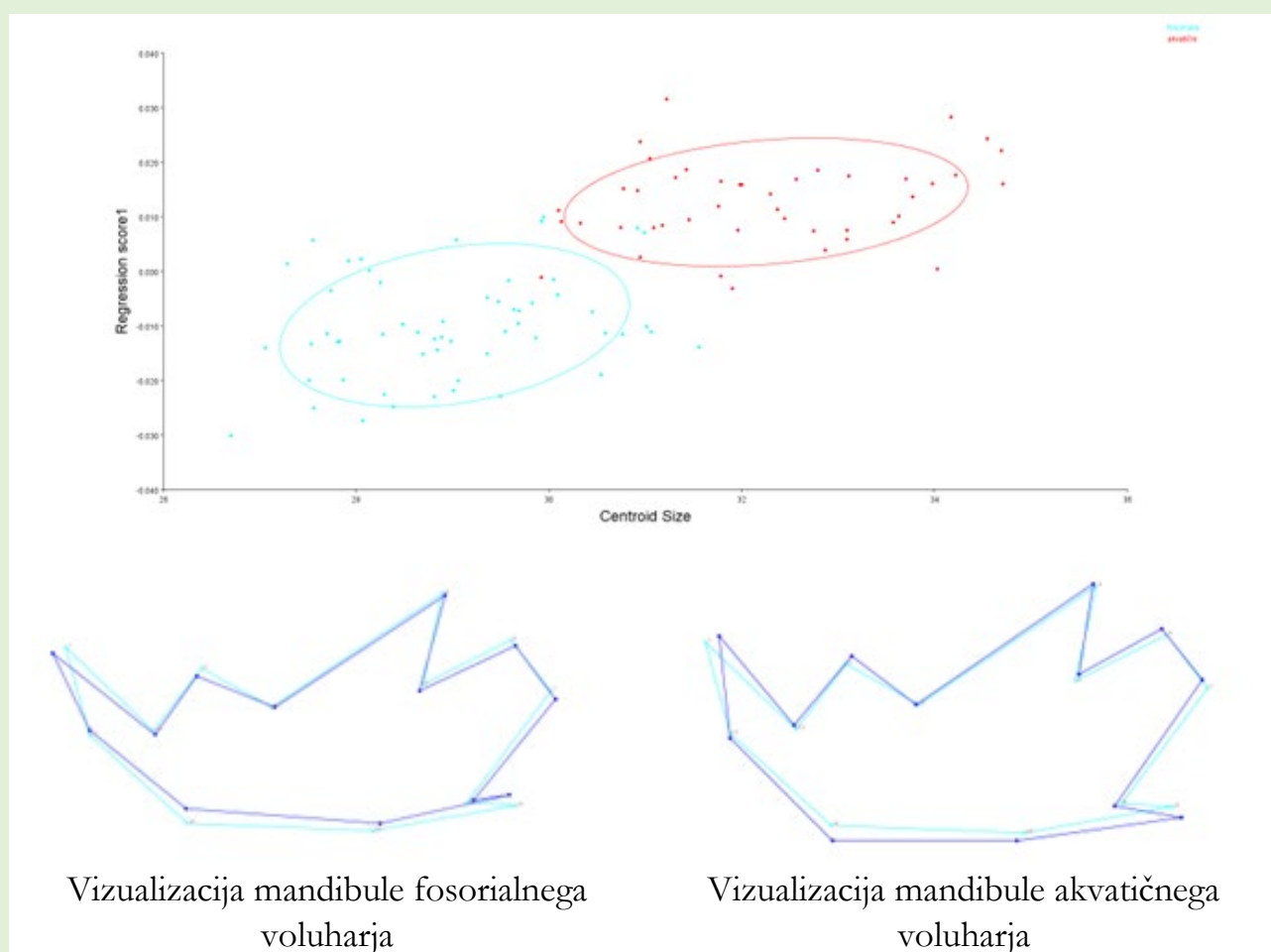
Morfometrična variabilnost mandibule velikega voluharja GM – PA

Meritve oslonilnih točk (Slika 11.12) vsebujejo informacijo o orientaciji in povečavi mandibule. V Prokrustovi analizi PA (ang. Procrustes Analysis) izvedemo poravnavo točk v isto izhodišče (premik in vrtenje) in skaliranje. Rezultat Prokrustove analize PA je: 1. grafična predstavitev centroidov oslonilnih točk in vseh meritev (Slika 11.13), na podlagi obsega oblaka točk okrog centroida sklepamo na variabilnost strukture; 2. informacija o velikosti mandibule; to je centroidna velikost CS in informacija o obliki mandibule, to so

Prokrustove koordinate (ang. Procrustes Coordinates), število koordinat se izračuna kot $nPK = 2n - 3$.



Slika 11.13 Poravnane meritve mandibule velikega voluharja.



Vizualizacija mandibule fosorialnega voluharja

Vizualizacija mandibule akvatičnega voluharja

Slika 11.14 Projekcija oblike mandibule (Regression score) na centroidno velikost (Centroid Size) zgoraj. Krivljenje mandibule fosorialnega in akvatičnega tipa voluharjev (spodaj). Svetlo modro povprečje, temno modro je lastnost tipa.

Odnos med velikostjo mandibule in njeno obliko ovrednotimo z regresijsko analizo, v kateri je centroidna velikost razlagalna spremenljivka, Prokrustove koordinate so odvisne spremenljivke. Permutacijski test je visoko značilen ($p < 0,0001$), centroidna velikost razloži 5,36 % variabilnosti v obliki. Rezultat regresijske analize je projekcija osebkov na vektor velikosti in vektor oblike (Slika 11.14) in vizualizacija oblikovnih sprememb v mandibuli (Slika 11.13). To pomeni, da lahko predstavimo dva neodvisna vidika variabilnosti, variabilnost v velikosti in variabilnost v obliki mandibule. vzdolž X osi so razporejeni osebki po velikosti mandibule, pri nižjih vrednosti so manjše živali – fosorialni tip; pri višjih vrednosti so večje živali – akvatični tip. vzdolž Y osi so osebki razporejeni glede na obliko mandibule. Pri nižjih vrednosti so fosorialne živali katerih mandibula ima ožje telo in bolj položno usmerjen sekalec; pri višjih vrednosti so akvatične živali, njihovo telo mandibule je čokato in širše in sekalec je bolj pravokotno postavljen na telo mandibule.



Evolucijska in ekogeografska pravila

Odzivi organizmov na okoljske oz. ekološke gradiente so praviloma podobni. Neodvisno so jih potrdili različni raziskovalci. Zato so se izoblikovala pravila, ki jih s skupnim imenom imenujemo **biološka pravila** (sin. biološki zakoni, ekogeografska pravila, makroekološka pravila). Dejansko gre za skrajšane posplošitve in poenostavljene razlage zapletenih pojavov, do katerih pridemo pri analiziranju ekologije, biogeografije in morfološke plastičnosti organizmov. V nadaljevanju so navedena pogosteje citirana pravila. Večinoma so poimenovana po raziskovalcih, ki so jih strokovno utemeljili.

Bioklimatska pravila – adaptacije na energetske bilanco

Allenovo pravilo – oblika telesa in telesna razmerja pri endotermnih živalih so oblikovana na način, da imajo živali iz hladnejših klimatov čim manj izpostavljenih površin, s čimer minimizirajo izgubo toplote prek telesne površine. Po drugi strani imajo živali iz tropskih klimatov izpostavljenih površin telesa za boljšo oddajanje toplote. Primer: puščavska lisica ima zelo dolge in izpostavljene uhlje, polarna lisica pa zelo kratke.

Bergmannovo pravilo – homeotermne živali z obsežno razširjenostjo imajo večjo telesno velikost v hladnejših klimatih in so manjše v toplejših klimatih. Primer: gozdne voluharice, ki živijo na višjih nadmorskih višinah pri nižjih povprečnih temperaturah zraka so večje od onih, ki živijo na nižjih nadmorskih višinah in višji povprečni temperaturi zraka.

Glogerjevo pravilo – endotermne živali, ki živijo v bolj vlažnem okolju, npr. bližje ekvatorju, so bolj intenzivno pigmentirane, kot tiste, ki živijo v manj vlažnem ali sušnem okolju.

Jordanovo pravilo – vrste, ki živijo v hladnejši vodi, imajo manjše število merističnih znakov kot vrste iz toplih voda, npr. manjše število plavutnic, vretenc ali lusk.

Damuthovo pravilo – tudi pravilo energetske ekvivalence; predstavniki iste filogenetske linije, ki živijo v vodi so vedno večji, kot njihovi kopenski sorodniki. Izhodišče je energetsko, saj je v vodnem okolju prevajanje toplote hitrejše in selekcijski pritisk deluje v smeri ugodnejšega razmerja med površino in volumnom. Primer: veliki voluharji, ki živijo v močvirnem okolju so večji od onih, ki žive na kopnem.

Evolucijska pravila – evolucijski vzorci

Copovo pravilo – pri živalskih vrstah se z dolžino evolucijske zgodovine povečuje njihova telesna velikost.

Globokomorski gigantizem – globokomorske živali razvijejo večjo telesno velikost kot njihovi evolucijski sorodniki iz priobalnih morij.

Dollov zakon ireverzibilnosti – v procesu evolucijskega razvoja se organizmi ne vrnejo v izhodiščno stanje, torej v obliko v kateri so že bili.

Eichlerjevo pravilo – vrstno bogastvo parazitov je povezano z vrstnim bogastvom gostiteljev.

Fosterjevo pravilo – tudi otoško pravilo ali otoški učinek pravi, da se na otokih z omejenimi viri večje živali (večje od ca. 1 kg) pomanjšajo, male živali (manjše od ca. 250 g) pa se povečajo.

Gauseov zakon – tudi princip kompetitivnega izključevanja pravi, da dve vrsti, ki sta v kompeticiji za iste vire, ne moreta sobivati s stabilnima populacijama. Kompeticija pripelje do izumrtja šibkejšega konkurenta, ali do njegovega evolucijskega ali vedenjskega premika v drugačno ekološko nišo.

Harrisonovo pravilo – telesna velikost parazitov je povezana s telesno velikostjo gostitelja.

Hennigovo pravilo – vrste z najbolj primitivnimi (izvornimi) lastnostmi so razširjene na prvobitnem območju evolucijskega izvora vrste, kjer je bil center izvora skupine.

Lackov princip – velikost legla pri pticah je omejena z največjim številom mladičev, za katere lahko starši zagotovijo dovolj hrane.

Rapoportovo pravilo – razpon geografske širine areala vrst, ki živijo pri nižji geografski širini (bližje ekvatorju), je manjši; vrste, ki živijo v višjih geografskih širinah (v zmernem pasu in proti poloma) imajo večji razpon geografske širine areala.

Renschevo pravilo – pri živalskih filogenetskih linijah, kjer so samci večji spol, spolni dimorfizem narašča z večanjem telesa; razlika v spolnem dimorfizmu se zmanjšuje, če so samice večji spol.

Schmaulhausnov zakon – na populacijo na skrajnem robu zmožnosti ima minimalna spremembe kritičnega resursa bistveno večji vpliv kot variabilnost kateregakoli drugega resursa.

Van Valenov zakon – verjetnost izumrtja vrst in višjih taksonov (npr. družin in redov) je v času konstantna za vse skupine. Taksoni niso zaradi evolucijsko starost niti bolj ogroženi niti odpornejši na izumrtje.

von Baerov zakon – v embrionalnem razvoju se najprej razvijejo splošne lastnosti skupine. Kasneje se razvijejo specialne lastnosti posameznega taksona. Embriji živali iz istega debla so si v zgodnjih fazah zelo podobni, z razvojem se vse bolj razlikujejo.

Willsonov zakon – organi ali segmenti telesa, ki nastopajo v večjem številu, imajo tendenco reduciranja števila in večje specializacije.

Zaključki

- Morfometrija je kvantitativna analiza oblike in velikosti organizmov s statističnimi orodji. Odnose analiziramo z univariatnimi in bivariatnimi (npr. ANOVA, korelacija, regresija) ter multivariatnimi (npr. analiza glavnih komponent) statističnimi metodami. Rezultate številčnih izračunov kombiniramo z grafičnimi predstavitvami odnosov med skupinami (npr. razsevni diagram, poligon, okvir z ročaji).
- Morfortip opišemo s parametri opisne statistike: $\bar{x} \pm StD, Min - Max, n$.
- S statističnim sklepanjem preverjamo ali se lastnosti morfortipov razlikujejo med skupinami, ki jih primerjamo (npr. glede na spol, starostni razred, geografski vzorec).
- Analiza glavnih komponent PCA je multivariatna metoda, namenjena zmanjšanju razsežnosti osnovnih spremenljivk v glavne komponente PC. Rezultate PCA prikažemo kot razsevni diagram s projekcijo osebkov na prvi dve glavni komponenti; rezultate PCA lahko vključimo v nadaljnje analize.
- Organizmu se v procesu rasti povečuje telesna velikost. Če vsi telesni segmenti rastejo z enako stopnjo, govorimo o izometrični rasti. Pri večini živali se posamezni telesni segmenti spreminjajo z različno hitrostjo, kar privede do spremembe oblike. Takšna rast je alometrična. Z rastjo se volumen organizma povečuje s tretjo potenco v^3 , površina pa s kvadratom dimenzije p^2 . Ta razkorak neposredno vpliva na fiziološke in mehanske lastnosti organizma.
- Fenotip je kombinacija vseh lastnosti organizma. Zajema vse ravni organizacije, fiziološke lastnosti, vedenje in rezultate vedenjskih aktivnosti. Fenotip je izraz genskega zapisa, vpliva okolja in medsebojne interakcije med genetskimi in okoljskimi dejavniki.
- Fenotipska variabilnost je pestrost v obliki in velikosti posameznikov v populaciji. Skupaj z genetsko variabilnostjo je temelj za delovanje selekcije. Fenotipska plastičnost je prilagajanje fenotipa okoljskim razmeram.
- Klina je postopno spreminjanja določene lastnosti vzdolž geografskega gradienta. Nastane zaradi nasprotnega delovanja selekcije in genskega pretoka. Razvije se bodisi s primarno diferenciacijo zaradi prostorske heterogenosti okolja in prilagajanja morfortipa ekološkemu gradientu. Klina se lahko pojavi s sekundarno diferenciacijo kot posledica ponovnega stika predhodno izoliranih populacij, ki sta se diferencirali v alopatriji. Morfološka klina je posledica ponovnega združevanja genomov dveh populacij, posledica česar je vmesni fenotip.
- Odzivi organizmov na okoljske gradiente so praviloma podobni. Iz tega so nastala ekogeografska pravila. Pravila so skrajšane posplošitve in poenostavljene razlage pojavov pri analiziranju ekologije, biogeografije in morfološke plastičnosti organizmov, npr. Bergmannovo pravilo.

Literatura

- Becker R.A., Wilks A.R. 2018. R version by Ray Brownrigg. Enhancements by Thomas P Minka and Alex Deckmyn. maps: Draw Geographical Maps. R package version 3.3.0. <https://CRAN.R-project.org/package=maps>.
- Cline (biology). 2020. Wikipedia. Pridobljeno dne 15. 5. 2020, z [https://en.wikipedia.org/wiki/Cline_\(biology\)](https://en.wikipedia.org/wiki/Cline_(biology)).
- Henry L., Wickham H. 2020. rlang: Functions for Base Types and Core R and 'Tidyverse' Features. R package version 0.4.7. <https://CRAN.R-project.org/package=rlang>.
- Hijmans R.J. 2020. raster: Geographic Data Analysis and Modeling. R package version 3.3-13. <https://CRAN.R-project.org/package=raster>.
- Ivanović A., Kalezić M. 2009. Evolucionarna morfologija. Teorijske postavke i geometrijska morfometrija. Biološki fakultet, Univerzitet u Beogradu. Beograd.
- Kahle D., Wickham H. 2013. ggmap: Spatial Visualization with ggplot2. The R Journal, 5(1): 144–161. URL <http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>.
- Kassambara A. 2020. ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. <https://CRAN.R-project.org/package=ggpubr>.
- Kassambara A., Mundt F. 2020. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7. <https://CRAN.R-project.org/package=factoextra>.
- Klingenberg C.P. 2011. MorphoJ: an integrated software package for geometric morphometrics. Molecular Ecology Resources, 11: 353–357.
- Klingenberg C.P. 2019. Phenotypic Plasticity, Developmental Instability, and Robustness: The Concepts and How They Are Connected. *Frontiers in Ecology and Evolution*, 7:56. doi: 10.3389/fevo.2019.00056.
- Legendre P., Legendre L. 2012. Numerical Ecology. Elsevier. Amsterdam.
- O'Donnell M.S., Ignizio D.A. 2012. Bioclimatic predictors for supporting ecological applications in the conterminous United States: U.S. Geological Survey Data Series 691, 10 p.
- Phenotypic plasticity. 2019. Wikipedia. Pridobljeno dne 30. 11. 2019, z https://en.wikipedia.org/wiki/Phenotypic_plasticity.
- R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Slowikowski K. 2020. ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.8.2. <https://CRAN.R-project.org/package=ggrepel>.
- Valladares F., Sanchez-Gomez D., Zavala A.M. 2006. Quantitative estimation of phenotypic plasticity: bridging the gap between the evolutionary concept and its ecological applications. *Journal of Ecology*, 94: 1103–1116. doi: 10.1111/j.1365-2745.2006.01176.x.
- Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
- Wickham H., François R., Henry L., Müller K. 2020. dplyr: A Grammar of Data Manipulation. R package version 1.0.2. <https://CRAN.R-project.org/package=dplyr>.
- Zelditch M.L., Swiderski D.L., Sheets H.D. 2012. Geometric Morphometrics for Biologists: A Primer. Elsevier: Academic Press. London.

12. poglavje



NUMERIČNA TAKSONOMIJA

Povzetek

12. poglavje: Numerična taksonomija

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

Taksonomija je biološka disciplina o poimenovanju, definiranju in klasifikaciji organizmov na podlagi njihovih skupnih lastnosti. Organizme združujemo v taksonne in jih razvrščamo v taksonomske kategorije. Glavne taksonomske kategorije so: domena, kraljestvo, deblo, razred, red, družina, rod in vrsta. Numerična taksonomija je klasifikacija organizmov na podlagi njihovih lastnosti z numeričnimi metodami.

Razlike med skupinami prepoznavamo in vrednotimo z univariatnimi in multivariatnimi statističnimi analizami. Z numeričnim pristopom se izognemo subjektivnemu opisovanju značilnosti posamezne taksonomske skupine, zato so ugotovitve veliko bolj verodostojne. Iz izmerjenih vrednosti oblikujemo matriko, kjer so objekti (osebki ali skupine) vrstice, njihove lastnosti so stolpci. Posamezni objekt je opisan z več spremenljivkami. Numerično analizo ene spremenljivke imenujemo univariatna analiza. Če istočasno izvajamo numerično analizo več spremenljivk, govorimo o multivariatni analizi.

Redukcijo razsežnosti multivariatnega podatkovnega niza izvedemo z analizo glavnih komponent, v kateri strnemo skupno variabilnost več spremenljivk v nekaj dimenzij, ki se imenujejo glavne komponente.

V diskriminantni analizi so skupine definirane vnaprej (*a priori*), analiza med spremenljivkami išče takšno kombinacijo lastnosti, ki podpira apriorno delitev na skupine.

Metoda voditeljev je multivariatna statistična metoda, s katero objekte cepimo v skupine. Optimiranje števila skupin preverjamo z metodo komolca, silhuetno metodo in permutacijskim testom.

Taksonomija (ang. Taxonomy) je biološka disciplina o opisovanju in klasifikaciji organizmov. Takson je taksonomska enota ne glede na hierarhični rang. Taksone na določenem rangju združimo v višji rang, s čimer vzpostavimo taksonomsko hierarhijo. Glavne taksonomske kategorije v sodobni taksonomiji so: domena, kraljestvo, debla, razred, red, družina, rod in vrsta. Carl Linnaeus (1707–1778) je bil švedski botanik, zoolog, taksonom in zdravnik, ki je uvedel binarno poimenovanje organizmov, nadaljnji razvoj področja je znan kot Linnejeva taksonomija.

Mednarodna komisija za zoološko nomenklaturu (ang. International Commission on Zoological Nomenclature ICZN) vodi enoten sistem zoološkega poimenovanja. S tem zagotavlja, da imajo živalske vrste edinstveno in splošno sprejeto ime. Pravila poimenovanja ureja **Mednarodni kodeks zoološke nomenklature** (ang. International Code of Zoological Nomenclature ICZN).

Numerična taksonomija (ang. Numerical Taxonomy) je klasifikacija organizmov na podlagi njihovih lastnosti z numeričnimi metodami. Razlike med skupinami prepoznavamo in vrednotimo z ustreznimi numeričnimi in statističnimi analizami. S tem nadgradimo subjektivno opisovanje značilnosti posamezne taksonomske skupine.

V tem poglavju predstavljamo tri numerične pristope v taksonomski analizi: 1. redukcija večrazsežnih meritev, 2. razvrščanje organizmov v skupine in 3. razlikovanje med skupinami. Začetni korak je izvedba meritev telesnih lastnosti. Praviloma izvedemo večje število bolj ali manj standardiziranih meritev na množici objektov. Izbrane lastnosti organizma (npr. telesna dolžina in masa, dolžina stopala, dolžina lobanje ipd.) povzamemo iz večjega števila osebkov. Kadar istočasno izvedemo analizo numeričnih vrednosti dveh ali več spremenljivk, govorimo o multivariatni analizi.

Podatke oblikujemo v pravokotno matriko, kjer so v vrsticah objekti O (osebki živali), izmerjene vrednosti x_{ij} njihovih lastnosti so v stolpcih X (spremenljivke). Posamezen objekt je opisan z več spremenljivkami, kot je predstavljeno v matriki:

| ID | O_1 | X_1 | X_2 | X_3 | ... | X_p |
|--------|----------|----------|----------|----------|-----|----------|
| Id_1 | O_{1j} | X_{11} | X_{12} | X_{13} | ... | X_{1p} |
| Id_2 | O_{2j} | X_{21} | X_{22} | X_{23} | ... | X_{2p} |
| Id_3 | O_{3j} | X_{31} | X_{32} | X_{33} | ... | X_{3p} |
| Id_4 | O_{4g} | X_{41} | X_{42} | X_{43} | ... | X_{4p} |
| Id_5 | O_{5g} | X_{51} | X_{52} | X_{53} | ... | X_{5p} |
| Id_6 | O_{6g} | X_{61} | X_{62} | X_{63} | ... | X_{6p} |
| Id_7 | O_{7g} | X_{71} | X_{72} | X_{73} | ... | X_{7p} |
| ... | ... | ... | ... | ... | ... | ... |
| Id_p | O_{1p} | X_{p1} | X_{p2} | X_{p3} | ... | X_{pp} |

Prvi stolpec je indikator ID z inventarno številko osebkov na katerem so izvedene meritve, po navadi gre za muzejsko inventarno številko. Indikator je pomembna informacija za sledljivost in možnost naknadnega preverjanja meritev in ni namenjen statističnemu vrednotenju.

Redukcijo razsežnosti izvedemo z metodo, imenovano analiza glavnih komponent, to je multivariatna statistična metoda iz družine ordinacijskih metod. V redukciji razsežnosti strnemo skupno variabilnost množice spremenljivk v nekaj novih dimenzij, pri čemer prva dimenzija, ki se imenuje prva glavna komponenta, razloži največji delež variabilnosti osnovnih spremenljivk. To omogoča, da v nadaljnjih analizah ne operiramo z množico spremenljivk, temveč jih izvajamo na manjšem številu dimenzij.

Metoda voditeljev je multivariatna statistična metoda iz družine klasičnih metod, s katero objekte cepimo v skupine. Algoritem objekte razvrsti v skupine, število skupin določi raziskovalec. Vrednotenje grupiranja je v presoji raziskovalca.

V diskriminantni analizi definiramo skupine vnaprej (*a priori*), algoritem pa išče najustreznejšo kombinacijo spremenljivk, ki podpira apriorno delitev na skupine. Metoda se imenuje diskriminantna analiza in je iz družine ordinacijskih metod, njen namen je iskanje razlik med skupinami in klasifikacija osebkov v eno od določenih skupin.

V tem poglavju je na primeru taksonomske analize puščavskih skakačev, glodavcev iz rodu *Pygeretmus*, predstavljena uporaba naslednjih analiz:

- Analiza glavnih komponent
- Metoda voditeljev
- Diskriminantna analiza.

Poglavje o numerični taksonomiji je narejeno po literaturi Jesenko in Jesenko 2007, Kassambara in Mundt 2020, Klingenberg 2011, Legendre in Legendre 2012, Sneath in Sokal 1973.



PROJEKT 12.1

Taksonomska analiza skakačev vrste *Pygeretmus pumilio*

Pritlikavi debelorepi skakač *Pygeretmus pumilio* je majhen skakač, z maso okrog 45 g (Slika 12.1) iz družine Dipodidae, ki živi v glinenih puščavskih in polpuščavskih habitatih centralne Azije in JV Evrope (Slika 12.2). Je nočno aktiven, prehranjuje se z zelišči in je fakultativni hibernator. Vrsta ima velik areal razširjenosti z okrog 5.400 km razpona

geografske dolžine in okrog 1.800 km razpona geografske širine. Vrsta je politipska s petimi podvrstami, ki so združene v dve skupini, imenovani *pumilio* in *potanini*. Taksonomija temelji na obarvanosti kožuha in obliki penisa.

Kryštufek in sod. (2019 in 2021) so v taksonomsko analizo vključili spodnje čeljustnice (mandibule). Material so pridobili v muzejih v Moskvi in St. Peterburgu. Mandibule so izmerili in analizirali z metodami geometrijske morfometrije. Oblikovali so 22 geografskih vzorcev oz. OTU (ang. Operational Taxonomic Unit), (Slika 12.2). Namen raziskave je preveriti, ali vsebujejo mandibule taksonomsko informacijo.



Slika 12.1 Pritlikavi debelorepi skakač (levo). Foto. Nedko Nedyalkov. Mandibula z labialne (a) in lingvalne (b) strani z označenimi strukturami, prirastišči za mišice in oslonilnimi točkami (oštevilčene pike), (desno), (vzeto iz Kryštufek in sod. 2019).



Slika 12.2 Razširjenost *Pygeretmus pumilio* po IUCN – rumeni poligon. 22 lokalitet, od koder izvirajo OTU z vzorci mandibul, izdelal Danijel Ivajnsič.



12.1 Analiza glavnih komponent

Analiza glavnih komponent PCA (ang. Principal Component Analysis) je najpogosteje uporabljena multivariatna metoda v morfometriji. Glavni namen metode je zmanjšati število osnovnih spremenljivk. Metoda PCA temelji na povezavah med osnovnimi spremenljivkami, torej na korelacijah (ali na kovariancah) med njimi. Višja kot je stopnja korelacije med spremenljivkami (pozitivna ali negativna), učinkovitejša bo analiza. Z metodo PCA izračunamo nove spremenljivke, ki se imenujejo **glavne komponente** PC (ang. Principal Components). Število PC je enako številu osnovnih spremenljivk (v primeru, ko je število objektov manjše od števila osnovnih spremenljivk, potem je število glavnih komponent enako številu objektov). Značilnost vseh PC je njihova orientiranost v multivariatnem prostoru. Prva glavna komponenta PC1 poteka skozi multivariatni prostor v smeri, ki razloži največji delež skupne variance osnovnih spremenljivk. Druga glavna komponenta PC2 je pravokotna na PC1 in razlaga naslednji največji skupni delež variance. Sledijo preostale komponente, dokler ni razložena vsa varianca. Glavne komponente so torej nove spremenljivke, razvrščene po deležu variance, ki jo razlaga, PC1 razloži največji delež skupne variabilnosti, sledi PC2, ki razloži naslednji najvišji delež itd. Ker so komponente pravokotne med seboj, vsaka od njih opisuje novo »dimenzijo« v variabilnosti podatkov.

Eden od pogojev za učinkovito analizo glavnih komponent PCA je multivariatna normalna porazdelitev spremenljivk (sin. multinormalnost; ang. Multivariate Normal Distribution). Multinormalnost preverimo s testom normalnosti posamezne spremenljivke, ki vključuje grafične in numerične analize: box plot, bi plot, qq plot, histogram ter analizo sploščenosti in asimetričnosti ali Mardianov test. Asimetričnost porazdelitve v primeru pozitivne asimetrije korigiramo s transformacijo z desetiškim logaritmom in v primeru negativne asimetrije s transformacijo s korenjenjem. Naslednji pogoj je primerljivost absolutnih vrednosti med posameznimi spremenljivkami. Če ima npr. ena od spremenljivk razpon vrednosti od 1 do 10 in druga od 100 do 1000, bo variabilnost absolutno večje spremenljivke zakrila variabilnost absolutno manjše spremenljivke. Zagotovitev enakovrednega prispevka vseh spremenljivk v multivariatni obravnavi dosežemo s **standardizacijo**, ki jo računamo po formuli:

$$z = \frac{x - \bar{x}}{StD},$$

kjer so z – standardizirana vrednost, x – konkretna vrednost meritve, \bar{x} – aritmetična sredina spremenljivke, StD – standardna deviacija spremenljivke.

V nadaljnjo analizo sprejmemo samo nekaj prvih glavnih komponent. Pri odločitvi, koliko glavnih komponent je smiselno vključiti naprej, si pomagamo s cenilkami: – lastna vrednost zadnje PC je večja od 1; – zadnja PC razloži vsaj 5 % variabilnosti; – kumulativna

vrednost razložene variabilnosti komponent je vsaj 80 %; – metoda komolca, ki poda informacijo, pri kateri PC se delež razlaga variance bistveno ne poveča več, mestu, kjer se krivulja prelomi pravimo komolec.

Rezultate analize glavnih komponent PCA po navadi prikažemo kot razsevni diagram s projekcijo osebkov (objektov) na prvi dve ali tri PC osi. Vrednosti glavnih komponent PC lahko vključimo v nadaljnje analize oz. testiranja. Dobra plat metode je, da so razmerja med objekti realno prikazana in da med komponentami ni korelacije.

Rezultati PCA so naslednji:

- **Lastne vrednosti** (ang. Eigenvalues) – so uteži in deleži razložene variance, za posamezno glavno komponento. Vsota lastnih vrednosti je enaka številu spremenljivk, vključenih v analizo, oz. številu dimenzij.
- **Uteži** (ang. Loadings) – so vrednosti korelacije med osnovnimi spremenljivkami in glavnimi komponentami. Prispevek osnovnih spremenljivk k izračunu posamezne PC merimo z velikostjo korelacijskega koeficienta med osnovnimi spremenljivkami in PC. S tem tudi osmislimo informacijo posamezne PC.
- **Glavne komponente** (ang. Principal Components – PC) – so nove spremenljivke, oz. vrednosti (ang. Scores) posameznega objekta. Predstavimo jih številčno, lahko jih vključimo v grafično predstavitev (bi plot, box plot) ali v nadaljnje analize.
- **Projekcija lastnih vrednosti** (ang. Scree Plot) – je projekcija deleža lastnih vrednosti po glavnih komponentah. Metoda določi »komolec« (ang. Broken Stick) za prepoznavanje med skupino pomembnih in skupino manj pomembnih glavnih komponent.
- **Razsevna projekcija objektov (osebkov) na prvi dve glavni komponenti** (ang. Scatter Plot) – s projekcijo predstavimo morfoprostor, ki ga zasedajo objekti. Objekte, ki pripadajo isti skupini, lahko povežemo v skupni poligon, npr. z metodo najmanjšega konveksnega poligona (ang. Convex Hull).

Za interpretacijo rezultatov je pomembno razumevanje vsebine glavnih komponent PC, ki povzemajo skupno informacijo o lastnostih osnovnih spremenljivk. Glavne komponente dobijo informacijsko vrednost s tem, ko jih vključimo v korelacijsko analizo z osnovnimi spremenljivkami. Močna korelacija med PC in osnovno spremenljivko pomeni, da je večina informacije te osnovne spremenljivke povzeta v obravnavani PC, oz. da ta spremenljivka, ali skupina spremenljivk, v največji meri definira obravnavano komponento. Če je korelacija pozitivna, jo interpretiramo tako, da z naraščanjem vrednosti PC narašča tudi vrednost izbrane spremenljivke, in obratno. Če je korelacija negativna, z naraščanjem vrednosti PC vrednost obravnavane spremenljivke upada. V morfometriji

praviloma vsebinsko osmislimo prvi dve komponenti PC1 in PC2, ki pogosto razložita več kot polovico variance (informacije) v celotnem naboru podatkov.

Prikaz uporabe PCA je predstavljen v spodnjem Primeru 12.3 in v poglavju Morfološka plastičnost.

Podrobnejša matematična razlaga je podana v Jesenko in Jesenko (2007), Legendre in Legendre (2012) ter Sokal in Rohlf (2012).



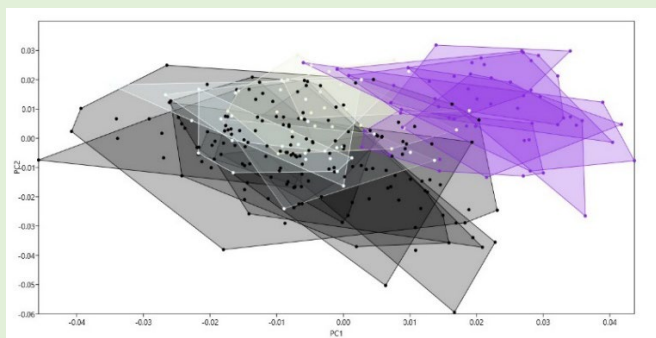
PRIMER 12.1

Analiza glavnih komponent

V analizo morfometrične variabilnosti smo vključili meritve 283 mandibul. Na vsaki smo izvedli meritve 14 oslonilnih točk. To pomeni, da je bilo skupno 28 spremenljivk, ki smo jih vključili v Prokrustovo analizo in postopek izvedli v programu MorphoJ. Prokrustove koordinate smo vključili v analizo glavnih komponent po metodi kovariance. Rezultat PCA je skupaj 26 dimenzij. Lastne vrednosti za prvih 10 so predstavljene v Preglednici 12.1. Vrednostih prvih dveh dimenzij smo predstavili kot bi plot, na katerem smo naknadno s poligoni povezali osebke, ki pripadajo istemu geografskemu vzorcu (Slika 12.3).

Preglednica 12.1 Lastne vrednosti in delež razložene variance za prvih 10 glavnih komponent

| Komponenta | Lastna vrednost | % razložene variance | % kumulativno razložena varianca |
|------------|-----------------|----------------------|----------------------------------|
| 1. | 0,00027811 | 15,952 | 15,952 |
| 2. | 0,00024301 | 13,938 | 29,890 |
| 3. | 0,00016164 | 9,271 | 39,161 |
| 4. | 0,00013366 | 7,666 | 46,827 |
| 5. | 0,00013180 | 7,560 | 54,387 |
| 6. | 0,00011826 | 6,783 | 61,170 |
| 7. | 0,00008845 | 5,073 | 66,243 |
| 8. | 0,00008151 | 4,675 | 70,918 |
| 9. | 0,00006492 | 3,724 | 74,642 |
| 10. | 0,00005583 | 3,202 | 77,844 |



Slika 12.3 Projekcija 283 osebkov (črne pike) in poligonov 22 OTU v morfoпростor prvih dve glavnih komponent PC1-PC2, ki skupaj razložita 29,9 % variabilnosti v obliki mandibul. Sivo: skupina *pumilio*, vijoličasto: skupina *potanini*.



12.2 Metoda voditeljev

Metoda voditeljev (ang. K-means) je multivariatna statistična metoda iz družine klasterških analiz. Namen metode je cepitev osnovne množice objektov v vnaprej izbrano število skupin. Pri tem domnevamo, da skupine v resnici obstajajo in so prepoznavne. Algoritem deluje na t.i. voditeljih, ki določajo lastnosti skupin in so centriodi skupin. Postopek optimalne cepitve poteka po korakih:

1. Izbira števila voditeljev; število voditeljev izbere raziskovalec v skladu z domnevo, ki jo preverja; lahko pa število skupin določi algoritem za določanje skupin.
2. Vsak objekt je vključen v skupino k najbližjemu voditelju, in s tem se oblikujejo skupine.
3. Skupinam izračunamo nove centroide.
4. K novim centroidom ponovno vključujemo objekte.
5. Postopka 3 in 4 iteriramo (ponavljamo), dokler se spreminja položaj centroida in se objekti selijo iz skupine v skupino.

Za srednje velike primere ($n \approx 1.000$) je potrebnih manj kot 100 iteracij. Rezultata analize sta: 1. prepoznanje skupin; 2. razporeditev objektov v skupine.

Metoda temelji na izračunu razdalj med centroidom in posameznim objektom, na kar močno vpliva absolutna vrednost spremenljivk. Razlike v absolutnih vrednostih med spremenljivkami odpravimo s standardizacijo. S tem postopkom uravnotežimo prispevek posamezne spremenljivke k cepitvi v skupine.

Šibkost in obenem prednost metode je potreba po vnaprejšnjem definiranju števila skupin. V primeru taksonomske raziskave, ko preverjamo hipotezo, ali se osebki delijo v dve skupini (npr. dve vrsti), odločitev o številu skupin izhaja iz raziskovalnega vprašanja. Ko pa ne poznamo števila dejanskih skupin, lahko ugibamo in preverjamo modele z različnim številom skupin. Na splošno iščemo takšno cepitev v skupine, kjer so si objekti znotraj skupine čim bolj podobni, objekti iz različnih skupin pa čim bolj različni. Podobnost objektov znotraj skupine se imenuje **kohezija**, različnost objektov med skupinami je **ločljivost**. Poznamo dve skupini metod za optimiranje števila skupin: 1. neposredne metode (metoda komolca in silhuetna metoda) in 2. metode statističnega testiranja (permutacijski test).

Silhuetna metoda (ang. Average Silhouette Method) temelji na koheziji. Kakovost cepitve, ki vključuje tako kohezijo kot ločljivost, ocenimo s silhuetnim koeficientom s , ki ima zalogo vrednosti od -1 do $+1$. Metoda temelji na razmerju med razdaljami med objekti znotraj skupine, glede razdalje do objektov v sosednji skupini. Razdalja znotraj skupine je

vsota vseh razdalj med naključnim objektom x_i iz skupine i in vsemi preostalimi objekti te skupine, to je razdalja a_i (notranja razpršenost). Razdalje do sosednje skupine je vsota vseh razdalj med objektom x_i in vsemi objekti iz najbližje sosednje skupine, to je razdalja b_i . Silhueta skupine i se izračuna po formuli:

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)},$$

kjer je s_i – silhuetni koeficient skupine i , a_i – notranja razpršenost skupine i , b_i – zunanja razpršenost skupine i . Silhueta cepitve se računa kot povprečna silhueta po enačbi:

$$s = \frac{1}{K} \sum_{i=1}^K S_i,$$

kjer je s – silhuetni koeficient, K – število skupin. Optimalno število skupin maksimira povprečno silhueto. Pričakujemo, da je $a_i < b_i$, zato je $s_i > 0$. Objekti s silhuetnim koeficientom blizu 1 kažejo na kakovostno kohezijo, vrednosti okrog 0 so mejni primeri, vrednosti manjše od 0 in v bližini -1 pa pomenijo, da je ločljivost slaba, oz. da skupine niso utemeljene.

Metoda komolca (ang. Elbow Method) temelji na čim manjši celokupni znotraj skupinski varianci (ang. Total Within-cluster Sum of Squares WSS). Celokupna WSS je merilo kompaktnosti skupin, za katero želimo, da je čim manjša, ker to omogoča razlikovanje med skupinami. V metodi komolca je WSS funkcija števila skupin, vsaka dodana skupina zmanjša varianco. Dodajanje novih skupin bistveno vpliva na spremembo variance, optimalno število skupin je doseženo, ko dodatek nove skupine ne vpliva več bistveno na varianco. V grafičnem prikazu je dosežen »komolec«.

Permutacijski test (ang. Gap Statistic Method). Iskanje števila skupin lahko izvedemo s permutacijskim testom. Preverjamo ničelno hipotezo, po kateri se število dobljenih skupin ne razlikuje od števila skupin, ki bi jih dobili, če bi bile vrednosti naključno porazdeljene. Po alternativni hipotezi se dobljena cepitev razlikuje od naključnega. Test izvedemo tako, da podatke iz spremenljivk, vključenih v analizo, med seboj naključno pomešamo. Izvedemo cepitev v skupine in ponovimo postopek več 100-krat. Skupine, ki jih dobimo iz dejanskih podatkov, primerjamo s skupinami iz naključno porazdeljenih vrednosti. Če ugotovimo, da se katera od rešitve iz naključnih porazdelitev pogosteje pojavlja, ohranimo ničelno hipotezo in sklenemo, da izbrano število skupin ni ustrezno. Kadar pa nobena od oblikovanih cepitev ne presega naključne vrednosti, sprejmemo alternativno hipotezo, da je cepitev na predlagane skupine verjetnejša od naključne cepitve.

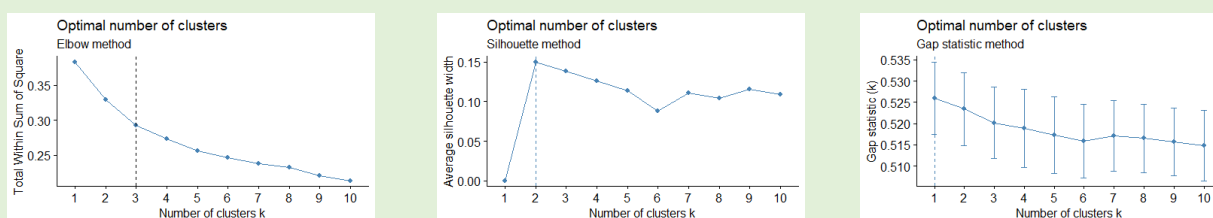
Podrobnejša matematična razlaga je podana v Jesenko in Jesenko (2007), Legendre in Legendre (2012) ter Sokal in Rohlf (2012).



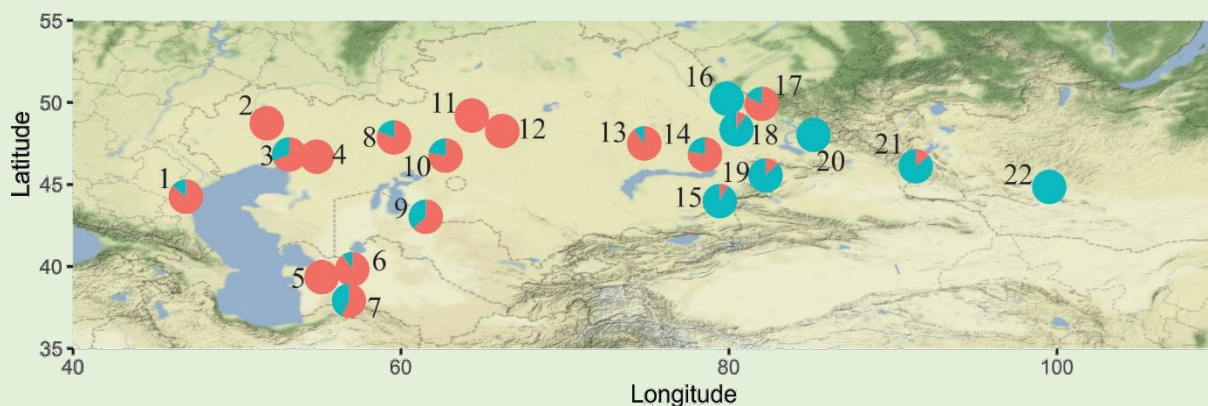
PRIMER 12.2

Metoda voditeljev – število skupin

V prepoznavanje števila skupin z metodo voditeljev smo vključili prvih 10 glavnih komponent iz predhodne analize, predstavljene v Primeru 12.1. Matrika je imela dimenzijo 283 vrstic (objekti) krat 10 stolpcev (spremenljivke). Cepitev v skupine smo preverili z vsemi metodami: s silhuetno metodo, z metodo komolca in s permutacijskim testom. Prepoznavanje skupin po različnih metodah je bilo različno, metoda komolca je prepoznala 3 skupine, silhuetna metoda 2 in permutacijski test eno skupino (Slika 12.4). Iz predhodnih taksonomskih študij skakačev, kjer so na podlagi obarvanosti kožuha in oblike penisa prepoznali dve skupini (*pumilio* in *potanini*), smo tudi v primeru te analize sprejeli odločitev za cepitvi v dve skupini.



Slika 12.4 Prepoznavanje optimalnega števila skupin z metodo komolca (levo), s silhuetno metodo (sredina) in permutacijskim testom (desno).



Slika 12.5 Projekcija 22 OTU. S krožnimi izseki je predstavljen delež skupine 1 (rdeč – *pumilio*) in skupine 2 (modro – *potanini*).

Metoda voditeljev, z vnaprej določenima dvema skupinama ($K = 2$), je glede na vrednosti prvih 10 GK, razvrstila vsak osebek v eno od obeh skupin. Za vsak OTU (geografski vzorec) posebej smo preverili deleže razporeditve po skupinah in jih predstavili s krožnimi izseki; rdeča barva je delež osebkov razvrščenih v 1. skupino, modra barva je delež osebkov razvrščenih v 2. skupino (Slika 12.5). Krožni izseki so razporejeni na geografski karti (Slika 12.5). Delitev v dve skupini se sklada z geografsko razširjenostjo obeh taksonomskih skupin. Areal skupine *pumilio* je zahodno, skupine *potanini* pa vzhodno od približno 80° geografske dolžine.

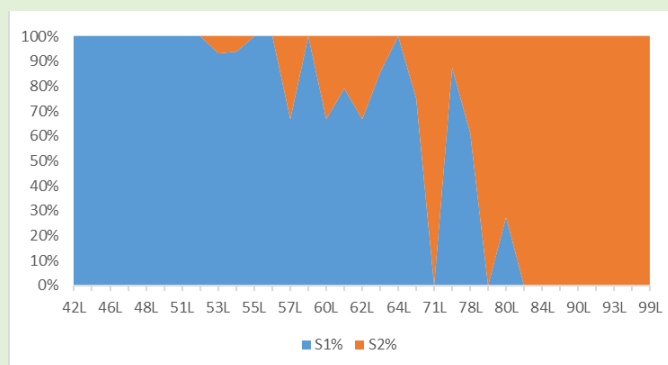
```
R >
#Izračun K mean klastra za k=2
rezultat <- kmeans(Py, 2)
rezultat
table(pu$ssp, rezultat$cluster)
```



PRIMER 12.3

Metoda voditeljev – prostorski vzorec

V drugem delu analize smo povečali število osebkov na 341 in ponovili analizo z metodo voditeljev s $K = 2$. Cepitev v dve skupini smo projicirali na geografsko dolžino (Slika 12.6). V skrajnih delih areala, na zahodu med 42° in 52° ter na vzhodu med 82° in 99° geografske dolžine, je metoda razvrstila osebkve samo v eno skupino, ali v *pumilio* ali *potanini*, v vmesnem območju med 53° in 81° , pa je prepoznavala postopen prehod med skupinama.



Slika 12.6 Frekvenca zastopanosti dveh klastrov vzdolž geografske dolžine od 42° do 99° .

```
R > package = »factoextra«
```

```
#Določanje števila klastrov. Vnos in pregled podatkov, pu = ves material Pp n=341
```

```
Py$ssp <- NULL
```

```
Py$LON <- NULL
```

```
#Metoda komolca
```

```
fviz_nbclust(Py, kmeans, method = "wss") + geom_vline(xintercept = 4, linetype = 2) + labs(subtitle = "Elbow method")
```

```
# Silhuettna metoda
```

```
fviz_nbclust(Py, kmeans, method = "silhouette") + labs(subtitle = "Silhouette method")
```

```
# Gap statistic
```

```
# nboot = 50 to keep the function speedy.
```

```
set.seed(123)
```

```
fviz_nbclust(Py, kmeans, nstart = 25, method = "gap_stat", nboot = 50) + labs(subtitle = "Gap statistic method")
```


Podatki, vključeni v diskriminantno analizo, morajo izpolnjevati naslednje pogoje. Število skupin je vsaj dve ali več; vsaka skupina mora vsebovati najmanj dva objekta, optimalno število je okrog 30 objektov v skupini; skupine morajo imeti primerljivo število objektov; skupno število objektov mora biti višje od števila razlagalnih spremenljivk. Zagotovitev enakovrednega prispevka vseh razlagalnih spremenljivk v multivariatni obravnavi dosežemo z njihovo \bar{x} standardizacijo.

V primeru obravnave samo dveh skupin se metoda imenuje enostavna diskriminantna analiza (ang. Simple Diskriminant Analysis), če pa so skupine tri ali več, je to multipla diskriminantna analiza (ang. Multiple Diskriminant Analysis), ali analiza kanoničnih spremenljivk (ang. Canonical Variate Analysis).

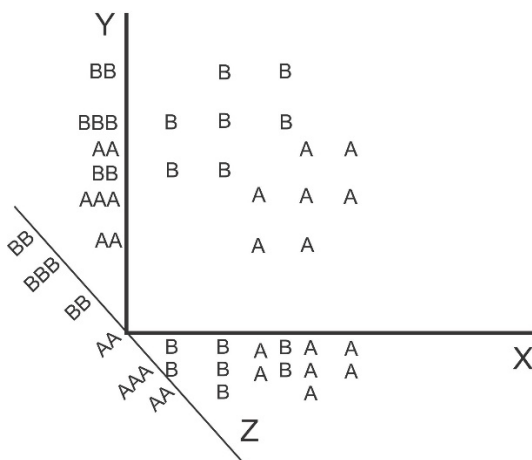
Diskriminantna analiza poteka v dveh stopnjah. 1. Prvi korak je preverjanje razlik v razlagalnih spremenljivkah glede na odvisno spremenljivko. Testiramo ali se njihova povprečja razlikujejo glede na skupine v odvisni spremenljivki. Rezultat testiranja je Wilksova lambda, ki podaja delež nerazložene variance med skupinami; manjša je Wilksova lambda, uspešnejša je analiza. Ta prvi korak v diskriminantni analizi je enak multivariatni analizi variance MANOVA. 2. Če se vrednosti razlagalnih spremenljivk med skupinami značilno razlikujejo, nadaljujemo analizo. Algoritem išče takšno linearno kombinacijo razlagalnih spremenljivk, ki maksimalno poveča razlike med skupinami. Iskanje linearne kombinacije imenujemo inverzna analiza. Skupine v odvisni spremenljivki so kriterij, ki ga razlagamo z razlagalnimi spremenljivkami.

V diskriminantni analizi oblikujemo linearne kombinacije razlagalnih spremenljivk v nove spremenljivke, imenovane **diskriminantne funkcije DF**. Funkcije DF so nove »sintetične« spremenljivke, ki razložijo razliko med skupinami. Njihovo število n pa je za ena manjše od števila skupin, torej:

$$nDF = K - 1.$$

Za diskriminantne funkcije je značilna orientiranost v multivariatnem prostoru. Prva diskriminantna funkcija DF1 poteka skozi multivariatni prostor v smeri, ki razloži največji delež razlik med skupinami. Diskriminantna funkcija DF2 je pravokotna na DF1 in razlaga naslednji največji delež razlik med skupinami, sledijo še preostale DF, dokler ni razložena vsa razlika med skupinami. Namen diskriminantnih funkcij je oblikovanje klasifikacijskega kriterija za uvrščanje objektov v napovedane skupine (identifikacija). Druga vloga diskriminantnih funkcij je prepoznavanje razlagalnih spremenljivk, ki najbolj prispevajo k uspešni diskriminaciji. Prikaz oblikovanja linearne kombinacije dveh spremenljivk v diskriminantno os, ki uspešno razlikuje med dvema skupinama, je predstavljen na Sliki

12.7. Slaba plat metode je, da so razlike med skupinami prikazane precejeno (preveč optimistično). Zato zanesljivost razlikovanja dodatno preverimo s prečnim preverjanjem in z Bonferronijevim popravkom.



Slika 12.7 Dve skupini, A in B, vsaka zastopana s 7 objekti, se prekrivata tako vzdolž osi X in Y , kar je prikazano s histogrami ob oseh, vendar sta skupini popolnoma ločeni vzdolž diskriminantne osi Z . Položaj objekta i vzdolž osi Z izračunamo po enačbi:

$$z_i = \frac{\sqrt{2}}{2}(x_i - y_i)$$

(Povzeto po Legendre in Legendre 2012).

Rezultati LDA so naslednji:

- **Wilksova lambda** in pripadajoče p vrednosti.
- **Diskriminantne funkcije DF** – linearne kombinacije razlagalnih spremenljivk, posameznemu objektu so prirejene nove vrednosti (ang. Scors). Vrednosti DF bodisi prikažemo numerično ali grafično z razsevnim diagramom ali okvirjem z ročaji bodisi vključimo v nadaljnje analize. Število diskriminantnih funkcij je za ena manjše od števila skupin.
- **Okvir z ročaji** (ang. Box Plot) – vrednosti izbrane DF prikažemo po skupinah grafično s parametri opisne statistike (povprečje \pm standardna deviacija \pm minimum / maksimum).
- **Razsevni diagram** (ang. Bi Plot) – rezultate LDA po navadi prikažemo kot razsevni diagram s projekcijo objektov na prvi dve ali tri DF. Objekte posamezne skupine povežemo z **minimalnim konveksnim poligonom** (ang. Convex Hull), s čimer predstavimo razmerja morfoprostorov med skupinami.
- **Lastne vrednosti** – (ang. Eigenvalues) in deleže razložene variance za posamezno DF.
- **Uteži** – (ang. Loadings) korelacije med osnovnimi spremenljivkami in DF. Jakost povezave oz. prispevek posamezne razlagalne spremenljivke k tvorbi posamezne DF. To je informacija o diskriminantni moči posamezne razlagalne spremenljivke.
- **Klasifikacija** – klasifikacija posameznih enot iz dejanske skupine v eno od v vnaprej določenih skupin.
- **Klasifikacijska preglednica** – (ang. Classification or Confusion Table) je pregled uspešnosti klasifikacij iz dejanskih v napovedane skupine.

- **Mahalanobisove razdalje** – (ang. Mahalanobis Distances) so razdalje med objekti v diskriminantnem prostoru. Prikažemo jo v obliki simetrične matrike Mahalanobisovih razdalj med centri (središči) skupin.
- **Bonferronijev popravek** – (ang. Bonferroni Correction) zadeva problem multiplih primerjav. Če istočasno preverjamo več, to je k ničelnih hipotez, je verjetnost, da bomo zavrnili vsaj eno ničelno hipotezo, enaka $1 - (1 - \alpha)^k$. Za korekcijo izvedemo Bonferronijev popravek: $\alpha_c = \alpha/k$, slednje pomeni, da vrednost 0,05 delimo s k in dobimo novo mejno vrednost za odločitev o hipotezi.
- **Prečno preverjanje** – (ang. Leave-one-out Cross-validation LOOCV) zadeva primerjanje med dejansko klasifikacijo in naključno klasifikacijo osebkov in je namenjena preverjanju relevantnosti klasifikacij.

Podrobnejša matematična razlaga je podana v Jesenko in Jesenko (2007), Legendre in Legendre (2012) ter Sokal in Rohlf (2012).



PRIMER 12.4

Diskriminantna analiza 1 – MANOVA

Vhodna matrika je imela dimenzijo 341 vrstic (osebkov) in 30 stolpcev (Prokrustovih koordinat) – te so bile razlagalne spremenljivke. Odvisna spremenljivka je bila taksonomska pripadnost oz. pet vnaprej določenih skupin (podvrst). V prvem delu diskriminantne analize izračunamo vrednost Wilksove lambde $\lambda = 0,165$ in MANOVA testa $F_{104, 1236} = 6,845$ $p < 0,0001$. Vrednost Wilksove lambde pomeni, da je ostalo 16,5 % medvrstne variabilnosti nerazložene, medtem ko so razlike med skupinami visoko značilne ($p < 0,05$, Preglednica 12.2). Zaradi multiplih primerjav med petimi podvrstami izvedemo Bonferronijev popravek. Izračunamo korigirano p vrednost na način, da mejno vrednost 0,05 delimo s 5, to je število skupin in nova kritična vrednost $p = 0,01$. Vrednosti v Preglednici 12.2 so vse manjše od 0,01 kar pomeni, da se vse skupine med seboj statistično značilno razlikujejo. Na podlagi potrjene razlike med skupinami nadaljujemo z drugim delom analize.

Preglednica 12.2 p vrednosti iz analize MANOVA

| | <i>bulatoides</i> | <i>iliensis</i> | <i>potanini</i> | <i>pumilio</i> |
|-----------------|-------------------|-----------------|-----------------|----------------|
| <i>iliensis</i> | 0,0001 | | | |
| <i>potanini</i> | <0,0001 | <0,0001 | | |
| <i>pumilio</i> | <0,0001 | <0,0001 | <0,0001 | |
| <i>selevini</i> | <0,0001 | <0,0001 | 0,0062 | <0,0001 |

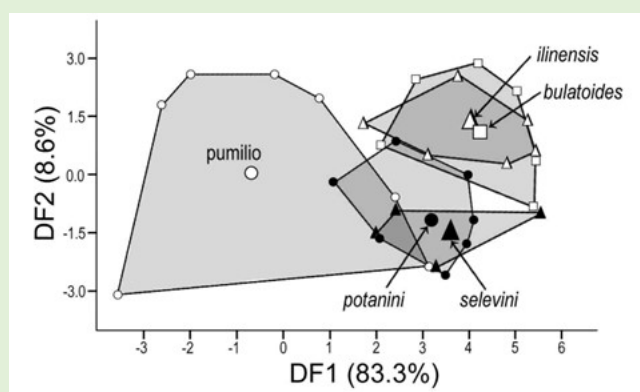




PRIMER 12.5

Diskriminantna analiza – klasifikacija

Lastna vrednost prve diskriminantne funkcije ima vrednost $DF1 = 2,65$ in razlaga 83,3 % variabilnosti med skupinami, lastna vrednost $DF2 = 0,27$ in razlaga 8,6 % variabilnosti. V drugem delu diskriminantne analize smo opravili klasifikacijo osebkov iz dejanskih v napovedane skupine (Preglednica 12.3). Klasifikacija je bila pravilna v 96,5 % primerov, prečno preverjanje pravilnosti klasifikacij je vrnilo 93,5 % pravilnih klasifikacij. Razlika med izvorno klasifikacijo in z izpustitvijo enega osebkov po metodi prečnega preverjanje je samo 3 %, kar je pokazatelj zelo uspešne klasifikacije. Na projekciji osebkov na prvi dve diskriminantni funkciji (Slika 12.8) vzdolž $DF1$ je vidno razločno razhajanje v morfoprostoru med skupinama *pumilio* (z eno samo podvrsto, *pumilio*) in skupino *potanini* (s podvrstmi *potanini*, *selevini*, *iliensis* in *bulatooides*). Skupina *pumilio* zaseda $DF1$ v razponu od -3,5 do 3,3; skupina *potanini* pa v razponu od 1,0 do 6,0, prekrivanje morfoprostora med skupinama je minimalno. Razdalje med centriidi skupin so predstavljene z Mahalanobisovimi razdaljami (Preglednica 12.4).



Slika 12.8 Projekcija osebkov na vrednosti prvih dveh diskriminantnih funkcij $DF1$ in $DF2$. Osebkov posamezne skupine so povezani z minimalnim konveksnim polinomom v skupni morfoprostoru. Centroidi posameznih skupin so označeni.

Preglednica 12.3 Klasifikacija osebkov iz dejanskih v napovedane skupine

| | ssp. | Napovedane skupine | | | | | Skupaj |
|------------------|--------------------|--------------------|-----------------|-----------------|-----------------|--------------------|--------|
| | | <i>pumilio</i> | <i>selevini</i> | <i>iliensis</i> | <i>potanini</i> | <i>bulatooides</i> | |
| Dejanske skupine | <i>pumilio</i> | 274 | 1 | 0 | 6 | 0 | 281 |
| | <i>selevini</i> | 0 | 8 | 0 | 0 | 0 | 8 |
| | <i>iliensis</i> | 0 | 1 | 12 | 1 | 0 | 14 |
| | <i>potanini</i> | 1 | 0 | 0 | 24 | 0 | 25 |
| | <i>bulatooides</i> | 1 | 0 | 0 | 0 | 12 | 13 |

Preglednica 12.4 Mahalanobisove razdalje med centriidi skupin

| ssp. | <i>bulatooides</i> | <i>iliensis</i> | <i>potanini</i> | <i>pumilio</i> | <i>selevini</i> |
|--------------------|--------------------|-----------------|-----------------|----------------|-----------------|
| <i>bulatooides</i> | 0 | | | | |
| <i>iliensis</i> | 2,904 | 0 | | | |
| <i>potanini</i> | 3,149 | 2,933 | 0 | | |
| <i>pumilio</i> | 5,223 | 4,816 | 4,011 | 0 | |
| <i>selevini</i> | 3,623 | 3,527 | 2,616 | 4,814 | 0 |



Zaključki

- Taksonomija je biološka disciplina o prepoznavanju (identifikaciji) in uvrščanju organizmov v sistem (klasifikaciji) na podlagi njihove sorodnosti in skupnih lastnosti.
- Numerična taksonomija je klasifikacija organizmov na podlagi njihovih merskih lastnosti z numeričnimi metodami.
- Razlike med skupinami prepoznavamo in vrednotimo z univariatnimi in multivariatnimi statističnimi analizami. Lastnosti objektov (osebkov ali skupin), ki jih zapišemo s številčnimi vrednostmi oblikujemo v matriko, kjer so objekti v vrsticah, preučevane lastnosti objektov pa v stolpcih.
- Posamezen objekt je opisan z več spremenljivkami; kadar obsega numerična obdelava istočasno analizo več spremenljivk, govorimo o multivariatni analizi.
- Redukcijo razsežnosti multivariatnega podatkovnega niza izvedemo z analizo glavnih komponent, v kateri skupno variabilnost množice spremenljivk strnemo v nekaj dimenzij, imenovanih glavne komponente.
- V diskriminantni analizi vnaprej (*a priori*) definiramo skupine; algoritem išče med spremenljivkami kombinacijo lastnosti, ki podpira apriorno delitev na skupine.
- Metoda voditeljev je multivariatna statistična metoda, s katero objekte cepimo v skupine. Optimiranje števila skupin preverjamo z metodo komolca, silhueto metoda in permutacijskim testom.

Reference

- Jesenko J., Jesenko M. 2007. Multivariatne statistične metode. Moderna organizacija. Kranj.
- Kassambara A., Mundt F. 2020. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7. <https://CRAN.R-project.org/package=factoextra>.
- Klingenberg C. P. 2011. MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources*, 11: 353–357.
- Kryštufek B., Janžekovič F., Shenbrot G., Ivajnsič D., Klenovšek T. 2019. Phenotypic plasticity under desert environment constraints: mandible variation in the dwarf fat-tailed jerboa, *Pygeretmus pumilio* (Rodentia: Dipodidae). *Canadian Journal of Zoology*, 97: 940–951.
- Kryštufek B., Shenbrot G., Klenovšek T., Janžekovič F. 2020. Geometric morphometrics of mandibular shape in the dwarf fat-tailed jerboa: relevancy for trinomial taxonomy. *Zoological Journal of the Linnean Society*, XX: 1–10.
- Legendre P., Legendre L. 2012. *Numerical Ecology*. Elsevier. Amsterdam.
- Sneath A.H.P., Sokal R.R. 1973. *Numerical Taxonomy. The Principles and Practice of Numerical Classification*. W.H. Freeman and Company. San Francisco.
- Sokal R.J., Rohlf F.J. 2012. *Biometry*. W.H. Freeman Company. New York.

13. poglavje

| | | | |
|---|---|--|---|
|  |  |  |  |
| 1 | 0 | 1 | 0 |
|  |  |  |  |
| | 0 | 0 | 1 |
|  |  |  |  |
| 1 | 1 | 0 | 0 |
|  |  |  |  |
| 1 | 0 | 1 | 1 |

FENETIKA IN KLADISTIKA

Povzetek

13. poglavje: Fenetika in kladistika

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

Filogeneza je evolucijski proces v katerem nastajajo nove vrste. Filogenija je hipoteza o najverjetnejšem poteku evolucijske zgodovine izbrane skupine organizmov. Prikažemo jo v obliki filogenetskega drevesa, ki ga zgradimo v postopku filogenetske rekonstrukcije. Razlikujemo tri metodološke koncepte, v katerih izvajamo filogenetske rekonstrukcije: filogenetski, kladistični in fenetski koncept. Filogenetika je metodološki pristop, v katerem filogenijo razlagamo z analizo verjetnosti v variabilnosti molekularnih znakov (v zaporedju nukleotidov in aminokislin). Kladistika je metodološki pristop filogenetskih rekonstrukcij, ki temelji na konceptu zadnjega skupnega prednika, oblikovanju monofiletskih skupin in iskanju skupnih izpeljanih znakov tako morfoloških kot molekularnih. Fenetika je metodološki pristop filogenetskih rekonstrukcij, ki temelji na statistični analizi variabilnosti v podobnosti in v velikosti morfoloških struktur. Vsi trije pristopi (filogenetika, kladistika in fenetika) imajo isti cilj, to je razumevanje in razlaga evolucijske zgodovine taksonov.

Fenetske odnose med taksoni analiziramo z metodami tradicionalne morfometrije in geometrijske morfometrije. Znaki so morfološke strukture od nivoja celičnih organelov, tkiv, organov, telesnih segmentov, do celotnega organizma. Morfološki znaki so kvalitativni (prisoten/odsoten ali v več stanjih) in kvantitativni (diskretni in zvezni). Podatke organiziramo v pravokotno matriko, kjer so taksoni v vrsticah in spremenljivke v stolpcih.

Morfološki znaki vsebujejo filogenetsko in fenetsko informacijo. Fenetska informacija je povezana s fenotipsko plastičnostjo v velikosti in oblike telesa kot odziv na biotske in abiotske dejavnike. Vir morfološke variabilnosti so tudi spolni dimorfizem, starostna variabilnost in alometrija. Dodatni vidiki variabilnosti zakrivajo filogenetsko informacijo v

morfoloških znakih. V fenetskem pristopu analiziramo podobnost med organizmi, iz podobnosti sklepamo na evolucijske odnose med njimi. Rezultat filogenetske rekonstrukcije prikažemo s fenetskim drevesom, ki ponazarja morfološko podobnost med predstavniki evolucijske linije. Fenetsko drevo gradimo z algoritmičnim pristopom, kjer drevo zgradimo na morfoloških razdaljah po metodi združevanja sosedov ali metodi UPGMA.

V kladističnem pristopu prepoznavamo naravne skupine organizmov, to so monofiletske skupine – imenovani kladi. Klad sestavljajo zadnji skupni prednik in vsi njegovi evolucijski potomci. Prepoznamo jih po tem, da si delijo skupno lastnost, ki izvira iz zadnjega skupnega prednika, pri sodobnih vrstah je lastnost lahko preoblikovana – sinapomorfni znak. Sinapomorfni morfološki znakom moramo potrditi homologijo. To pomeni, da morfološka lastnost izhaja iz skupnega zadnjega prednika, nastane iz istega zarodnega tkiva in ima enotno organizacijo. Oblika in funkcija homolognega znaka lahko variirata od zelo podobne do zelo različne. Npr. prednja ekstremiteta vseh placentalnih sesalcev izhaja iz istega zarodnega materiala in ima identično notranjo organizacijo, po obliki in funkciji pa se lahko med taksoni zelo razlikuje (npr. noga konja, plavut kita ali prhut netopirja).

Rezultat kladistične analize je kladogram. Z njim je vzpostavljena hierarhija monofiletskih skupin, hierarhično podrejene skupine so ugnezdene v višje skupine, za vsako skupino je definirana sinapomorfna lastnost. Kladogram gradimo po načelu varčnosti. Zgradimo ga z iskanjem optimalnega drevesa, to je z metodo največje varčnosti. Metoda zgradi vsa možna drevesa in na podlagi kriterija najmanjšega števila evolucijskih sprememb izberemo optimalno drevo.

V evolucijskem procesu se spreminja genetski zapis osebkov v populaciji. Te spremembe so dedne in se v populaciji prenašajo z zaporedjem generacij. Z genetskim zapisom so kodirane informacije za lastnosti strukturnih in regulatornih genov. Genetski zapis se spreminja v času, oziroma genetske lastnosti izkazujejo variabilnost tako znotraj populacije in še bolj med populacijami. Vir genetske variabilnosti so mutacije, rekombinacije in stohastični dogodki. Glavna vzroka za evolucijski nastanek nove vrste sta genetski drift in naravni izbor. Posledica njunega delovanja je sprememba v frekvenci alelov, določeni aleli v posameznih populacijah postajajo redkejši in se izgubijo in v nekaterih drugih populacijah postanejo pogostejši. Proces genetskega drifta in naravnega izbora vodi v kopičenje genetske razlike med populacijami. Ko populacija postane reproduktivno nekompatibilna z drugimi populacijami, nastane nova vrsta. Evolucijski pritisk, ki povzroča spremembo dednih lastnosti v zaporedju generacij, in s tem nastajanje novih vrst, je ključen za nastanek biodiverzitete (Evolution 2023).

Zgornji odstavek je zelo strnjen in poenostavljen zapis evolucijskega procesa in je vsebina evolucijskih in makroevolucijskih raziskav. Za vsebino te knjige je ključna informacija, da v evolucijskem procesu nastajajo nove vrste in s tem biodiverziteta. Iz makroekološkega vidika nas zanimajo evolucijski odnosi med vrstami in predvsem vrednotenje biodiverzitete iz perspektive filogenije. **Filogeneza** (ang. Phylogenesis) je proces, v katerem nastajajo nove vrste. In **filogenija** (ang. Phylogeny) je razlaga oziroma hipoteza o evolucijski zgodovini izbrane skupine organizmov, po navadi jo prikažemo v obliki **filogenetskega drevesa** (ang. Phylogenetic Tree) in postopek se imenujemo **filogenetska rekonstrukcija** (ang. Phylogenetic Reconstruction). Razlikujemo tri metodološke koncepte, to so filogenetski, kladistični in fenetski koncept, s katerimi pristopamo k razlagi filogenije in k filogenetskim rekonstrukcijam (Phylogenesis 2023, Phylogenetic tree 2003).

Filogenetika (ang. Phylogenetics) je metodološki pristop k filogenetskim rekonstrukcijam, ki variabilnosti molekularnih znakov razlaga z analizo verjetnosti. Filogenetski znaki so zaporedja nukleotidov v dednini ali zaporedja aminokislin v proteinih. Istočasno s terminom filogenetika poimenujemo vedo o evolucijski zgodovini organizmov in njihovih evolucijskih odnosih znotraj skupin in med njimi. Opomba: Priporočamo pozornost pri pravilni rabi termina filogenetika, saj ima dvojni pomen (kot metodološki pristop za molekularne znake in je veda o filogeniji), obenem je soroden terminoma filogeneza in filogenija, od katerih se vsebinsko razlikuje (Phylogenetics 2023).

Kladistika (ang. Cladistic) je metodološki pristop k filogenetskim rekonstrukcijam, ki temelji na konceptu zadnjega skupnega prednika, oblikovanju monofiletskih skupin in iskanju skupnih izpeljanih znakov tako morfoloških kot molekularnih (Cladistic 2023).

Fenetika (ang. Phenetics) je metodološki pristop k filogenetskim rekonstrukcijam, ki temelji na statistični analizi variabilnosti v obliki in v velikosti morfoloških struktur (Phenetics 2023).

Fenetika vs. kladistika. V fenetski filogeniji je merilo evolucijske sorodnosti morfološka podobnost oz. fenotip. Prispevek morfoloških znakov je izenačen. Filogenetski signal je praviloma zakrit zaradi drugih virov morfološke variabilnosti. Razlike med skupinami postavljamo s pomočjo statističnega testiranja variance. Kladistični pristop izhaja iz popolnoma druge perspektive, meje in razlikovanje med monofiletskimi skupinami postavljamo na podlagi sinapomorfnih znakov.

Vsi trije pristopi (filogenetika, kladistika in fenetika) imajo isti cilj, to je razumevanje in razlaga evolucijske zgodovine taksonov. Glavna razlika med koncepti je v znakih, ki jih vključuje, in njihovi metodološki obravnavi. Lastnosti posameznega pristopa so predstavljene v Preglednici 13.1. Koncepti se tudi ne izključujejo, v marsikaterih primerih se dopolnjujejo in prehajajo eden v drugega. Podrobnejši pristop k fenetiki in kladistiki je predstavljen v nadaljevanju tega poglavja, filogenetski koncept je predstavljen v Poglavju 14.

Preglednica 13.1 Pregled lastnosti treh konceptov v izdelavi filogenetskih rekonstrukcij: filogenetski, kladistični in fenetski koncept (ML – največje verjetje, BI – Bayesovo sklepanje, UPGMA – metoda združevanja neuteženih aritmetičnih sredin, NJ – združevanje sosedov, PA – največja varčnost, PCA – analiza glavnih komponent)

| Lastnost | Filogenetika | Kladistika | Fenetika |
|--------------------|---|---|--|
| Idejni pristop | Z modeliranjem substitucij nukleotidov in verjetnostjo razlagamo evolucijske spremembe v dednini. | Iščemo zadnjega skupnega prednika in oblikujemo monofiletske skupine. | Podobnost morfoloških struktur je merilo evolucijske sorodnosti. |
| Metode | - Iskanje substitucijskega modela. - Analiza verjetnosti. - Izračun genetskih razdalj. - ML, BI, UPGMA, NJ | - Prepoznavanje sinapomorfnih znakov in oblikovanje monofiletskih skupin. - Ugotavljanje homologij. - Načelo varčnosti. - PA, ML | - Analiza variance. - Redukcija razsežnosti s PCA. - Izračun morfoloških razdalj. - UPGMA, NJ |
| Znaki | - Zaporedja nukleotidov - Zaporedja aminokislin | - Molekularni znaki - Nominalni morfološki znaki | - Nominalni in intervalni morfološki znaki |
| Filogenetsko drevo | Filogram | Kladogram | Fenogram |
| Informativnost | Genetska podobnost med taksoni. | Hierarhija monofiletskih skupin. | Morfološka podobnost med taksoni. |

Poglavje je razdeljeno na dva sklopa, prvi sklop je namenjen predstavitvi fenetskega koncepta in filofenetike. V drugem sklopu so predstavljene vsebine kladističnega pristopa.

13.1 Fenetika

Filofenetika (ang. Phylophenetics) je pristop k filogenetskim rekonstrukcijam na morfoloških znakih. Za morfološke znake velja, da vsebujejo predvsem informacijo o organizaciji in obliki telesa. **Fenotip** (ang. Phenotype) je opredelitev velikosti, oblike in organizacije telesa, zato govorimo o fenetskih znakih. Fenotip se realizira v procesu ontogenetskega razvoja, dedni zapis odločilno vpliva na fenotip, ni pa edini dejavnik. Na razvoj fenotipa ob genotipu, vplivajo tudi abiotski in biotski dejavniki okolja, v katerem se organizem razvija, interakcija med genotipom in okoljskimi vplivi ter stohastični dogodki. Spremembe v obliki in velikosti organizmov, ki so posledica interakcije z biotskimi in abiotskimi dejavniki okolja in niso dedne, so **modifikacije** (ang. Modification) oziroma fenotipska plastičnost (glej Poglavje 11). Modifikacije ne prispevajo k filogenetski informaciji, so moteče in njihovo informacijo skušamo odstraniti iz filogenetske analize. Dodaten šum k filogenetskim rekonstrukcijam na morfoloških znakih vnaša spolni dimorfizem, starostna variabilnost in alometrija. Vsi navedeni vidiki variabilnosti zelo omejujejo uporabnost fenetske informacije v filogenetski rekonstrukciji. Imajo pa navedeni vidiki morfološke variabilnosti veliko uporabno vrednost v drugih kontekstih, predvsem v analizi odziva organizma na okolje in medvrstne odnose. V tem poglavju je poudarek na metodah in postopkih za filogenetske rekonstrukcije na morfoloških znakih, in to raziskovalno področje se imenuje filofenetika.

Filogenetska rekonstrukcija je hipoteza, s katero razlagamo evolucijski razvoj živih bitij. Danes privzemamo, da molekularni znaki (predvsem zaporedje nukleotidov v izbranem odseku DNK in zaporedje aminokislin v izbrani beljakovini) omogočajo vpogled v potek evolucije. Pristop in raziskovalno področje, v katerem filogenetske rekonstrukcije izvajamo na molekularnih znakih, je filogenetika in je predstavljena v Poglavju 14.

Filogenetsko rekonstrukcijo prikažemo z izrisom **filogenetskega drevesa** (sin. genealoško drevo, ang. Phylogenetic Tree). Drevo je matematični graf, v evolucijskem kontekstu je drevo sinonim za predstavitev evolucijskih odnosov med obravnavanimi taksoni. Topologija drevesa ponazarja potek nastajanja vrst, listi na filogenetskem drevesu so konkretni osebki. Veje predstavljajo evolucijsko linijo in smer pretoka genetske informacije. V razvejitvah se veje dihotojno cepijo na dva evolucijska potomca. Razvejitve predstavljajo evolucijski dogodek, to je nastanek nove vrste oz. speciacijo. Drevo v filofenetski analizi se imenuje fenogram in po navadi ni koreninjeno. Fenogram lahko koreninimo po metodi sredinske točke (podrobneje o filogenetskih drevesih in njihovem vrednotenju v Poglavju 14.5). Izvedba filogenetske rekonstrukcije poteka v dveh osnovnih korakih: 1. prvi korak je identifikacija in mapiranje homolognih znakov in 2. drugi korak je rekonstrukcija evolucijske zgodovine na teh homolognih znakih z metodami računske filogenetike in filogenetskega sklepanja ter njena predstavitev v obliki drevesa. Filofenetske rekonstrukcije lahko izvajamo na stanju morfoloških znakov od nivoja

celičnih organelov, tkiv, organov, organskih sistemov do celotnega organizma. Redkeje v filofenetske rekonstrukcije vključimo fiziološke znake, metabolne poti ali vedenjske vzorce.

Mapiranje morfoloških znakov zapisujemo v obliki pravokotne matrike, znaki so v stolpcih, OTU so v vrsticah in vrednosti znakov v celicah. Filogenetsko sklepanje pravilom izvajamo po algoritmičnem pristopu in temelji na razdaljah. V pristopu, temelječem na razdaljah, izračunamo razdaljo med dvema OTU. Razdalja je relativna vrednost, s katero numerično ovrednotimo različnost med dvema OTU (npr. osebkoma, vzorcema, populacijama, vrstama ...) na podlagi izbranih meritev. Matriko razdalj prikažemo v obliki drevesa. Vsak korak v navedeni proceduri ima svoje specifične in različne metodološke izpeljanke. Značilni postopki so predstavljeno s primeri.

Temeljna literatura iz katere smo črpali vsebino poglavja: Freeman in Herron (2004), Hall (2011), Legendre in Legendre (2012), Lemey in sod. (2009), Liem in sod. (2001), Sneath in Sokal (1973).

Izračuni in grafični izdelki so bili narejeni s programskimi orodji: MorphoJ (Klingenberg 2011), PAST (Hammer in sod. 2001) ter v R okolju (R Core Team 2022) in s paketi: "ade4" (Dray in sod. 2020), »ecodist« (Goslee in Urban 2007), »factoextra« (Kassambra in Mundt 2020) in "vegan" (Oksanen in sod. 2019).

Vsebina poglavja je razdeljena na:

- Fenetska analiza na kvalitativnih (nominalnih) morfoloških znakih.
- Fenetska analiza na kvantitativnih (intervalnih) morfoloških znakih.

13.1.1 Fenetska analiza na kvalitativnih morfoloških znakih

Skupini organizmov $OTU_1, OTU_2, \dots, OTU_m$ definiramo nabor morfoloških znakov Z_1, Z_2, \dots, Z_p z vrednostmi x_{mp} in oblikujemo matriko $m \times p$:

| | Z_1 | Z_2 | Z_3 | ... | Z_p |
|---------|----------|----------|----------|-----|----------|
| OTU_1 | x_{11} | x_{12} | x_{13} | ... | x_{1p} |
| OTU_2 | x_{21} | x_{22} | x_{23} | ... | x_{2p} |
| ... | ... | ... | ... | ... | ... |
| OTU_m | x_{m1} | x_{m2} | x_{m3} | ... | x_{mp} |

Za vsak OTU_i izmerimo vrednost znaka Z_i . Posamezni OTU je zastopan z enim osebkom. Spremenljivke z dihonomnim značajem zapisujemo z vrednostmi 1 (prisotno) ali 0 (odsotno). Meritve organiziramo v pravokotno matriko, kjer so OTU v vrsticah, znaki so v stolpcih, v celicah so vrednosti x_{ij} , ki so bodisi 0 bodisi 1 (glej Preglednico 13.2).

Posamezni programi oz. računski postopki zahtevajo obrnjeno razporeditev vrednosti, vrstice so znaki, stolpci so OTU. V primeru znaka s tremi ali več stanji imamo dve možnosti zapisa podatkov in izračuna cenilke. Prva možnost je, da naredimo nabor vseh možnih kategorij stanja (npr. 4 stanja), za posamezno kategorijo stanja oblikujemo svojo spremenljivko in jo vrednotimo po načelu prisotno odsotno (t.i. slepa spremenljivka, ang. Dummy Variable). Izračun cenilk izvedemo za binarne znake. Druga možnost je, da posameznemu stanju znaka določimo kodo stanja npr. s črkami (a, b, c, d) ali besednimi oznakami (glej Preglednico 13.4) in asociacije izračunamo z Gowerjevim splošnim koeficientom podobnosti.

Cenilke za binarne znake

Primerjamo dva OTU v zaporedju posameznih morfoloških znakov Z_1, Z_2, \dots, Z_p . Vsak par primerjanih OTU₁ in OTU₂ ima za posamezni znak Z_i eno izmed kombinacij:

$$\begin{aligned} Z_{i, 1, 1} &= a, \\ Z_{i, 1, 0} &= b, \\ Z_{i, 0, 1} &= c \text{ in} \\ Z_{i, 0, 0} &= d, \end{aligned}$$

ki so predstavljene v preglednici:

| | | OTU ₂ | | |
|------------------|---|------------------|-------|-------------------|
| | | Z _i | 1 | |
| OTU ₁ | 1 | a | b | a+b |
| | 0 | c | d | c+d |
| | | a + c | b + d | p = a + b + c + d |

Razdaljo na podlagi asociacijskega pristopa izračunamo s formulo:

$$d = \sqrt{1 - s} \text{ ali } d = 1 - s,$$

kjer je d – razdalja med dvema OTU, s – asociacijski koeficient.

V nadaljevanju so predstavljene cenilke za asociacijske koeficiente, vgrajene v paket »ade4« okolja R (Dray in sod. 2020), ki jih izvedemo z ukazom `dist.binary()`. Zaporedna številka pred posameznim koeficientom (od 1 do 10) se nanaša na izbor koeficienta asociacije v paketu »ade4«:

1 – **Jaccardov koeficient:**

$$s1 = \frac{a}{a + b + c}$$

2 – **Koeficient enostavnega ujemanja** (ang. Simple Matching Coefficient):

$$s2 = \frac{a + d}{a + b + c + d}$$

3 – **Sokal & Sneath S5 koeficient:**

$$s2 = \frac{a}{a + 2(b + c)}$$

4 – **Rogers & Tanimoto koeficient:**

$$s4 = \frac{a + d}{a + 2(b + c) + d}$$

5 – **Dice in Sorensen koeficient:**

$$s5 = \frac{2a}{2a + b + c}$$

6 – **Hamannov koeficient:**

$$s6 = \frac{a - (b + c) + d}{a + b + c + d}$$

7 – **Ochiai koeficient:**

$$s7 = \frac{a}{\sqrt{(a + b)(a + c)}}$$

8 – **Sokal & Sneath S13 koeficient:**

$$s8 = \frac{ad}{\sqrt{(a + b)(a + c)(d + b)(d + c)}}$$

9 – **Pearsonov Phi koeficient:**

$$s9 = \frac{ad - bc}{\sqrt{(a + b)(a + c)(d + b)(d + c)}}$$

10 – **Gower & Legendre S2 koeficient:**

$$s10 = \frac{a}{a + b + c + d}$$

Oblikovanje matrike znakov in izračun koeficienta razdalje na binarnih znakih je predstavljen v Primeru 13.1.



PRIMER 13.1

Filogenetski odnosi med kraljestvi – morfološki binarnih znaki

Organizmi/celice. V filogenetsko rekonstrukcijo vključimo predstavnike posameznega kraljestva. 1. Vzorec kraljestva živali so celice ustne sluznice človeka. 2. Predstavnik protistov (Alveolata) je paramecij. 3. Celica luskolista čebule je predstavnica rastlin. 4. Povrhnjice kukmaka je predstavnica gliv. 5. Predstavnica bakterij Cianobakterije so modrozeleni cepljivke z drevesne skorje.

Znaki. Za vsak organizem je podano stanje znakov, prisotnost znaka je označena z ena 1. Če organizem znaka nima, je označeno z nič 0: organizem je enocelični; celica ima celično jedro; celica je obdana s celulozno celično steno; celica je obdana s hitinsko celično steno; celica je obdana s polisaharidno-aminokislinsko celično steno; v celici so prisotni plastidi (Preglednica 13.2).

Preglednica 13.2 Mapiranje znakov, stanje šestih znakov za predstavnike iz petih različnih kraljestev

| Kraljestvo | Predstavnik | Enocelični | Celično jedro | Endoplazmatski retikulum | Celična stena – celuloza | Celična stena – hitin | celična stena – polisaharid aminokislina | Plastidi |
|------------|----------------|------------|---------------|--------------------------|--------------------------|-----------------------|--|----------|
| Animalia | človek | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| Protista | paramecij | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| Plantae | čebula | 0 | 1 | 1 | 1 | 1 | 0 | 0 |
| Fungi | kukmak | 0 | 1 | 1 | 1 | 0 | 1 | 0 |
| Bakteria | cianobakterije | 1 | 0 | 0 | 1 | 0 | 0 | 1 |

Izračun podobnosti med organizmi

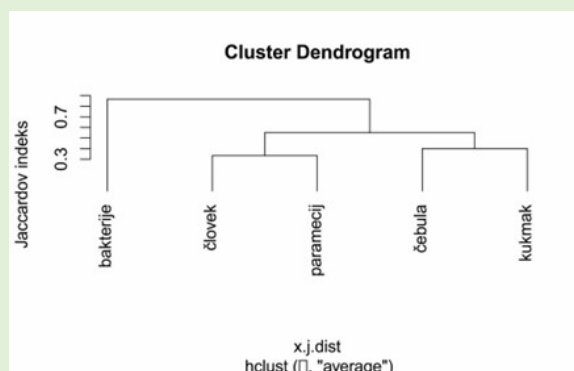
V primeru binarnega stanja znakov prisoten/odsoten (ang. PA Data – Present / Absent) izračun podobnosti med organizmi izvedemo z Jaccardovim koeficientom. V biologiji in ekologiji je Jaccardov koeficient pogosto uporabljena mera za vrednotenje podobnosti med pari primerjanih objektov.

Podobnost med organizmi je predstavljena z Jaccardovim koeficientom (Preglednica 13.3). Gre za asimetrični koeficient, v izračun katerega dvojne ničle niso vključene. Vrednosti so podane kot razdalje $d = 1 - s$. Najkrajša razdalja je 0,577. Izkazuje podobnost

para človek-paramecij, kar pomeni, da si ta dva OTU delita največ skupnih lastnosti. Nasproten primer (najdaljša razdalja) je med OTU človek-bakterija, ki imata najvišjo možno razdaljo 1. To pomeni, da si ne delita niti ene skupne lastnosti. Razmerja med OTU so predstavljena z drevesom (Slika 13.1). OTU so v dveh skupinah, eno oblikujejo prokarionti, predstavnik je cianobakterija in drugo evkarionti, v katero so vključeni preostali organizmi. Znotraj skupine Eukarionta sta vzpostavljeni dve skupini, »živali sensu lato« (človek in paramecij) in rastline (čebula in kukmak).

Preglednica 13.3 Razdalje med posameznimi OTU na podlagi Jaccardovega koeficienta

| | človek | paramecij | čebula | kukmak | bakterija |
|-----------|--------|-----------|--------|--------|-----------|
| človek | 0 | | | | |
| paramecij | 0,5773 | 0 | | | |
| čebula | 0,7071 | 0,7745 | 0 | | |
| kukmak | 0,7071 | 0,7745 | 0,6324 | 0 | |
| bakterija | 1,0000 | 0,8944 | 0,9128 | 0,9128 | 0 |



Slika 13.1 UPGMA drevo odnosov med petimi OTU. Vhodni podatki za izračun drevesa so zbrani v simetrični matriki razdalj v Preglednici 13.3. Vrednost koeficienta kofenetske korelacije je 0,966.

```
R > package = »vegan«, »ade4«
# Jaccardov koeficient + Hierarhični klastering, data = kralj
d.kralj<-dist.binary(kralj, method=1) #Izračun koeficienta
d.kralj #Izpis vrednosti
d.kralj.clust <- hclust(d.kralj, method = "average") # Izračun drevesa
plot(d.kralj.clust, hang = -1, axes = TRUE, ylab = "Jaccardov indeks") # Izpis drevesa
coph<-cophenetic(d.kralj.clust) #Kofenetska korelacija
cor(d.kralj, coph) #Korelacija
```



Cenilka za mešan tip spremenljivk

Primerjamo dva OTU, ki ju opišemo z naborom morfoloških znakov Z_1, Z_2, \dots, Z_p . Znak Z_j je spremenljivka, katere vrednost je lahko intervalna, ordinalna, razmernostna ali nominalna. Za sočasno obravnavo znakov različnih tipov uporabimo Gowerjev splošni koeficient podobnosti. Podobnost med OTU₁ in OTU₂ je povprečje podobnosti vseh znakov Z_p . Za posamezen znak j se parcialna vrednost podobnosti s_{12j} med dvema OTU za vrednosti x_{1j} in x_{2j} izračuna, kot sledi:

- Za vsak posamezen *binarni znak* velja, da je $s_j = 1$, če sta stanji identični; ali $s_j = 0$, če sta stanji različni. Gower je predvidel dve obliki koeficienta, simetričnega $s_j = 1$, tudi v primeru dvojne ničle; oz. asimetričnega $s_j = 0$ v primeru dvojne ničle. Veljajo odnosi:
 - $w_{12j} = 1$, če sta x_1 in x_2 znana; če kateri od znakov manjka, je $w_{12j} = 0$;
 - $s_{12j} = 0$, če je $x_{1j} \neq x_{2j}$;
 - $s_{12j} = 1$, če $x_{1j} = x_{2j}$ ali $x_{1j} = x_{2j} = 0$, dvojne ničle (obojestranska odsotnost) so vključene v primeru simetričnega koeficienta;
 - $s_{12j} = 0$, če $x_{1j} = x_{2j} = 0$, dvojne ničle so izključene iz izračuna v primeru asimetričnega koeficienta.
- Za posamezen *kvalitativni znak*, ko sta stanji skladni, $s_j = 1$ oz. $s_j = 0$, če stanji nista skladni. Dvojne ničle se obravnavajo po simetričnem načelu. Veljajo odnosi:
 - $w_{12j} = 1$, če sta x_1 in x_2 znana, če kateri od znakov manjka, je $w_{12j} = 0$;
 - $s_{12j} = 0$, če $x_{1j} \neq x_{2j}$;
 - $s_{12j} = 1$, če $x_{1j} = x_{2j}$.
- Za posamezni *semikvantitativni znak* glede na tip znaka uporabimo:
 - »klasični« pristop; če so vrednosti kvantitativne, izračunamo po formuli, navedeni pri kvantitativnem znaku;
 - »metrični« pristop, vrednosti rangiramo kot 1., 2., 3. itd. in uporabimo enačbo iz kvantitativnega pristopa;
- Za *kvantitativni znak* (zvezni ali diskretni); izračunamo absolutno razliko med dvema vrednostma $|x_{1j} - x_{2j}|$. Dobljeno vrednost delimo z razponom R_j ($R = \max - \min$). Podobnost med objektoma v znaku s_{12j} izračunamo po formuli:

$$s_{12j} = 1 - \frac{|x_{1j} - x_{2j}|}{R_j}.$$

Gowerjev splošni koeficient podobnosti S_{15} , pri primerjanju OTU₁ in OTU₂ ima obliko:

$$S_{15(1,2)} = \frac{\sum_{j=1}^p w_{12j} s_{12j}}{\sum_{j=1}^p w_{12j}}$$

z zalogo vrednosti od 0 do 1 (maksimalna podobnost).

Vrednost w_{12j} je tehtani prispevek posameznega znaka in ima vrednost 0 ali 1. Vrednost 0 ne prispeva k izračunu podobnosti in vrednost 1 prispeva bistveno. Vrednost w_{12j} uporabljamo pri vključevanju OTU z manjkajočimi (neznanimi) posameznimi vrednostmi.

Primer izračuna Gowerjevega splošnega koeficienta podobnosti je predstavljen v Primeru 13.2 na vsebini o filogenetskih odnosih med mišmi (družina Muridae).



PRIMER 13.2

Filogenetski odnosi med mišmi Muridae – mešani znaki

V analizi podatkov mešanega tipa (kvantitativni (intervalni), semikvantitativni (rangiranja), kvalitativni in binarni podatki (nominalni)) izračun podobnosti izvedemo z Gowerjevim splošnim koeficientom podobnosti.

Preglednica 13.4 Mešani tip podatkov za 7 vrst miši (koda vrste je zgrajena iz prvih treh črk rodovnega in treh črk vrstnega imena, polna imena vrst so navedena v besedilu Primera 13.4)

| Vrsta | TGL | CbL | Korenine m1 | Korenine M1 | Nadočasna letev | Barva kožuha | Areal |
|--------|-----|------|----------------|----------------|--------------------|-----------------|---------------|
| AOSYL | 94 | 22,8 | 2 | 4 | Ne | rjav | Evroazijski |
| AOFLA | 112 | 25,3 | 2 | 4 | Ne | rjav | Evropski |
| AOAGR | 102 | 23,5 | 2 | 4 | Da | rjav | Evroazijski |
| MUSMUS | 81 | 19,3 | 2 | 3 | Ne | siv | Palearktični |
| MICMIN | 69 | 17,1 | 3 | 5 | Ne | rjav | Evroazijski |
| RATRAT | 186 | 40,3 | 4 | 5 | Da | črn | Kozmopolitski |
| RATNOR | 241 | 46,8 | 4 | 5 | Da | siv | Kozmopolitski |

Preglednica 13.5 Postopek izračuna Gowerjevega splošnega koeficienta podobnosti pri primerjanju dveh vrst belonogih miši: AOSYL-AOFLA

| Tip podatka | Kvanti. zvezni | Kvanti. zvezni | Kvanti. diskretni | Kvanti. diskretni | Binarni | Kvalitativni | Kvalitativni | |
|----------------|-------------------|-------------------|----------------------|----------------------|------------------|-----------------|--------------|------------------|
| Vrsta | TGL | CbL | Korenine m1 | Korenine M1 | Nadoče. letev | Barva kožuha | Areal | |
| AOSYL | 94 | 22,8 | 2 | 4 | 0 | rjav | Evroazijski | |
| AOFLA | 112 | 25,3 | 2 | 4 | 0 | rjav | Evropski | |
| w_{12j} | 1 | 1 | 1 | 1 | 1 | 1 | 1 | N=7 |
| R_{12j} | 172 | 29,7 | 2 | 2 | | | | |
| $ y_{12j} $ | 18 | 2,5 | 0 | 0 | | | | |
| $ y_{12j} /R$ | 0,105 | 0,084 | 0 | 0 | | | | |
| S_{12j} | 0,895 | 0,916 | 1 | 1 | 1 | 1 | 0 | Sum = 5,811 |
| Gower = Sum/N; | | | | | | | | $S_{15} = 0,830$ |

Postopek vrednotenja podobnosti predstavljamo s primerjavo med sedmimi vrstami miši. V analizo smo vključili: - kvantitativne variable, tako zvezne (dolžina telesa-TGL in dolžina lobanje-CbL) kot diskretne (število korenin na spodnjem m1 in število korenin na zgornjem prvem meljaku M1); - kvalitativne (barva kožuha in areal) ter - binarne (prisotnost odsotnost nadočesne letve). Podatki so predstavljeni v Preglednici 13.4. Primer izračuna splošnega Gowerjevega koeficienta podobnosti za vrsti *Apodemus sylvaticus* in *Apodemus flavicollis* je predstavljen v Preglednici 13.5 in ima vrednost $S_{15} = 0,83$.

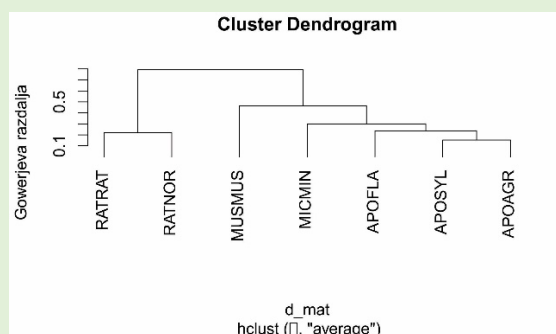
Gowerjev splošni koeficient podobnosti G ima zalogo vrednosti od 0 do 1, pri čemer 1 pomeni popolno podobnost med OTU. Najbolj podobni sta vrsti *Apodemus sylvaticus* in *Apodemus agrarius* z vrednostjo Gowerjevega koeficienta $G = 0,847$. Najmanj podobni sta vrsti *Apodemus sylvaticus* in *Rattus norvegicus*, $G = 0,197$ (Preglednica 13.6).

Za predstavitev odnosov z drevesom smo koeficient podobnosti preoblikovali v Gowerjevo razdaljo v postopku: $d = 1 - G$.

Klastrska analiza razporedi vrste v dve skupini. Prvo skupino sestavljata obe vrsti podgan *Rattus rattus* in *Rattus norvegicus*, v drugi skupini so predstavniki treh rodov miši, in sicer rod *Mus* z vrsto *M. musculus*, rod *Micromys* z vrsto *M. minutus* in rod *Apodemus* s tremi vrstami *A. flavicollis*, *A. sylvaticus* in *A. agrarius* (Slika 13.2).

Preglednica 13.6 Vrednosti Gowerjevega splošnega koeficienta podobnosti med različnimi vrstami iz družine miši Muridae

| Vrsta | APOSYL | APOFLA | APOAGR | MUSMUS | MICMIN | RATRAT | RATNOR |
|--------|--------|--------|--------|--------|--------|--------|--------|
| APOSYL | 1 | | | | | | |
| APOFLA | 0,830 | 1 | | | | | |
| APOAGR | 0,847 | 0,697 | 1 | | | | |
| MUSMUS | 0,615 | 0,588 | 0,462 | 1 | | | |
| MICMIN | 0,809 | 0,639 | 0,656 | 0,479 | 1 | | |
| RATRAT | 0,197 | 0,224 | 0,349 | 0,097 | 0,291 | 1 | |
| RATNOR | 0,120 | 0,147 | 0,272 | 0,163 | 0,214 | 0,780 | 1 |



Slika 13.2 UPGMA fenetsko drevo odnosov med različnimi vrstami iz družine miši Muridae. Vhodni podatek za izračun drevesa je simetrična matrika Gowerjevih razdalj. Vrednost koeficienta kofenetske korelacije je 0,962.

```
R > package = »vegan«
# Izris dendrogama za mešane podatke, data = muroidea
d_Gm<-1-muroidea #Preračun podobnosti v razdaljo
d_mat<-as.dist(d_Gm) #Pretvorba v distančno matriko
hc_mat<-hclust(d_mat, method = "average") #izračun dendrograma
plot(hc_muroidea, hang = -1, axes = TRUE, ylab="Gowerjeva razdalja") #Izris
cop<-cophenetic(hc_muroidea) #Izračun kofenetske korelacije
cor(d_mat, cop) #Izpis
```



13.1.2 Fenetska analiza na kvantitativnih morfoloških znakih

V filogenetskih raziskavah primerjamo odnose med organizmi $OTU_1, OTU_2, \dots, OTU_{mp}$; znaki $Z_1, Z_2, Z_3, \dots, Z_p$ so kvantitativne morfološke spremenljivke z intervalnimi vrednostmi x_{ij} , ki so bodisi diskretne bodisi zvezne s porazdelitvijo blizu normalni. V naslednjih dveh poglavjih 13.1.3 in 13.1.4 je podrobneje predstavljena analiza na dveh vrstah zveznih spremenljivk, ki so lahko:

- enorazsežne vrednosti morfoloških struktur, npr. dolžine oz. razdalje – analizo enorazsežnih spremenljivk imenujemo tradicionalna morfometrija;
- dvo- ali trirazsežne vrednosti morfoloških struktur npr. koordinate (X, Y, Z) – ta pristop imenujemo geometrijska morfometrija.

Cenilke za kvantitativne znake

Za oba tipa spremenljivk velja, da razdalje med OTU računamo na povprečjih. V izračun lahko vključimo tudi varianco (oz. standardno deviacijo). Izračune razdalj na kvantitativnih znakih med OTU izvajamo na standardiziranih vrednostih. Razdalje med paroma OTU, ki temelji na povprečjih, so Evklidova, Bray-Curtisova in Manhattanova razdalja.

Evklidova razdalja, izračunamo jo s formulo:

$$E_{ij} = \sqrt{\sum_k (x_{ki} - x_{kj})^2},$$

kjer so: E_{ij} – Evklidova razdalja med OTU_i in OTU_j ; x_{ki} – vrednost znaka k za OTU_i ; x_{kj} – vrednost znaka k za OTU_j .

Bray-Curtisova razdalja, izračunamo jo s formulo:

$$d_{ij} = \frac{\sum_k |x_{ki} - x_{kj}|}{\sum_k (x_{ki} + x_{kj})},$$

kjer so: d_{ij} – Bray-Curtisova razdalja med OTU_{*i*} in OTU_{*j*}; x_{ki} – vrednost znaka *k* za OTU_{*i*}; x_{kj} – vrednost znaka *k* za OTU_{*j*}.

Manhattanova razdalja (sin. City block razdalja), izračunamo jo s formulo:

$$M_{ij} = \frac{\sum_k |x_{ki} - x_{kj}|}{p},$$

kjer so: M_{ij} – Manhattanova razdalja med OTU_{*i*} in OTU_{*j*}; x_{ki} – vrednost znaka *k* za OTU_{*i*}; x_{kj} – vrednost znaka *k* za OTU_{*j*}; *p* – število znakov.

Razdalje, pri katerih upoštevamo tudi varianco, izračunavamo v multivariatni analizi, npr. diskriminantni analizi. Posamezni OTU je zastopan z vzorcem, iz katerega izračunamo razdaljo med dvema centroidoma, npr. Mahalanobisovo razdaljo in njeno izpeljanko Prokrustovo razdaljo.

Mahalanobisova razdalja je multidimenzionalna posplošitev razdalj med točkami *P* in centroidom *D*. Podana je v standardnih deviacijah. Kadar Mahalanobisove razdalje reduciramo v Evklidovo razdaljo, dobimo standardizirano Evklidovo razdaljo. Izračunamo jo s formulo:

$$d_{x,y} = \sqrt{\sum_{i=1}^N \frac{(x_i - y_i)^2}{s_i^2}},$$

kjer so: $d_{x,y}$ – Mahalanobisova razdalja med OTU *X* in *Y*; za znak *i*; s_i^2 – varianca znaka *i*.

Prokrustova razdalja je modifikacija Mahalanobisove razdalje, v katero namesto enorazsežnih vrednosti vključimo vrednosti koordinat (*X*, *Y*) kot dvorazsežne vrednosti.

Standardizacija

V morfometrični analizi se srečujemo z zveznimi spremenljivkami, med katerimi je absolutna razlika izrazito velika; npr. dolžina roke proti dolžini prsta. Variabilnost spremenljivke z absolutno večjimi vrednostmi »zasenči« variabilnost spremenljivke z absolutno manjšimi vrednostmi. Absolutnim vrednostim meritev pravimo surove vrednosti. Pred vključitvijo spremenljivk v multivariatno analizo izenačimo njihovo

variabilnost tako, da njihove vrednosti podamo v enotah standardne deviacije. Postopek se imenuje **z-standardizacija**. Standardizacija je oblika transformacija podatkov, pri kateri izločimo absolutne razlike med vrednostmi, ohranimo pa relativna razmerja med njimi. Z – transformirani podatki imajo nabor vrednosti praviloma od –3 do +3 s povprečjem 0 (nič). Standardizacijo izvedemo s formulo:

$$z = \frac{x_i - \bar{x}_i}{Std_i},$$

kjer so: z – standardizirana vrednost; x_i – posamezna meritev spremenljivke X_i ; \bar{x}_i – aritmetična sredina spremenljivke X_i ; Std_i – standardna deviacija spremenljivke X_i .

Alometrija

V obdobju rasti organizma se spreminjajo njegova velikost in oblika, pa tudi razmerja med njima. Posameznimi segmenti telesa imajo različne stopnje rasti, kar povzroči spremembo oblike. Razmerje v stopnji rasti med posameznimi deli telesa imenujemo alometrično razmerje oz. alometrija. Ko se opazovana struktura relativno poveča, govorimo o pozitivni alometriji. Če se relativno manjša, govorimo o negativni alometriji (podrobneje v Poglavju 11.4).

Vpliv statične in ontogenetske alometrije, ki moti filogenetsko rekonstrukcijo, lahko odstranimo iz analize s pomočjo regresije. To izvedemo z regresijo osnovnih vrednosti meritev na komponento velikosti. Ostanki regresije so velikosti prosti.

Pristopi analiziranja posameznih tipov podatkov (tradicionalna morfometrija, geometrijska morfometrija in na nukleotidnih sekvencah) so narejeni na modelni skupini belonogih miši rod *Apodemus*, pristop je podrobneje predstavljen v Projektu 13.1.

13.1.3 Tradicionalna morfometrija in izločitev velikosti

Variabilnost enorazsežnih morfometričnih spremenljivk vsebuje informacijo o velikosti strukture in o njeni obliki. Nabor numeričnih in statističnih metod omogoča razstavitev informacije na velikostno in oblikovno komponento. Tako lahko ločeno obravnavamo obliko in velikost organizmov. Velikost je izrazito fenotipsko plastična, zato variabilnost v velikosti pogosto odraža odziv organizma na lokalne okoljske razmere. V filogenetskih rekonstrukcijah je fenotipska variabilnost v velikosti moteča, saj je povezana z variabilnostjo okolja in praviloma nima filogenetske informacije. Prvi cilj filogenetskih rekonstrukcij na morfoloških znakih je izločitev virov variabilnosti, ki nimajo filogenetske vrednosti. Najprej izločimo variabilnost v velikosti, ki predstavlja odziv na dejavnike okolja

in zakriva filogenetski signal. V naslednjem koraku izločimo tudi spolno in starostno variabilnost, če sta prisotni.

Analiza glavnih komponent PCA je ključnega pomena za razstavitev morfometrične informacije iz podatkovnega niza na informacijo o velikosti in informacijo o obliki organizma. Rezultat PCA so glavne komponente PC, ki hierarhično razlagajo deleže variabilnosti podatkovnega niza. Prva glavna komponenta PC1 razloži najvišji delež variabilnosti, druga glavna komponenta PC2 razloži naslednji najvišji delež variabilnosti, ki še ni bil razložen s PC1, PC3 razloži preostali najvišji delež variabilnosti, ki ga nista razložili PC1 in PC2, in tako naprej do zadnje PC. Glavne komponente v multidimenzionalnem prostoru ležijo pravokotno ena na drugo, zato razlagajo neodvisne dimenzije variabilnosti. Vse glavne komponente skupaj razložijo celotno variabilnost v podatkovnem nizu. Število glavnih komponent je praviloma enako številu osnovnih spremenljivk. Manjše je v primeru, če je število meritev manjše od števila spremenljivk, potem je izenačeno s številom meritev. Raziskave na živalih izkazujejo, da je najvišji delež njihove variabilnosti v telesni velikosti. Npr. višji človek ima po navadi tudi daljše okončine in večjo telesno maso. Telesna velikost je tako najbolj variabilna metrična lastnost živali, zato PC1 razlaga najvišji delež skupne variabilnosti. Iz teh dveh lastnosti je izpeljan sklep, da prva glavna komponenta PC1 predvsem razlaga variabilnost organizmov v njihovi telesni velikosti. V morfometričnih analizah živali je sprejeta posplošitev, da je PC1 komponenta telesne velikosti, preostale PC pa so nosilke velikosti proste (oblikovne) informacije o organizmu. Da PC1 dejansko povzema velikost organizma, preverimo z izpolnjevanjem naslednjih kriterijev: 1. uteži glavnih komponent (ang. Loadings) med osnovnimi spremenljivkami in PC1 morajo imeti isti predznak (+ ali -); 2. uteži morajo biti visoke ($> 0,5$) in 3. uteži se ne smejo preveč razlikovati v absolutnih vrednostih.

Po razstavitvi informacije na velikost (PC1) in na obliko (PC2, ..., PCn) lahko obe komponenti telesne variabilnosti obravnavamo ločeno. V filogenetskih analizah nas zanima informacija o obliki, saj domnevamo, da je oblika nosilka informacije o evoluciji organizma. Razviti sta dve metodi za odstranitev vpliva velikosti iz enorazsežnih meritev in ločeno analizo oblike: **izločitev PC1** in **Burnabyjeva metoda**. Pri metodi izločitve PC1 v nadaljnje filogenetske analize vključimo glavne komponente od vključno 2. naprej (PC2, ..., PCn). S tem, ko ne vključimo PC1, smo izločili informacijo o velikosti organizma. Burnabyjeva metoda izločitve velikosti temelji na ortogonalni (pravokotni) regresiji osnovnih podatkov na PC1 kot komponento velikosti. V primeru izometričnega odnosa med spremenljivkami bi posamezne meritve zavzele položaj na regresijski premici, in tako bi velikost razložila vso variabilnost v obliki. Dejanska opažanja odstopajo od regresijske premice in so posledica neodvisne variabilnosti oblike od velikosti. Z regresijsko metodo iz osnovnih podatkov odstranimo variabilnost, ki je povezana s PC1 (oz. z velikostjo). Rezultat regresije so ostanki, ki ohranijo informacijo, ki ni povezana z velikostjo. Te ostanke interpretiramo kot proste velikosti, tako da povzemajo obliko organizma.

Prva glavna komponenta PC1 je torej komponenta velikosti in je ključnega pomena v ekomorfoloških analizah, kjer omogoča obravnavo odziva telesne velikosti v gradientu ekoloških dejavnikov (glej Poglavlje 11 o Fenotipski plastičnosti).

Pristop s tradicionalno morfometrijo je predstavljen v Primerih 13.3 in 13.4.



PROJEKT 13.1

Filogenetski odnosi v rodu *Apodemus*

Rod belonogih miši *Apodemus* vključuje majhne do srednje velike miši z dolžino trupa od 65 do 120 mm (Slika 13.3). Glava je podolgovata z dolgim gobčkom, brez posebnih modifikacij. Rep je razmeroma dolg od 66 do 122 mm, zmerno dlakav, na koncu pogosto s čopičem dlak. Kožuh je običajno gost in mehak, načeloma temno rjave do sive barve, z rumenorjavimi do belimi lisami. Uhlji so razmeroma veliki, merijo od 9 do 16 mm. Ker so prilagojene za skakanje, imajo razmeroma dolg zadnji par nog. Stopala in dlani so gola, razen v bližini pet. Na sprednjih nogah imajo po štiri prste, na zadnjih pet. Prvi zgornji meljak ima štiri korenine, prvi spodnji pa dve. Naseljujejo gozdnate habitate in tudi grmišča in polja.

V obravnavo smo vključili šest evropskih vrst rodu *Apodemus*: *A. agrarius*, *A. epimelas*, *A. alpicola*, *A. sylvaticus*, *A. flavicollis* in *A. uralensis*. Filogenetske rekonstrukcije smo izvedli s tremi metodološkimi pristopi: tradicionalno morfometrijo, geometrijsko morfometrijo in analizo nukleotidnih zaporedij.

Tradicionalna morfometrija se nanaša na uni- in multi–variatno analizo enorazsežnih meritev. Na Sliki 13.4 so predstavljene meritve posameznih znakov na lobanji miši. Meritve praviloma izvajamo na objektih (lobanjah) s kljunastim merilom in vrednosti zapisujemo v milimetrih. Informacija se v glavnem nanaša na velikost lobanje, oblikovno informacijo izluščimo iz podatkovnega niza.

Geometrijska morfometrija se nanaša na uni- in multi–variatno analizo dvorazsežnih 2D meritev (lahko tudi 3D). Na Sliki 13.8 so predstavljene oslonilne točke posameznih znakov na lobanji miši. Meritve praviloma izvajamo na fotografijah objektov s posebnimi programi, npr. tpsDig in vrednosti zapisujemo kot koordinate (X, Y). Informacija se nanaša na velikost in obliko lobanje. V Prokrustovi analizi meritve vseh objektov poravnamo v isto koordinatno izhodišče in informacijo razstavimo na velikost (centroidna velikost) in na obliko (Prokrustove koordinate). V nadaljevanju velikost in obliko analiziramo ločeno.

Tradicionalna in geometrijska morfometrija imata različen koncept zbiranja meritev. Pri analizi uporabljamo identična statistična orodja. Na univariatnem nivoju razlike med skupinami preverjamo s *t*-testom in ANOVA *F*-testom in povezave med spremenljivkami s korelacijsko analizo. V multivariatni obravnavi najpogosteje uporabljamo: analizo glavnih komponent, diskriminantno analizo in klastersko analizo in odnose med spremenljivkami analiziramo z regresijsko analizo.

Analiza nukleotidnih zaporedij se nanaša na določen gen. Pri glodalcih je to pogosto mitohondrijski gen dihalne verige, imenovan citokrom b. V tem delu zbiranja primarnih podatkov ne predstavljamo. Nukleotidna zaporedja gena za citokrom b za vrste rodu *Apodemus* smo dobili iz genske banke.



Slika 13.3 Dimasta miš *Apodemus agrarius*.

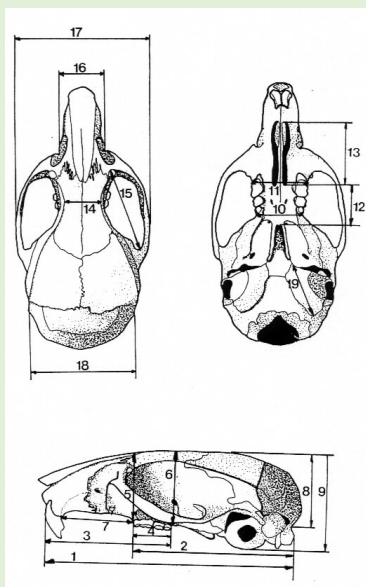
Foto. Marijan Govedič.



PRIMER 13.3

Filogenetski odnosi v rodu *Apodemus* –
tradicionalna morfometrija

Naš namen je preverjanje podobnosti, in s tem domnevne sorodnosti med evropskimi vrstami belonogih miši rodu *Apodemus* na osnovi enorazsežnih lobanjskih znakov. Tradicionalna morfometrična analiza lobanje je standardni postopek v raziskavah sesalcev. Izmerili smo 13 lobanjskih spremenljivk (Slika 13.4) pri šestih vrstah rodu *Apodemus*. Njihove povprečne vrednosti so navedene v Preglednici 13.7. Najvišjo vrednost ima spremenljivka VL1 (npr. pri vrsti *epimelas* 28,1 mm) in najmanjšo spremenljivka VL4 (*uralensis* 3,5 mm). V absolutnem merilu je VL1 osem krat večja od VL4. V multivariatni analizi to pomeni bistveno večjo utež, kar privede do rezultatov, ki lahko popačijo dejanska razmerja med vrstami. Zato spremenljivke standardiziramo, njihove vrednosti izrazimo kot standardne deviacije (Preglednica 13.8). Standardizirane \bar{x} vrednosti vključimo v izračun Evklidovih razdalj med vrstami.



Slika 13.4 Lobanja belonogih miši *Apodemus* z označenimi spremenljivkami.

* VL1 – kondilobazalna dolžina, VL2 – dolžina od zatilničnega čvrša do prednjega roba prvega meljaka M1, VL3 – dolžina od sekalca do tretjega meljaka IM3, VL4 – dolžina zgornjega niza meljakov, VL5 – višina rostruma, VL6 – višina lobanje, VL7 – dolžina diasteme, VL8 – višina možganskega dela lobanje, VL9 – višina možganskega dela lobanje z bobničnim mehurjem, VL10 – razdalja med levim in desnim meljakom M3, VL11 – razdalja med levim in desnim meljakom M1, VL12 – razdalja od nebne špranje do konca trdega neba, VL13 – dolžina nebne špranje, VL14 – interorbitalna širina, VL15 – širina očnice, VL16 – širina rostruma, VL17 – širina ličnih lokov, VL18 – širina lobanje, VL19 – dolžina bobničnega mehurja.

* *Apodemus epimelas* – APOEPI, *Apodemus flavicollis* – APOFLA, *Apodemus sylvaticus* – APOSYL, *Apodemus uralensis* – APOURA, *Apodemus alpicola* – APOALP, *Apodemus agrarius* – APOAGR

Preglednica 13.7 Aritmetična sredina (mm) 13 morfometričnih znakov lobanje pri petih vrstah rodu *Apodemus**

| | VL1 | VL2 | VL3 | VL4 | VL7 | VL8 | VL9 | VL13 | VL14 | VL16 | VL17 | VL18 | VL19 |
|--------|------|------|------|-----|-----|-----|-----|------|------|------|------|------|------|
| APOEPI | 28,1 | 18,2 | 14,8 | 5 | 8,1 | 9,6 | 11 | 6,9 | 4,8 | 5,3 | 15,3 | 12,5 | 5,2 |
| APOAGR | 23,6 | 15,1 | 12,4 | 4,1 | 7 | 7,7 | 9,1 | 4,9 | 4,4 | 5,3 | 12,7 | 10,9 | 4,6 |
| APOFLA | 25,2 | 16,2 | 13,3 | 4,3 | 7,5 | 8,5 | 9,9 | 5,5 | 4,4 | 5,2 | 13,9 | 11,6 | 5,2 |
| APOSYL | 23 | 14,9 | 12 | 3,8 | 6,9 | 8 | 9,3 | 5,5 | 4,1 | 4,8 | 12,7 | 11,1 | 4,7 |
| APOURA | 20,9 | 13,4 | 10,9 | 3,5 | 6 | 7,3 | 8,4 | 4,4 | 3,9 | 4,1 | 11,5 | 10,1 | 4 |
| APOALP | 24,7 | 15,2 | 13,5 | 4,1 | 7,9 | 8,2 | 9,4 | 5,7 | 4,2 | 4,9 | 13,2 | 11,2 | 4,4 |

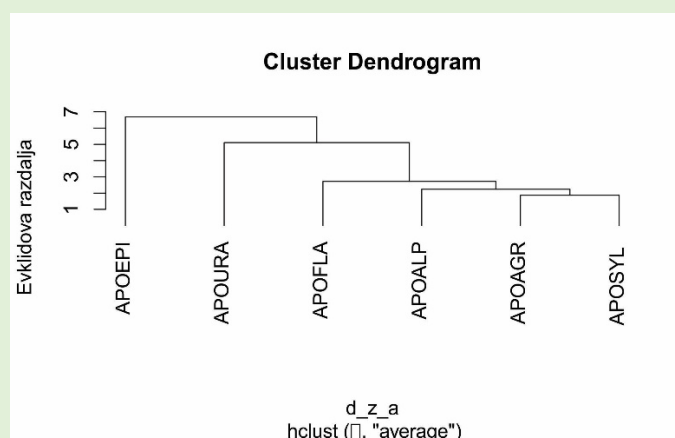
Preglednica 13.8 z-standardizirane vrednosti iz Preglednice 13.7

| | VL1 | VL2 | VL3 | VL4 | VL7 | VL8 | VL9 | VL13 | VL14 | VL16 | VL17 | VL18 | VL19 |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| APOEPI | 1,59 | 1,69 | 1,47 | 1,70 | 1,13 | 1,74 | 1,70 | 1,68 | 1,61 | 0,80 | 1,62 | 1,60 | 1,107 |
| APOAGR | -0,27 | -0,25 | -0,31 | -0,07 | -0,30 | -0,65 | -0,48 | -0,69 | 0,32 | 0,80 | -0,40 | -0,42 | -0,179 |
| APOFLA | 0,39 | 0,44 | 0,36 | 0,33 | 0,35 | 0,36 | 0,44 | 0,02 | 0,32 | 0,58 | 0,53 | 0,46 | 1,107 |
| APOSYL | -0,52 | -0,38 | -0,60 | -0,66 | -0,43 | -0,27 | -0,25 | 0,02 | -0,65 | -0,29 | -0,40 | -0,17 | 0,036 |
| APOURA | -1,39 | -1,31 | -1,42 | -1,25 | -1,60 | -1,16 | -1,28 | -1,28 | -1,29 | -1,82 | -1,33 | -1,43 | -1,465 |
| APOALP | 0,19 | -0,19 | 0,51 | -0,07 | 0,87 | -0,02 | -0,13 | 0,26 | -0,32 | -0,07 | -0,01 | -0,04 | -0,607 |

Standardizirane vrednosti iz Preglednice 13.7 smo vključili v izračun Evklidovih razdalj (Preglednica 13.9) in slednje vpeljali v izračun UPGMA drevesa. Najkrajšo razdaljo, in s tem največjo podobnost, izkazujeta vrsti *agrarius* – *sylvaticus* (razdalja = 1,86). Najdaljša razdalja je 10,4 pri paru *epimelas* – *uralensis*, ki sta največja in najmanjša vrsta rodu. Zaključimo lahko, da so razdalje med vrstami zaznamovane z velikostjo oz. da velikost živali zelo vpliva na prikaz odnosov med njimi. V drevesu morfometričnih odnosov med vrstami belonogih miši ne vidimo divergence, ampak sukcesivno dodajanje vrst (Slika 13.5). Najbolj sta si podobni vrsti *agrarius* in *sylvaticus*, njima se pridružuje vrsta *alpicola*, nato vrsta *flavicollis* in na koncu vrsta *epimelas*.

Preglednica 13.9 Simetrična matrika Evklidovih razdalj med posameznimi vrstami rodu *Apodemus*

| | APOEPI | APOAGR | APOFLA | APOSYL | APOURA | APOALP |
|--------|--------|--------|--------|--------|--------|--------|
| APOEPI | 0 | | | | | |
| APOAGR | 6,57 | 0 | | | | |
| APOFLA | 4,14 | 2,77 | 0 | | | |
| APOSYL | 6,80 | 1,86 | 3,00 | 0 | | |
| APOURA | 10,40 | 4,57 | 6,68 | 3,87 | 0 | |
| APOALP | 5,56 | 2,31 | 2,40 | 2,17 | 5,33 | 0 |



Slika 13.5 UPGMA drevo morfometričnih odnosov med vrstami rodu *Apodemus*. Vhodni podatki za izračun drevesa so simetrična matrika Evklidovih razdalj, navedeni so v Preglednici 13.9. Vrednost koeficienta kofenetske korelacije je 0,817.

```
R > package = »vegan«
##Izdelava dendrograma iz zveznih spremenljivk, data = aq
z_aq<-scale(aq) #z standardizacija
d_z_aq<-dist(z_aq, method = "euclidean") #Izračun razdalj
clus_d_z_aq<-hclust(d_z_aq, method = "average") #Izračun dendrograma
plot(clus_d_z_aq, hang = -1, axes = TRUE, ylab="Evklidova razdalja") #Izris dendrograma
cop<-cophenetic(clus_d_z_aq) #Kofenetska korelacija
cor(d_z_aq, cop) #Korelacija
```



PRIMER 13.4

Filogenetski odnosi v rodu *Apodemus* –
tradicionalna morfometrija – izločitev velikosti

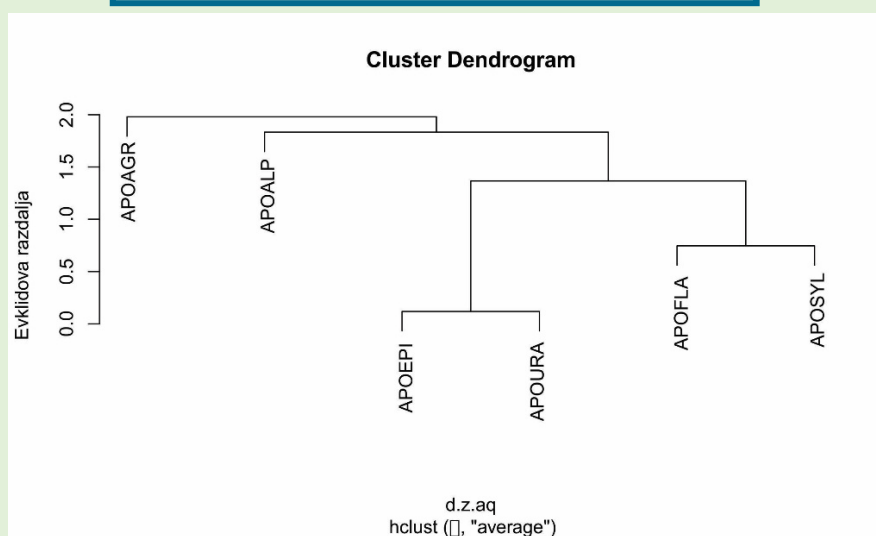
Na podobnost med vrstami zelo vpliva velikost živali, zato izločimo njen vpliv in ovrednotimo odnose na podlagi oblike. Izločitev velikosti in analizo oblike pričnemo s standardizacijo osnovnih podatkov (Preglednici 13.7 in 13.8) in nadaljujemo z njihovo vključitvijo v analizo glavnih komponent PCA. Izvedemo izračun glavnih komponent (Preglednica 13.10). Izločimo PC1, štiri naslednje glavne komponente (od PC2 do PC5)

vključimo v izračun Evklidovih razdalj (Preglednica 13.11). Matriko razdalj vpeljemo v klasterško analizo, s tem odnose med vrstami prikažemo v obliki drevesa (Slika 13.6). Vključitev oblikovnih PC v klasterško analizo pokaže največjo podobnost para *epimelas-uralensis* in *flavicollis-sylvaticus* (Slika 13.6).

Preglednica 13.10 Vrednosti 2., 3., 4. in 5. glavne komponente

| | PC 2 | PC 3 | PC 4 | PC 5 |
|---------------|-------|-------|-------|-------|
| APOEPI | -0,54 | -0,33 | -0,40 | 0,08 |
| APOAGR | 1,24 | 0,32 | -0,43 | 0,13 |
| APOFLA | 0,58 | -0,34 | 0,43 | -0,36 |
| APOSYL | -0,08 | -0,37 | 0,58 | 0,35 |
| APOURA | -0,61 | -0,41 | -0,37 | -0,14 |
| APOALP | -0,59 | 1,12 | 0,19 | -0,06 |

```
R > package = »factoextra«
#PCA, data = z-povprečjih Apodemus
» Izračun PC in izločitev PC1
pc<-dudi.pca(z_aq, nf = 13, scannf = FALSE) #PCA
fviz_eig(pc) #Graf eigenvalue
fviz_pca_var(pc) #Graf variabel
eig.val<-get_eigenvalue(pc) #Izračun eigenvalue
eig.val #Izpis eigenvalue
resultat.var<-get_pca_var(pc) #Izračun za variable
resultat.var$coord #Izpis za variable
res.ind<-get_pca_ind(pc) #Izračun za PCje
res.ind$coord #Izpis za osebkje
gk<-as.data.frame(res.ind$coord) #Izvoz PCjev
pc2.4<-gk[,2:4] #Izbor PC2-4
```



Slika 13.6 UPGMA drevo odnosov med belonogimi mišmi, izdelan na oblikovnih komponentah PC2–PC5. Vrednost koeficienta kofenetske korelacije je 0,924.

Preglednica 13.11 Simetrična matrika Evklidovih razdalj med vrstami rodu *Apodemus*

| | APOEPI | APOAGR | APOFLA | APOSYL | APOURA | APOALP |
|--------|--------|--------|--------|--------|--------|--------|
| APOEPI | 0 | | | | | |
| APOAGR | 1,89 | 0 | | | | |
| APOFLA | 1,46 | 1,36 | 0 | | | |
| APOSYL | 1,11 | 1,81 | 0,98 | 0 | | |
| APOURA | 0,24 | 2,01 | 1,45 | 1,19 | 0 | |
| APOALP | 1,57 | 2,10 | 1,91 | 1,67 | 1,63 | 0 |

```
R > package = »vegan«
# Izdelava dendrograma, data = pc2.4 (iz prejšnje analize)
d.z.aq<-dist(pc2.4, "euclidean") # Izračun Evklidovih razdalj
d.z.aq # prikaz vrednosti
clus.d.z.aq<-hclust(d.z.aq, method = "average") # Izračun dendrograma
plot(clus.d.z.aq, ylab = "Evklidova razdalja") # izpis dendrograma
#Cophenetska korelacija
cop<-cophenetic(clus.d.z.aq)
cor(d.z.aq, cop)
```

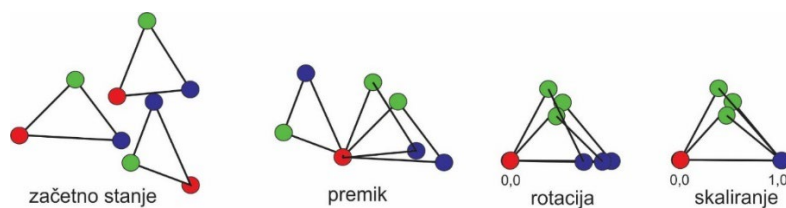


13.1.4 Geometrijska morfometrija

Geometrijska morfometrija je pristop, kjer so podatki podani v kartezičnih koordinatah. Pridobimo jih z določanjem oslonilnih točk, ki jih postavljamo na mesta homolognih anatomskih struktur. Meritve izvajamo na fotografijah in točke postavljamo v dvorazsežnem prostoru 2D na računalniškem monitorju ali v 3D prostoru skeniranih objektov. Vsaka oslonilna točka je opisana s koordinatami (X, Y oz. X, Y, Z), ki v seštevku povzemajo velikost in obliko objekta. Skozi celotno analizo se ohranjajo podatki o prostorskih razmerjih med točkami, kar omogoča prostorsko vizualizacijo sprememb objekta. Proces digitalizacije (fotografiranje) in digitizacije (merjenje) lahko spremlja več motečih fenomenov, na katere moramo biti pozorni in jih moramo izločiti pred statističnim vrednotenjem. Objekti raziskave so običajno kompleksne tridimenzionalne strukture, ki jih pri zajemanju 2D slik (fotografiranje) reduciramo v dve dimenziji. Zato je ključnega pomena pravilna orientacija objektov. Na vseh fotografijah moramo dodati merilo za umeritev merilne skale. Prvi korak v geometrijski morfometriji je Prokrustova analiza, v kateri izločimo razlike med objekti, ki so nastale v procesu digitalizacije. Konfiguracijo točk postavimo v skupno izhodišče s premikom in rotacijo, nato sledi tudi skaliranje (Slika 13.7). V Prokrustovi analizi se informacija o objektu razdeli v dve ločeni komponenti. Spremenljivka, ki opisuje velikost objekta, je centroidna velikost CS (ang. Centroid Size), spremenljivke, ki opisujejo obliko objekta, so Prokrustove koordinate. Centroidna velikost je definirana kot kvadratni koren vsote kvadriranih razdalj med

oslonilnimi točkami in njihovim težiščem (centroidom). Prokrustove koordinate so nove vrednosti kartezijskih koordinat z izločenim vplivom začetnih pogojev (postavitev vseh objektov v isto izhodišče in umeritev vpliva goriščne razdalje) in z odstranjenjo velikostjo (Klenovšek 2014).

Razstavitev informacije na velikost in obliko omogoča njuno ločeno analiziranje. Vrednosti centroidne velikosti in Prokrustovih koordinat vpeljemo v nadaljnje statistične analize. V Primeru 13.5 je predstavljen proces morfometrične analize in filogenetske rekonstrukcije z metodami geometrijske morfometrije.



Slika 13.7 Prokrustova analiza na primeru trikotnikov: začetno stanje, premik, rotacija v isto izhodišče in skaliranje. Trikotniki so objekti in krogi oslonilne točke. Homologe točke so obarvane z isto barvo (prirejeno po Bookstein 1991).

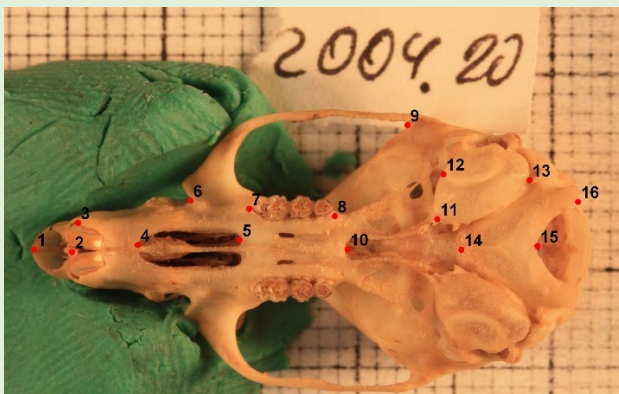


PRIMER 13.5

Filogenetski odnosi v rodu *Apodemus* –
tradicionalna morfometrija – izločitev velikosti

Geometrijska morfometrija temelji na meritvah in analizi oslonilnih točk (Slika 13.8). Z oslonilnimi točkami določimo položaj homolognih morfoloških struktur v kartezijskem koordinatnem sistemu. Vrednosti meritev zapišemo z navedbo X in Y koordinate. Primer meritve lobanje je naveden v Preglednici 13.12.

Meritve vseh obravnavanih osebkov poravnamo s Prokrustovo analizo (Slika 13.7) v programu MorphoJ (Klingenberg 2011). Rezultat so podatki o centroidni velikosti in o obliki lobanje – Prokrustove koordinate. Slednje vpeljemo v analizo glavnih komponent, s čimer zmanjšamo razsežnost podatkovnega niza.

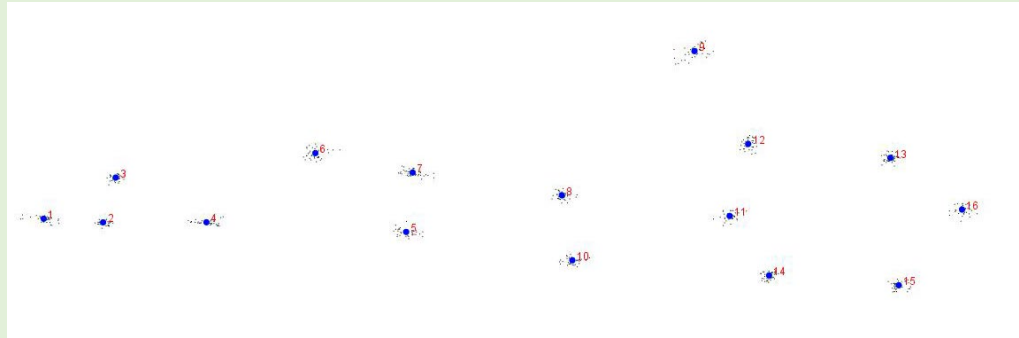


Slika 13.8 Ventralna stran lobanje belonoge miši z označenimi 16 oslonilnimi točkami: 1 vrh – stik nosnih kosti; 2 središčnica stika sekalcev; 3 proksimalni lateralni rob alveole; 4 proksimalni rob nebne špranje; 5 distalni rob nebne špranje; 6 rob infraorbitalnega kanala; 7 prednji rob alveole M1; 8 zadnji rob alveole M3; 9 notranji rob ličnega loka – posteriorni rob orbite; 10 rob trdega neba; 11 proksimalni podaljšek bobničnega mehurja; 12 prednji rob bobničnega mehurja; 13 zadnji rob bobničnega mehurja; 14 sutura zatilnice in baziokcipitalne kosti; 15 ventralni rob velike line; 16 rob zatilničnega čvrša. Sestavini fotografije sta tudi merilna skala, ki omogoča skaliranje, in inventarna številka, ki omogoča sledenje.

Foto. Boris Kryštufek.

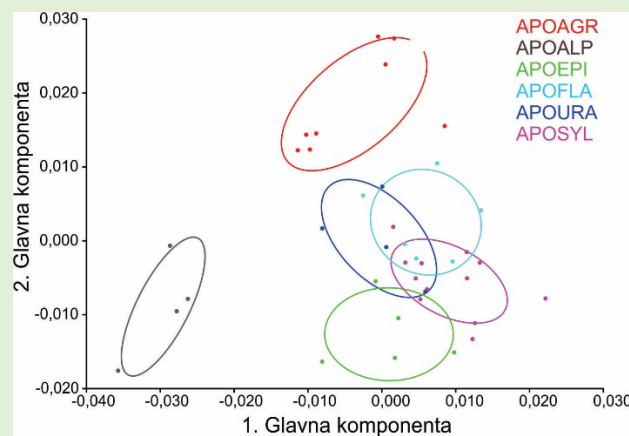
Preglednica 13.12 Vrednosti 16 oslonilnih točk (LM) meritve ventralne strani lobanje (Slika 13.8) za osebek z ID številko IMAGE=2004.20; in SCALE koeficientom, s katerim umerimo merilno skalo

LM=16
 500 1572
 685 1581
 704 1729
 1008 1597
 1593 1614
 1328 1860
 1624 1797
 2107 1738
 2442 2204
 2139 1561
 2624 1733
 2695 1949
 3156 1920
 2809 1544
 3229 1527
 3403 1758
 IMAGE=2004.20.JPG
 SCALE=0.009



Slika 13.9 Poravnane oslonilne točke. Velike pike – povprečne točke, majhne pike – posamezne meritve.

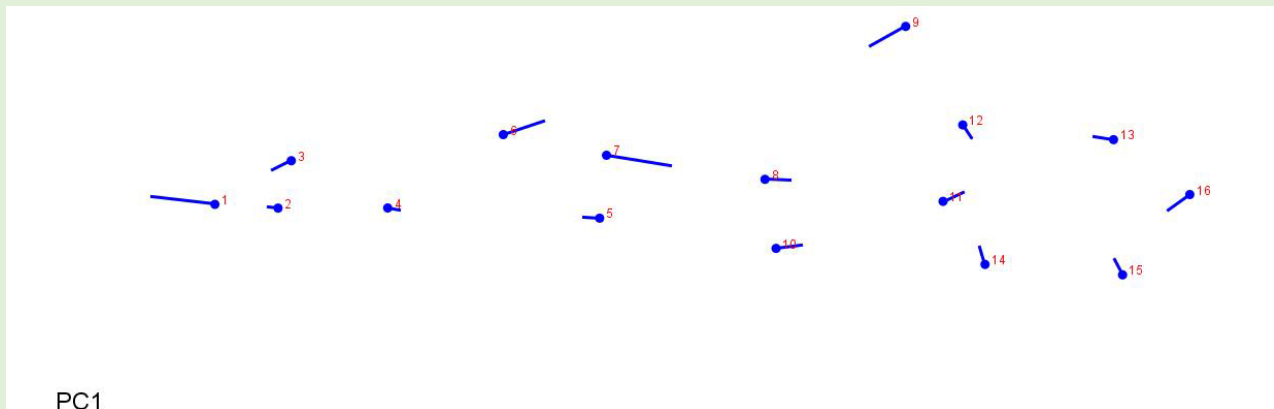
Odnosi med vrstami so prikazani kot projekcija osebkov na prvi dve glavni komponenti (Slika 13.10), ki skupaj razložita 46,8 % oblikovne variabilnosti ventralne strani lobanje (PC1–25,1 % in PC2–21,7 % variabilnosti). Vrsta *A. alpicola* se loči od ostalih vrst vzdolž prve osi, saj je izolirana v negativnem območju, daleč stran od preostalih vrst. Preostale vrste so razporejene okrog povprečja. Vzdolž druge osi si sledijo od negativnega proti pozitivnemu območju: *A. epimelas* (podrod *Karstomys*), *A. sylvaticus*, *A. uralensis* in *A. flavicolis* (podrod *Sylvaemus*) ter vrsta *A. agrarius* (podrod *Apodemus*) z najvišjo pozitivno vrednostjo PC2.



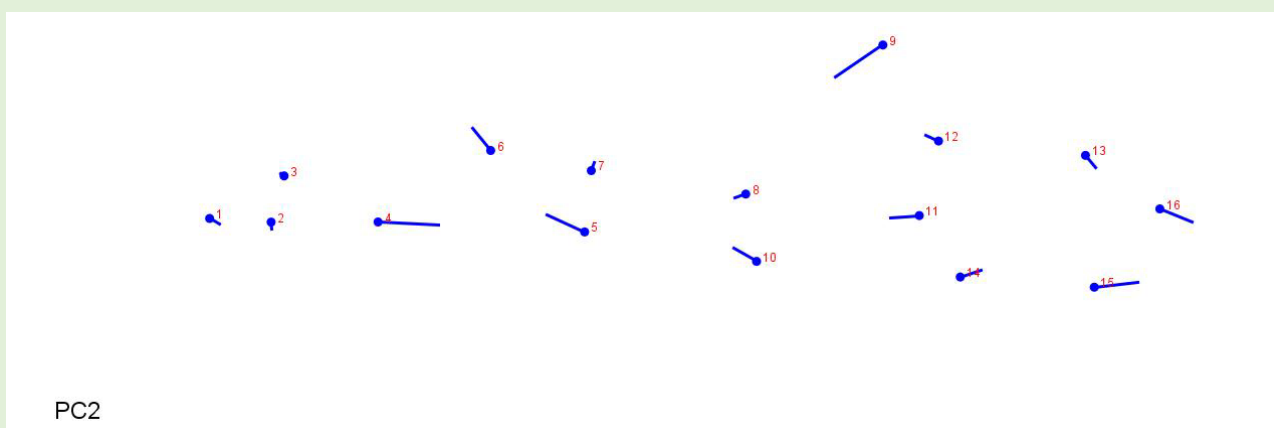
Slika 13.10 Projekcija 39 osebkov 6 vrst rodu *Apodemus* na prvi dve glavni komponenti. Razporeditev osebkov posameznih vrst je prikazana z 80 % elipsami zaupanja. Položaj posameznega osebka v morfoprostoru je prikazan s piko.

Smer in jakost oblikovnih sprememb, kot jih razlagata PC1 in PC2, je prikazana z vektorji ob posamezni točki (Sliki 13.11 in 13.12). Variabilnost vzdolž PC1 opisuje razlike med *A. alpicola* in preostalimi vrstami. Pika predstavlja povprečje vseh vrst in vektorji lastnosti *A. alpicola* glede na povprečje rodu. Izstopajo tri morfološke spremembe: 1. daljšanje nosnih

kosti LM1. 2. območje meljakov in ličnega loka, niz meljakov opisujeta točki LM7 in 8, ki sta pomaknjeni nazaj, lični lok LM6 in 9 sta usmerjeni ena proti drugi, kar pomeni, da je območje očesa zmanjšano pri *A. alpicola*. 3. tretji segment zadeva bobnični mehur. Vse točke LM11, 12, 13, 14, 15 in 16 so usmerjene navznoter (mediano), kar pomeni manjšanje bobničnega mehurja pri *A. alpicola*.



Slika 13.11 Variabilnost oslonilnih točk vzdolž PC1.



Slika 13.12 Variabilnost oslonilnih točk vzdolž PC2.

Divergenca med podrodovi *Apodemus-Sylvaemus* in *Karstomys* je očitna vzdolž PC2. Na Sliki 13.12 so z vektorji predstavljene lastnosti vrste *A. agrarius*, ki ima kratko nebno špranjo LM4–5, ožjo lobanjo, ki jo opisuje LM19, in večji bobnični mehur. Vsi vektorji bobničnega mehurja LM11, 12, 13, 14, 15 in 16 se razhajajo.

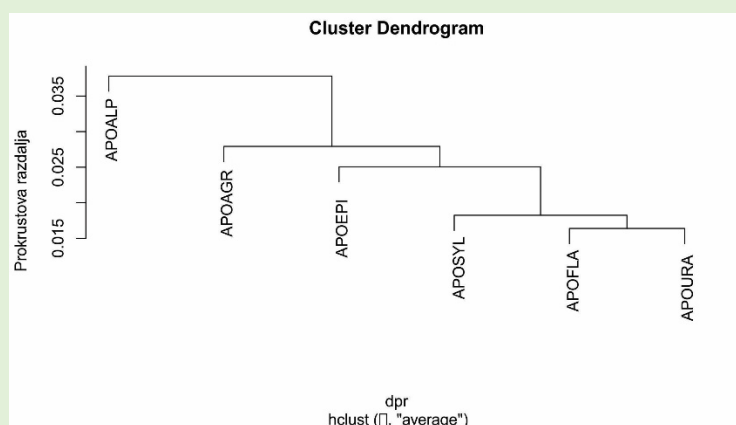
Izračunali smo Prokrustove razdalje (Preglednica 13.13) med 6 vrstami rodu *Apodemus*. Najdaljše razdalje med pari $> 0,0339$, in s tem največjo razliko, izkazuje vrsta *A. alpicola*. Po drugi strani izkazujeta najkrajšo razdaljo 0,0164, in s tem najvišjo podobnost vrsti *A. flavicollis* in *A. uralensis*.

Preglednica 13.13 Prokrustove razdalje med vrstami rodu *Apodemus*

| | APOAGR | APOALP | APOEPI | APOFLA | APOURA | APOSYL |
|--------|--------|--------|--------|--------|--------|--------|
| APOAGR | 0 | | | | | |
| APOALP | 0,0392 | 0 | | | | |
| APOEPI | 0,0334 | 0,0365 | 0 | | | |
| APOFLA | 0,0246 | 0,0397 | 0,0258 | 0 | | |
| APOURA | 0,0250 | 0,0339 | 0,0273 | 0,0164 | 0 | |
| APOSYL | 0,0287 | 0,0397 | 0,0221 | 0,0191 | 0,0174 | 0 |

Prokrustove razdalje smo vključili v analizo razvrščanja v skupine (klastrska analiza), ki je izrisala drevo na Sliki 13.13. Samostojno skupino oblikuje vrsta *A. alpicola*, ki se v ventralni strani lobanje najbolj razlikuje od preostalih vrst. Med preostalimi vrstami izrazito odstopa *A. agrarius*.

Kofenetska korelacija ima vrednost $r = 0,95$, kar pomeni, da dendrogram zelo dobro povzema odnose med vrstami, kot jih izkazuje matrika Prokrustovih razdalj.



Slika 13.13 UPGMA drevo podobnosti med 6 vrstami rodu *Apodemus*. Vhodni nabor za izračun drevesa je simetrična matrika vrednosti Prokrustovih razdalj, navedene v Preglednici 13.13). Vrednost koeficienta kofenetske korelacije je 0,924.



13.1.5 Metoda združevanja sosedov

Metoda združevanja sosedov NJ (ang. Neighbor-Joining Method) gradi topologijo drevesa z združevanjem najbližjih sosedov iz matrike razdalj med taksoni. Klastrska metoda združevanje sosedov zgradi nekoreninjeno aditivno drevo od spodaj navzgor (od listov proti korenini) s pristopom AGNES (glej Poglavje 7.5.1). Po navadi jo uporabimo za gradnjo filogenetskega drevesa na molekularnih podatkih (zaporedjih nukleotidov ali

aminokislinskih zaporedjih), izračun je možen tudi na morfoloških in drugih fenetskih podatkih. Pravzaprav je NJ metoda poenostavitev metode najmanjše evolucije (ang. Minimum Evolution Method), pri kateri konstrukcija optimalnega drevesa vključuje izračun vseh možnih topologij drevesa in oceno njihovih dolžin vej. Z metodo NJ prikažemo stopnjo podobnosti med taksoni.

Za genetske razdalje, ki jih vključimo v NJ veljajo priporočila: Jukes-Cantorjevo razdaljo uporabimo, kadar podatki izkazujejo multiple substitucije na istem mestu; Kimura 2 razdaljo uporabimo v primerih, kadar so tranzicije in transverzije prisotne v različnih deležih; Tamura-Neijevo razdaljo uporabimo, kadar deleži nukleotidov odstopajo od 0,25. Predhodno preverimo vrednost povprečne Jukes-Cantor razdalje. V primerih, ko presega vrednost 1 (average JC > 1), takšni podatki niso primerni za vključitev v NJ (za razlago glej Poglavlje 14).

Vhodni podatki za gradnjo drevesa NJ so oblikovani v simetrično matriko razdalj D med vsemi pari taksonov (OTU x OTU). Algoritem gradnje drevesa se začne z nerazrešenim drevesom z osrednjo razvejitvijo, iz katerega izhajajo vsi taksoni v obliki zvezde (Slika 13.14.A). Algoritem se izvaja s ponovitvami do popolne razrešitve drevesa in izračuna dolžine vseh vej. Postopek gradnje drevesa po metodi NJ vključuje naslednje korake:

1. Oblikujemo pravokotno matriko M velikosti $n \times z$, n taksonom in z znakov.
2. Iz matrike M izračunamo simetrično matriko D razdalj med taksoni velikosti $n \times n$ med n taksoni.
3. D matriko vključimo v izračun simetrične matrike Q $n \times n$ s povprečnimi razdaljami med taksoni.
4. V Q matriki poiščemo par taksonov i in j , med katerima ima razdalja q_{ij} najnižjo vrednost. Ustvarimo novo razvejitev u , s katerim sta taksona i in j povezana, in razvejitev u povežemo z osrednjo razvejitvijo (Slika 13.14.B).
5. Izračunamo dolžino veje med taksonom i in razvejitvijo u ter taksonom j in razvejitvijo u . Veji naneseemo na graf.
6. Izračunamo razdaljo vsakega preostalega taksona k do nove razvejitve u in matriko D posodobimo v novo matriko D_1 tako, da je namesto taksonov i in j nova razvejitev u (združena taksona i in j).
7. Sledi ponovitev algoritma v točki 3, izračun matrike Q_1 in iskanje razdalje z najnižjo vrednostjo med pari taksonov.
8. Algoritem od koraka 3 do 7 ponavljamo do združitve vseh taksonov in razrešitve drevesa. Število ponovitev je enako $n - 2$.

9. Zvezdni graf z vsakim dodanim taksonom oz. razvejitvijo posodobimo v skladu z dolžinami vej (Slika 13.14).

Q matrika

Iz diagonalne matrike razdalj D velikosti $n \times n$ taksonov izračunamo matriko povprečnih razdalj Q velikosti $n \times n$ s formulo:

$$q_{ij} = (n - 2)d_{ij} - \sum_{k=1}^n d_{ik} - \sum_{k=1}^n d_{jk} ,$$

q_{ij} je razdalja med taksonoma i in j v matriki Q , d_{ij} je razdalja med taksonoma i in j iz matrike D , d_{ik} so razdalje med taksonom i in vsemi preostalimi taksoni ter d_{jk} so razdalje med taksonom j in vsemi preostalimi taksoni.

Dolžina vej med združenim parom taksonov in nova razvejitev

Za vsak par, ki ga združimo v novi razvejitvi, izračunamo dolžino veje za vsak takson do razvejitve po formulah:

$$\delta_{fu} = \frac{d_{fg}}{2} + \frac{1}{2(n-2)} \left[\sum_{k=1}^n d_{fk} - \sum_{k=1}^n d_{gk} \right] \text{ in}$$

$$\delta_{gu} = d_{fg} - \delta_{fu}$$

δ_{fu} je dolžina veje med taksonom f in razvejitvijo u ter δ_{gu} je dolžina veje med taksonom g in razvejitvijo u .

f in g je par taksonov, združen v novi razvejitvi u . Veji, ki povezujeta taksona f ter g z razvejitvijo u in imata dolžino δ_{fu} in δ_{gu} dodamo k drevesu, in tako postopno preoblikujemo drevo (Slika 13.14). Slednje ne vpliva na združevanje naslednjih sosedov.

Razdalja preostalih taksonov do nove razvejitve

Za vsak takson, ki še ni bil vključen v predhodnem koraku, izračunamo razdaljo do nove razvejitve po formuli:

$$d_{uk} = \frac{1}{2} (d_{fk} + d_{gk} - d_{fg}),$$

kjer je n nova razvejitev, k je naslednja razvejitev, do katerega želimo izračunati razdaljo, f in g sta nov združeni par. Z izračunom novih razdalj posodobimo matriko razdalj D v novo simetrično matriko D_1 . Matrika D_1 je za en takson manjša, saj smo v predhodnem koraku dva taksona združili v novi razvejitvi. Sledi iteracija do končne razrešitve drevesa. Postopek izračunov in gradnje drevesa je predstavljen v Primeru 13.13.



PRIMER 13.6

Filogenetska rekonstrukcija amniotov – združevanje sosedov NJ

Naš namen je predstavitev filogenetskih odnosov na morfoloških znakih. V filogenetsko sklepanje smo vključili šest taksonov amniotov in dodatno zunanjo skupino. Pričakovani rezultat je nekoreninjeno in neusmerjeno aditivno filogenetsko drevo, ki ponazarja podobnost med taksoni. Predstavniki taksonov, ki smo jih vključili v analizo, so: želva, kuščar, krokodil, kokoš, kenguru in zajec ter riba kot zunanja skupina. Morfološke lastnosti taksonov smo opisali s 14 znaki (Preglednica 13.14) po binarnem načelu prisotno/odsotno (0/1).

Podatki in izračun razdalje med taksoni

Prvi korak v analizi združevanja sosedov NJ je izračun razdalje med vsemi pari taksonov. Binarno vrednosti 0/1 iz Preglednice 13.14 smo vključili v izračun Jaccardovega koeficienta podobnosti S_1 (za izračun glej Poglavje 13.1.1). V naslednjem koraku smo koeficient podobnosti spremenili v Jaccardovo razdaljo s preračunom $d = 1 - S_1$. Razdalja med kengurujem in zajcem ima vrednost $d_{fg} = 0,333$. Razdalje med vsemi pari so oblikovane v simetrično matriko razdalj D (Preglednica 13.15).

Preglednica 13.14 Pravokotna matrika morfoloških lastnosti sedmih taksonov

| Skupina | Takson | Luske | Amnion | Dlaka | Endotermija | 1 slušna kost | Diapsidna lobanja | Oklep | Okno v mandibuli | Kloaka | Perje | Panizzijeva odprtina | Euiviparni | Placenta | Marzupij |
|--------------|----------|-------|--------|-------|-------------|---------------|-------------------|-------|------------------|--------|-------|----------------------|------------|----------|----------|
| Osteichthyes | Riba | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Chelonia | Želva | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Lepidosauria | Kuščar | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Archosauria | Krokodil | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| Aves | Kokoš | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Marsupialia | Kenguru | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Placentalia | Zajec | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |

Preglednica 13.15 Simetrična matrika razdalj D med taksoni, pred vsakim taksonom je črka, s katero bo takson naveden v računskih zapisih

| D | a-Riba | b-Želva | c-Kuščar | d-Krokodil | e-Kokoš | f-Kenguru | g-Zajec |
|------------|--------|---------|----------|------------|---------|-----------|---------|
| a-Riba | 0 | | | | | | |
| b-Želva | 1,000 | 0 | | | | | |
| c-Kuščar | 1,000 | 0,600 | 0 | | | | |
| d-Krokodil | 1,000 | 0,667 | 0,500 | 0 | | | |
| e-Kokoš | 1,000 | 0,714 | 0,571 | 0,429 | 0 | | |
| f-Kenguru | 1,000 | 0,857 | 0,875 | 0,889 | 0,778 | 0 | |
| g-Zajec | 1,000 | 0,857 | 0,875 | 0,889 | 0,778 | 0,333 | 0 |

Izračun Q matrike

Drugi korak je izračun simetrične matrike Q . Vrednost n je število taksonov, v tem primeru jih je 7, vrednost k pomeni vse taksoni. Za vsak par izračunamo q razdaljo, prikazan je postopek izračuna razdalje med zajcem in kengurujem, formula ima obliko:

$$q_{fg} = (n - 2) * d_{fg} - \text{SUM}(d_{fk}) - \text{SUM}(d_{gk}) = (7 - 2) * 0,333 - 4,732 - 4,732 = -7,798$$

q razdalja med zajcem in kengurujem ima vrednost $q_{fg} = -7,798$, razdalje med vsemi pari so podane v Preglednici 13.16.

Preglednica 13.16 Simetrična matrika q razdalj med vsemi pari taksonov, v krepkem tisku je najnižja vrednost

| Q | a-Riba | b-Želva | c-Kuščar | d-Krokodil | e-Kokoš | f-Kenguru | g-Zajec |
|------------|--------|---------|----------|------------|---------|---------------|---------|
| a-Riba | 0 | | | | | | |
| b-Želva | -5,695 | 0 | | | | | |
| c-Kuščar | -5,695 | -6,117 | 0 | | | | |
| d-Krokodil | -5,695 | -5,783 | -6,294 | 0 | | | |
| e-Kokoš | -5,695 | -5,545 | -5,937 | -6,500 | 0 | | |
| f-Kenguru | -5,695 | -4,831 | -4,419 | -4,198 | -5,113 | 0 | |
| g-Zajec | -5,695 | -4,831 | -4,419 | -4,198 | -5,113 | -7,798 | 0 |

V matriki q razdalj poiščemo par z najnižjo vrednostjo, to je razdalja med zajcem in kengurujem $q_{fg} = -7,798$. Par z najnižjo vrednostjo združimo v razvejitvi u (Slika 13.14.B).

Izračun dolžine vej do razvejitve

Zajca in kenguruja (taksona f in g) smo združili, in s tem oblikovali novo razvejitev u (Slika 13.14.B). Dolžino veje $f-u$ in $g-u$ izračunamo s formulo:

$$\delta_{fu} = d_{fg}/2 + ((\text{SUM}(d_{fk}) - \text{SUM}(d_{gk})) / 2(n - 2)) = 0,167 + ((4,732 - 4,732) / 2(7 - 2)) = 0,167$$

$$\delta_{gu} = d_{fg} - \delta_{fu} = 0,333 - 0,167 = 0,167$$

Vrednost δ_{fu} je dolžina veje od kenguruja do razvejitev u in vrednost δ_{gu} je dolžina veje od zajca do razvejitev u . Povezave in dolžine vej nanesemo na drevo (Slika 13.14.B).

Prva posodobitev matrike razdalj D

Ko smo izračunali novo razvejitev u , sledi izračun posodobljenih razdalj med preostalimi taksoni in novo razvejitvijo, s tem matriko D posodobimo v matriko D_1 (Preglednica 13.17). Razdalje med razvejitvijo u in preostalimi taksoni smo izračunali v naslednjih postopkih:

$$\text{Razdalja } d_{au} \text{ do ribe: } d_{au} = (d_{af} + d_{ag} - d_{fg})/2 = (1 + 1 - 0,33)/2 = 0,833.$$

$$\text{Razdalja } d_{bu} \text{ do želve: } d_{bu} = (d_{bf} + d_{bg} - d_{fg})/2 = (0,857 + 0,857 - 0,33)/2 = 0,691.$$

$$\text{Razdalja } d_{cu} \text{ do kuščarja: } d_{cu} = (d_{cf} + d_{cg} - d_{fg})/2 = (0,875 + 0,875 - 0,33)/2 = 0,708.$$

$$\text{Razdalja } d_{du} \text{ do krokodila: } d_{du} = (d_{df} + d_{dg} - d_{fg})/2 = (0,889 + 0,889 - 0,33)/2 = 0,722.$$

$$\text{Razdalja } d_{eu} \text{ do kokoši: } d_{eu} = (d_{ef} + d_{eg} - d_{fg})/2 = (0,778 + 0,778 - 0,33)/2 = 0,611.$$

Preglednica 13.17 Simetrična matrika D_1 razdalj med taksoni, taksona Kenguru in Zajec sta združena v u

| D1 | a-Riba | b-Želva | c-Kuščar | d-Krokodil | e-Kokoš | u |
|------------|--------|---------|----------|------------|---------|---|
| a-Riba | 0 | | | | | |
| b-Želva | 1,000 | 0 | | | | |
| c-Kuščar | 1,000 | 0,600 | 0 | | | |
| d-Krokodil | 1,000 | 0,667 | 0,500 | 0 | | |
| e-Kokoš | 1,000 | 0,714 | 0,571 | 0,429 | 0 | |
| u | 0,833 | 0,691 | 0,708 | 0,722 | 0,611 | 0 |

Vrednosti iz matrike D_1 vključimo v izračun matrike Q_1 (Preglednica 13.18). V matriki Q_1 je najnižja vrednost razdalje med paroma u in ribo $q_{au} = -5,171$.

Preglednica 13.18 Simetrična matrika Q_1 razdalj med taksoni, taksona Kenguru in Zajec sta združena v u , v krepkem tisku je najnižja vrednost

| Q1 | a-Riba | b-Želva | c-Kuščar | d-Krokodil | e-Kokoš | u |
|------------|---------------|---------|----------|------------|---------|---|
| a-Riba | 0 | | | | | |
| b-Želva | -4,505 | 0 | | | | |
| c-Kuščar | -4,505 | -4,651 | 0 | | | |
| d-Krokodil | -4,505 | -4,385 | -4,697 | 0 | | |
| e-Kokoš | -4,505 | -4,195 | -4,412 | -4,929 | 0 | |
| u | -5,171 | -4,289 | -3,864 | -3,754 | -4,446 | 0 |

Drugi izračun dolžine vej do razvejitev

Ribo (takson a) smo združili s taksonom u in s tem oblikovali novo razvejitev v (Slika 13.14.C). Dolžino veje $a-v$ in $u-v$ izračunamo s formulo:

$$\delta_{av} = d_{av}/2 + ((\text{SUM}(d_{ak}) - \text{SUM}(d_{uk})) / 2(n - 2)) = 0,417 + ((4,833 - 3,565) / 2(6 - 2)) = 0,575$$

$$\delta_{uv} = d_{uv} - \delta_{av} = 0,833 - 0,575 = 0,258$$

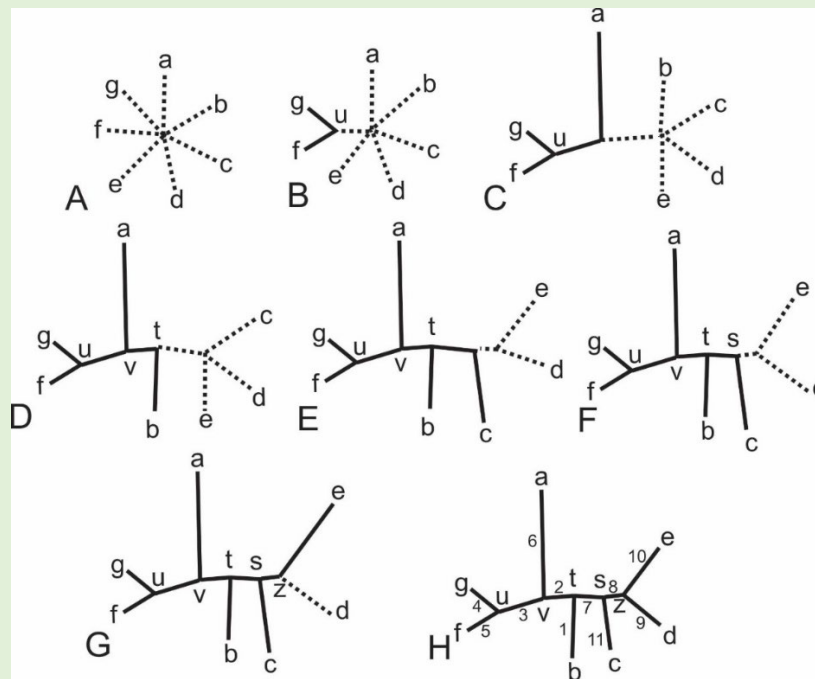
Vrednost δ_{av} je dolžina veje od ribe do razvejitev v in vrednost δ_{uv} je dolžina veje od razvejitev u do nove razvejitev v . Povezave in dolžine vej naneseemo na drevo (Slika 13.14.C).

Druga in naslednje iteracije

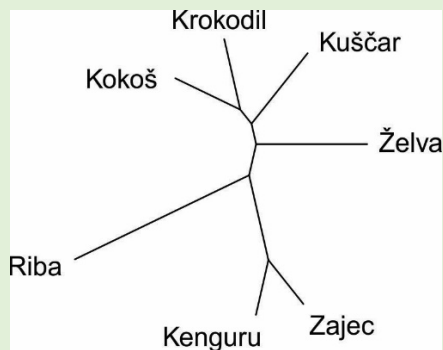
Potem ko smo izračunali novo razvejitev v , sledi izračun posodobljenih razdalj med preostalimi taksoni in novo razvejitev, s tem matriko D_1 posodobimo v matriko D_2 . Vrednosti iz matrike D_2 vključimo v izračun matrike Q_2 . V matriki Q_2 poiščemo najkrajša razdalja med paroma. Naslednji korak je izračun dolžine vej in nato sledi nova iteracija. Število iteracij do razrešitve drevesa je $n - 2$. Primer, ki ga rešujemo s sedmimi taksoni, bo razrešen v petih iteracijah. Predstavljen je izračun začetne in prve iteracije. Posamezni koraki v izgradnji radialnega drevesa so predstavljeni na Slikah 13.14.A–H. Konča oblika topologije drevesa je predstavljena na Slikah 13.14.H in 13.15 ter dolžina vej v Preglednici 13.19.

Preglednica 13.19 Povezave in dolžine vej nekoreninjenega aditivnega drevesa po metodi združevanja sosedov (črke se nanašajo na razvejitvi $s-z$ in taksoni $a-g$ na Sliki 13.14.H)

| Veja | Razvejitev | Takson, Razvejitev | Dolžina veje |
|------|------------|--------------------|--------------|
| 1 | 8 - t | 2 - Želva, b | 0,3234 |
| 2 | 8 - t | 11 - v | 0,1052 |
| 3 | 11 - v | 12 - u | 0,2582 |
| 4 | 12 - u | 6 - Zajec, g | 0,1667 |
| 5 | 12 - u | 7 - Kenguru, f | 0,1667 |
| 6 | 11 - v | 1 - Riba, a | 0,5751 |
| 7 | 8 - t | 9 - s | 0,0540 |
| 8 | 9 - s | 10 - z | 0,0710 |
| 9 | 10 - z | 4 - Krokodil, d | 0,2037 |
| 10 | 10 - z | 5 - Kokoš, e | 0,2249 |
| 11 | 9 - s | 3 - Kuščar, c | 0,2505 |



Slika 13.14 Postopna gradnja nekoreninjenega in aditivnega drevesa po metodi združevanja sosedov. A Začetna razporeditev taksonov v obliki zvezde; H Končna oblika radialnega drevesa. Taksoni so označeni s črkami *a–g* (za razlago glej Preglednico 13.15) in istočasno predstavljajo sedem zunanjih razvejitev, notranjih razvejitev je pet in so označena s črkami *u, v, t, s* in *z*. Vej je skupno 11, notranje veje so oštevilčene 3, 2, 7 in 8; zunanje veje so z oznakami 1, 5, 4, 6, 10, 9 in 11. Dolžina vej je navedena v Preglednici 13.19. Črtkana črta je začetno stanje, polna črta je izračunana dolžina in povezuje dve razvejitvi. Z vsakim korakom od slike B do H je dodan en par ali ena povezava.



Slika 13.15 Nekorinjeno in aditivno drevo, narejeno v R okolju s paketoma »vegan« in »ade4«. Vsebinsko je drevo identično kot na Sliki 13.14.H.

Filogenetsko drevo, zgrajeno na morfološki znakih po metodi združevanja sosedov, je najprej združilo taksona iz skupine sesalcev (Kenguru + Zajec). Podoben par sta oblikovala taksona iz skupine arhozavrov (Krokodil + Kokoš). Topologijo drevesa na področju plazilcev dopolnjujeta Kuščar in Želva, ki sta se postopno pridružila (((Krokodil + Kokoš) + Kuščar) + Želva). K skupini sesalcev se je priključila riba ((Kenguru + Zajec) + Riba). V zadnji peti iteraciji pride do povezave plazilcev in sesalcev (((Krokodil + Kokoš) + Kuščar) + Želva) + ((Kenguru + Zajec) + Riba)) (Sliki 13.14 in 13.15). Newick

format zapisa nekoreninjenega drevesa ima obliko: cluster = (((Kokoš: 0.2249, Krokodil: 0.2037): 0.7098, Kuščar: 0.2504): 0.0270, (((Zajec: 0.1667, Kenguru: 0.1667): 0.2582, Riba: 0.5751): 0.1052, Želva: 0.3234): 0.0270); 0;

Nekoreninjeno in aditivno filogenetsko drevo na Sliki 13.15 smo naredili v okolju R. Prvi korak je bil uvoz podatkov iz Preglednice 13.14 in njihova vključitev v izračun Jaccardovih razdalj (Preglednica 13.15) s paketom »vegan« in ukazom *vegdist()* (Oksanen in sod. 2019). Razdalje smo vključili v izračun in izris drevesa s paketom »ade4« in ukazom *nj()* (Dray in sod. 2020).

```
R > package = »vegan« , »ade4«
library(vegan)
library(ape)
placent # Podatki
jac<-vegdist(placent, method = "jaccard")
jac # Izračun Jaccardovih razdalj
nj<-nj(jac) # Izračun drevesa
plot(nj, "unrooted") # Print drevesa
nj$edge.length # Izpis dolžine vej
```



13.1.6 UPGMA metoda

UPGMA oz. metoda neuteženih aritmetičnih sredin (ang. Unweighted Pair Group Method with Arithmetic Mean UPGMA) je enostavno hierarhično združevanje v skupine (sin. klastrska analiza). UPGMA algoritem gradi koreninjeno ultrametrično drevo v pristopu AGNES od spodaj navzgor (glej Poglavje 7.5.1). Vhodni podatki za gradnjo UPGMA drevesa so urejeni v simetrično matriko D razdalj med taksoni. Algoritem na vsaki stopnji združi dve skupini (A in B) v novo skupino X . Predstavnik skupine A in B sta taksona i in j ali združeni taksoni v , med katerima je razdalja najmanjša $\min d_{ij}$ in ju poveže v novo razvejitev u . Razdaljo med novo razvejitvijo u in še ne vključenimi taksoni izračunamo s formulo:

$$d_{AB,X} = \frac{|A|*d_{AX}+|B|*d_{BX}}{|A|+|B|}.$$

Člena A in B imata vrednost enako številu taksonov, ki sestavljajo skupino A oz. skupino B . Nova razvejitev u , v katerem sta združena taksona i in j , pomeni zmanjšanje matrike razdalj za en takson. Zato posodobimo osnovno matriko razdalj D v novo matriko D_1 , ki

je za ena manjša od predhodne. Razdalje še ne vključenih taksonov k do razvejitve u v prvi posodobitvi izračunamo s formulo:

$$d_{uk} = \frac{1*d_{ik}+1*d_{jk}}{1+1}.$$

V drugi in naslednjih posodobitvah matrike razdalj upoštevamo, da skupino lahko sestavlja več taksonov, kar vpliva na vrednosti členov A in B .

Dolžina veje med razvejitvijo u in taksonom i oz. taksonom j je enaka in je enaka razdalji med taksonoma:

$$\delta_{iu} = \delta_{ju} = d_{ij}.$$

Združevanje skupin ponavljamo do razrešitve celotnega drevesa oz. združitve vseh taksonov. Vzporedno gradimo graf – drevo, postopek je opisan v Primeru 13.7.

Pri uporabi UPGMA drevesa za prikaz filogenije predpostavljamo, da je frekvenca evolucijskih sprememb v času enaka ali vsaj podobna za vse vključene taksoni. V računskem kontekstu to pomeni, da je razdalja od korenine do lista pri vseh taksonih enaka. Takrat dolžine vej v drevesu izpolnjujejo pogoj ultrametričnosti. Predpostavko o ultrametričnosti drevesa preverimo z **izpolnjevanjem pogoja treh točk** (ang. Three-point Condition). Trije taksoni v drevesu, povezani v klad $((a + b) + c)$, morajo izpolnjevati pogoj:

$$\begin{aligned} \delta_{ab} &\leq \max(\delta_{ac}, \delta_{bc}) \text{ ali} \\ \delta_{ac} &\leq \max(\delta_{ab}, \delta_{bc}) \text{ ali} \\ \delta_{cb} &\leq \max(\delta_{ab}, \delta_{ac}). \end{aligned}$$

Zapisano z besedami pomeni, da je med sestrskima taksonoma a in b dolžina vej krajša ali največ enaka dolžini vej katerega od sestrskih taksonov do zunanega taksona c . Še drugačna oblika zapisa: vsaj dve dolžini vej sta enaki, tretja veja pa je krajša ali so vse tri veje enako dolge. Pogoj preverjamo na poljubnih taksonih. Če je pogoj ultrametričnosti razdalj kršen, je drevo popačeno in ne prikaže relevantne filogenije. Z drugimi besedami, UPGMA metoda je zelo občutljiva na različno stopnjo evolucijskih sprememb.

**PRIMER 13.7****Filogenetska rekonstrukcija amniotov –
UPGMA metoda**

Namen in cilj tega primera je podoben, kot smo ga zapisali v Primeru 13.6, t.j. predstavitev filogenije amniotov. Pričakovani rezultat je ultrimetrično filogenetsko drevo, ki ponazarja hierarhično in dihotočno cepitev filogenetske linije in vzpostavlja relacijo med predniki in potomci. Vhodni podatki za izris filogenetskega drevesa po metodi UPGMA so isti kot v metodi NJ, opisani v Primeru 13.6. Osnovna matrika podatkov je podana v Preglednici 13.14. Razdalje med taksoni so izračunane z Jaccardovim koeficientom in podane v Preglednici 13.15.

V prvem koraku v matriki D poiščemo najkrajšo razdaljo, izkazuje jo par kenguru–zajec, in znaša $d_{fg} = 0,333$, povežemo ju v skupino u . Dolžina vej med taksonoma in razvejitevjo ima vrednost $\delta_{fu} = \delta_{gu} = d_{fg} = 0,333$ in ju nanese na graf (Slika 13.16.A).

Posodobimo simetrično matriko razdalj z izračunom razdalje med preostalimi taksoni in u in oblikujemo novo matriko D_1 (Preglednica 13.20):

Razdalja do ribe $d_{au} = d_{af} + d_{ag} = (1 + 1)/2 = 1$.

Razdalja do želve $d_{bu} = d_{bf} + d_{bg} = (0,857 + 0,857)/2 = 0,857$.

Razdalja do kuščarja $d_{cu} = d_{cf} + d_{cg} = (0,875 + 0,875)/2 = 0,875$.

Razdalja do krokodila $d_{du} = d_{df} + d_{dg} = (0,889 + 0,889)/2 = 0,889$.

Razdalja do kokoši $d_{eu} = d_{ef} + d_{eg} = (0,778 + 0,778)/2 = 0,778$.

Preglednica 13.20 Simetrična matrika razdalj D_1 med taksoni, najkrajša razdalja je v krepkem tisku

| D1 | Riba-a | Želva-b | Kuščar-c | Krokodil-d | Kokoš-e | u |
|------------|---------------|----------------|-----------------|-------------------|----------------|----------|
| Riba-a | 0 | | | | | |
| Želva-b | 1 | 0 | | | | |
| Kuščar-c | 1 | 0,600 | 0 | | | |
| Krokodil-d | 1 | 0,667 | 0,500 | 0 | | |
| Kokoš-e | 1 | 0,714 | 0,571 | 0,429 | 0 | |
| u | 1 | 0,857 | 0,875 | 0,889 | 0,778 | 0 |

V matriki D_1 poiščemo par z najkrajšo razdaljo. To sta krokodil in kokoš razdalja med katerima je $d_{de} = 0,429$ in ju povežemo v razvejitev v . Nadaljujemo z nanosom dolžine vej na graf in preračunom novih razdalj preostalih taksonov do razvejitve v in s tem posodobitve matrike razdalj v D_2 . Postopek ponavljamo do končne razrešitve drevesa (Preglednice 13.21 in 13.22, predstavljen je postopek samo do matrike D_3). Vzporedno veje nanašamo na graf (Slika 13.16). Nexus format končnega drevesa ima obliko `utree_cluster = (((Želva: 0,6603, (Kuščar: 0,5357, (Krokodil: 0,4286, Kokoš: 0,4286): 0,1071): 0,1246): 0,1894, (Kenguru: 0,3333, Zajec: 0,3333): 0,5164): 0,1503, Riba:1): 0`.

Dolžine vej za posamezni takson od lista do prve razvejitve so v razponu od 0,33 do 1 (Preglednica 13.23).

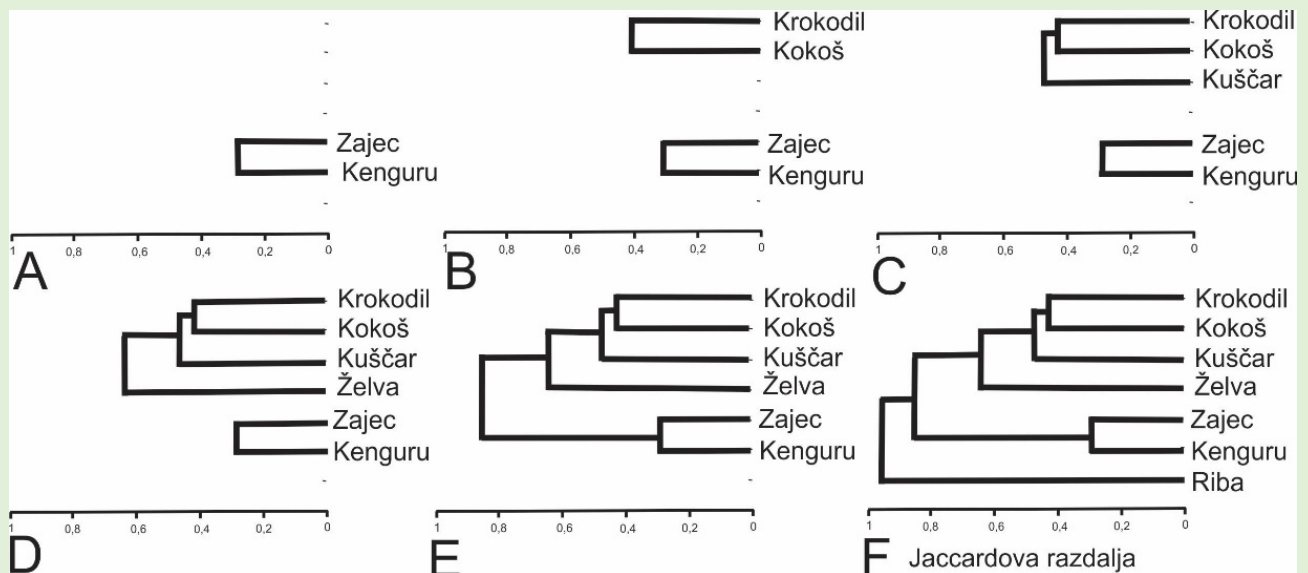
Izračun drevesa smo ponovili v okolju R s paketom »vegan« (Oksanenin sod. 2019), rezultat je predstavljen na Sliki 13.17. Filogenetsko drevo, zgrajeno na fenetskih podatkih (Sliki 13.16 in 13.17), je izrisalo odnose med taksoni, ki so skladni z monofiletskim konceptom in z današnjim znanjem o razvoju amniotov.

Preglednica 13.21 Simetrična matrika razdalj D_2 med taksoni, najkrajša razdalja je v krepkem tisku

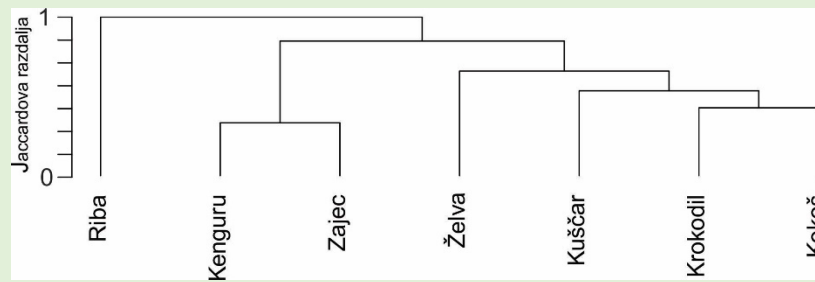
| D2 | Riba-a | Želva-b | Kuščar-c | v | u |
|----------|--------|---------|--------------|-------|---|
| Riba-a | 0 | | | | |
| Želva-b | 1 | 0 | | | |
| Kuščar-c | 1 | 0,600 | 0 | | |
| v | 1 | 0,691 | 0,536 | 0 | |
| u | 1 | 0,857 | 0,875 | 0,833 | 0 |

Preglednica 13.22 Simetrična matrika razdalj D_3 med taksoni, najkrajša razdalja je v krepkem tisku

| D3 | Riba-a | Želva-b | z | u |
|---------|--------|--------------|-------|---|
| Riba-a | 0 | | | |
| Želva-b | 1 | 0 | | |
| z | 1 | 0,660 | 0 | |
| u | 1 | 0,857 | 0,854 | 0 |



Slika 13.16 Gradnja ultrametričnega drevesa po metodi UPGMA. A Priprava merilne skale in položaja listov ter vnos prvega para taksonov (kenguru + zajec). B Vnos drugega para (krokodil + kokoš). C Vnos petega taksona ((krokodil + kokoš) + kuščar). D Vnos šestega taksona (((krokodil + kokoš) + kuščar) + želva). E Vnos povezave med plazilci in sesalci (((krokodil + kokoš) + kuščar) + želva) + (kenguru + zajec)). F V zadnjem koraku so se amnioti povezali z ribo kot zunanjo skupino (((((krokodil + kokoš) + kuščar) + želva) + (kenguru + zajec)) + riba).



Slika 13.17 UPGMA ultrametrično filogenetsko drevo zgrajeno v okolju R. Vsebinsko je drevo identično kot na sliki 13.16.F.

Preglednica 13.23 Dolžina veje za posamezni takson od lista do razvejitve

| Takson | Dolžina veje |
|------------|--------------|
| a-Riba | 1 |
| b-Želva | 0,660 |
| c-Kuščar | 0,536 |
| d-Krokodil | 0,429 |
| e-Kokoš | 0,429 |
| f-Kenguru | 0,333 |
| g-Zajec | 0,333 |

Preverimo tudi izpolnjevanja pogoja ultrametričnosti. Iz drevesa na Sliki 13.17 naključno izberimo trojico taksonov npr. (krokodil + kokoš) + želva), dolžine vej odčitamo iz Preglednice 13.23. Pogoj $\delta_{de} \leq \max(\delta_{db}, \delta_{eb})$; $0,66 \leq \max(0,66, 0,429)$ je izpolnjen in s tem smo potrdili, da je drevo na Slikah 13.16.F in 13.17 ultrametrično.

```
R > package = »vegan«
library(vegan)
placent # Podatki
jac<-vegdist(placent, method = "jaccard")
jac # Izračun Jaccardovih razdalj
hc_jac<-hclust(jac, method = "average")
plot(hc_jac, hang = -1, axes = TRUE)
hc_jac$height # Izpis dolžine vej
```



13.2 Kladistika

Kladistika (ang. Cladistics) je klasifikacija organizmov v skupine na podlagi hipoteze o **zadnjem skupnem predniku** (ang. Most Recent Common Ancestor MRCA). Kladistični pristop ob predpostavki o zadnjem skupnem predniku vključuje tudi spoznanja: 1. da se filogenetske linije v času in prostoru spreminjajo kar vidimo kot diskretne spremembe v

stanju znakov in jih lahko izmerimo; 2. da kladogram omogoča oblikovanje monofiletskih skupin; 3. da filogeneza poteka po načelu varčnosti.

Prepoznavanje zadnjega skupnega prednika in razumevanje evlucijskih odnosov med njegovimi potomci je istočasno povezano s prepoznavanjem naravnih skupin oz. razlikovanje med skupinami. V kladističnem kontekstu naravno skupino organizmov imenujemo **klad** (ang. Clade) (Cladistic 2023).

Skupina organizmov, ki jo sestavljajo zadnji skupni prednik, in vsi njegovi evlucijski potomci je **monofiletska skupina** (sin. klad, ang. Monophyletic Group, sin. Clade) (Monophyly 2023). Kadar oblikovanje skupine krši zgornje pravilo, imamo opravka s polifiletsko ali parafiletsko skupino (Slika 13.18). Ključni lastnosti monofiletske skupine so **sinapomorfije** (ang. Synapomorphy) in **homologije** (ang. Homology). Sinapomorfija je skupen izpeljani znak, ki je lahko molekularna, morfološka, funkcionalna ali fiziološka lastnost, lastna predstavnikom monofiletske skupine. Sinapomorfen znak je torej lasten zadnjemu skupnemu predniku in vsem njegovim evlucijskim potomcem, medtem ko zunanje skupine ali starejši predniki nimajo te lastnosti (Apomorphy and synapomorphy 2023). Znak je lahko pri posamezni skupini različno preoblikovan, zato moramo ugotoviti ali sta dve stanji homologni. Homologijo dveh primerjanih zaporedij nukleotidov potrdimo z ugotovijo, da obe zaporedji izhajata iz skupnega prednika. Morfološki homologni znak izhaja iz istega zarodnega tkiva in ima enotno organizacijo. Njegova oblika in funkcija variirata od zelo podobne do zelo različne (Homology 2023). Npr. prednja ekstremiteta vseh placentalnih sesalcev izhaja iz istega zarodnega materiala in ima identično notranjo morfološko organiziranost kosti. Zadnji domnevni skupni prednik placentalnih sesalcev, približno 66 milj. let star rovk podobna žival z imenom *Protungulatum donnae*, je imela petprstno okončino z nadlahtnico, podlahtnico, koželjnico ter dlančnicami, zapestnicami in prstnicami v funkciji hoje po tleh in plezanja. Pregled vseh evlucijskih potomcev razkrije, da se je prednja noga pri placentalnih sesalcih različno adaptirala, npr. pri nekaterih je še vedno v funkciji hoje po tleh (npr. medved), zmanjšalo se je število prstov (npr. konj), funkcionalno se je spremenila v oprijemalno roko (npr. človek), spremenila se je v plavut (npr. delfin) ali v prhut (npr. netopir). V procesu filogenetskega sklepanja moramo razlikovati med homologijo in **homoplazijo** (ang. Homoplasy). Pri homoplaziji gre za navidezno podobnost molekularnih ali morfoloških znakov, ki so posledica konvergentnega razvoja in/ali niso obravnavani znotraj monofiletske skupine. Molekularna homoplazija pomeni enako stanje dveh primerjanih znakov, identična nukleotida na istem mestu, ki nista rezultat skupnega prednika (nista homologna), ampak izhajata iz različnih prednikov in sta posledica neodvisne konvergentne mutacije. Primer morfološke homoplazije je npr. neposredno primerjanje

prednje okončine ptičih in netopirjih z domnevo, da sta obe adaptaciji za letenje rezultat istega evlucijskega dogodka. Adaptacija na letanje se je razvila neodvisno posebej pri ptičih in posebej pri netopirjih, tudi v različnem časovnem obdobju. Zato gre za navidezno podobnost iste morfološke strukture – prednje ekstremitete. Odnos med homologijo in homoplazijo je lahko tudi relativen. Z oblikovanjem monofiletske skupine tetrapodov z vključitvijo zadnjega skupnega prednika dvoživk, plazilcev, ptičev in sesalcev in vseh njihovih potomcev (Slika 13.20), prednja okončina netopirja in ptiča postane sinapomorfija in jo obravnavamo kot homologni znak. Filogenetski odnosi med vretenčarji, predstavljeni po kladističnem in tradicionalnem pristopu, so prikazani na Sliki 13.19 in 13.20.

Kladistična analiza poteka v naslednjih korakih, povzeto po Forey (2023):

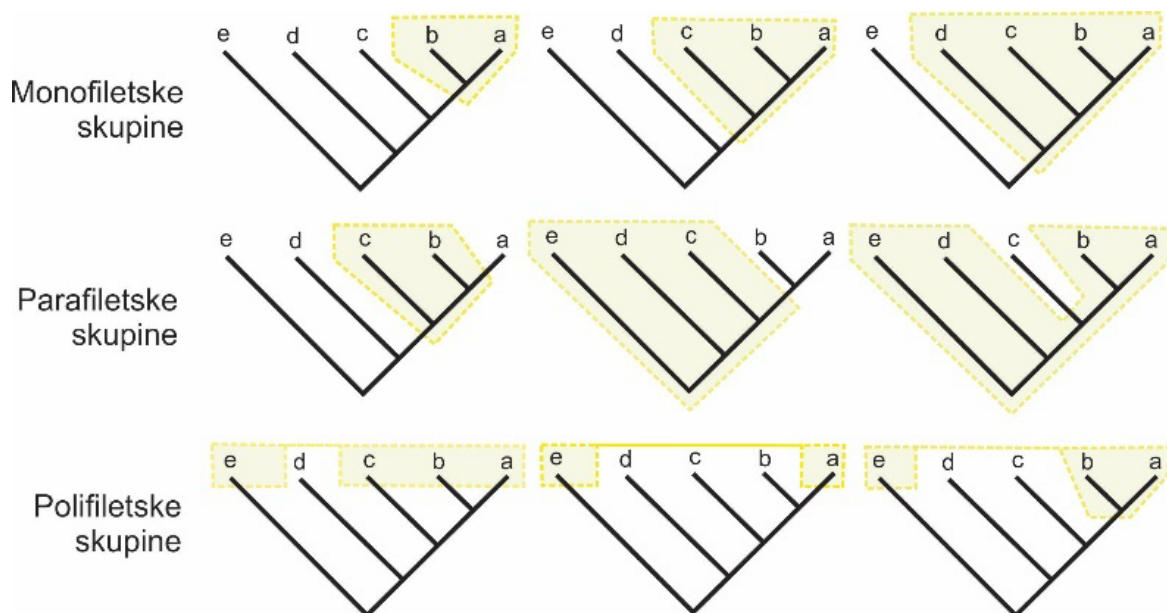
1. **Izbira taksonov.** Taksoni, vključeni v ugotavljanje evlucijskih odnosov, morajo oblikovati klad, to pomeni, da so monofiletska skupina.
2. **Izbira znakov.** Znaki so bodisi morfološki nominalni ali nukleotidi. V primeru analize nukleotidnih znakov jih organiziramo v poravnana in enako dolga ter homologna zaporedja. V analizi morfoloških struktur so vsi znaki nominalni, to pomeni da stanja lahko nedvoumno kategoriziramo bodisi v binarni obliki: da/ne oz. 1/0 bodisi v več kategorij npr.: a/b/c/d oz. 0/1/2/3. Spremenljivko z intervalno porazdelitvijo moramo predhodno preoblikovati v kategorične vrednosti, to so lahko povprečja posamezne skupine ali oblikujemo ranžirno vrsto in vrednosti razporedimo v range.
3. **Določanje polarnosti znakov.** Za vsak znak določimo polarnost, to pomeni njegovo izvorno stanje. S stališča taksona to pomeni ali je konkretna lastnost pri obravnavanem taksonu v izvornem ali spremenjenem (razvitem) stanju. Določanje polarnosti ni vedno nujno. Drug način določanja polarnosti je vključitev zunanje skupine, pri kateri se privzame, da ima vse znake v izvorni obliki.
4. **Oblikovanje skupin** po načelu sinapomorfij.
5. **Gradnja evlucijskih odnosov** po načelu varčnosti.
6. **Gradnja kladograma** po načelih: 1. vsi taksoni so na vrhu drevesa, ničesar ni na razvejitvah; 2. vsako razvejitev na kladogramu ima označeno sinapomorfijo, ki je skupna vsem taksonom nad to razvejitvijo; 3. posamezna sinapomorfija lahko nastopi samo enkrat, izjema so identični znaki, ki so se razvili neodvisno pri različnih taksonih in predstavljajo evlucijski paralelizem.

V okviru kladističnega pristopa je v nadaljevanju predstavljena metoda največje varčnosti, ki jo najpogosteje uporabimo za gradnjo kladograma, njena uporaba je prikazana v Primeru 13.8.

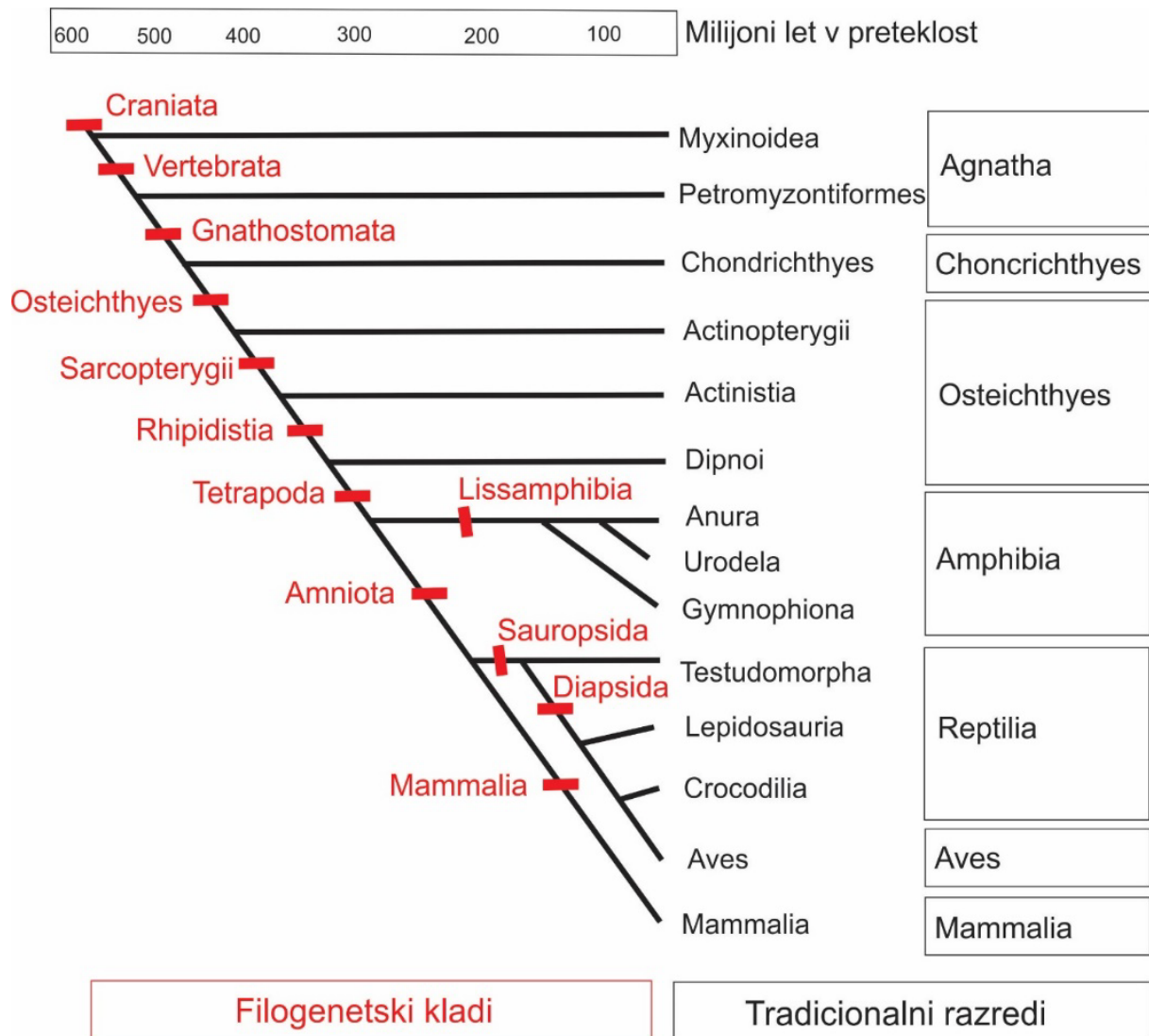
Monofiletske skupine v kladističnem pristopu so hierarhično organizirane in ugnezdene ena v drugo (Slika 13.20). Vsaka podstopnja v hierarhiji je podskupina nadrejene stopnje. Poglejmo primer klasifikacije izbranega taksona v kladističnem konceptu. Poimenovanje oz. terminologija posameznega taksona izhaja iz monofiletske skupine, ki ji pripada. Npr. beloprski jež je član naslednjih monofiletskih skupin, hierarhija sledi od kraljestva navzdol do vrste in je povzeta po NCBI GenBank (<https://www.ncbi.nlm.nih.gov/>):

Eukaryota, Opisthokonta, Metazoa, Eumetazoa, Bilateria, Deuterostomia, Chordata, Craniata, Vertebrata, Gnathostomata, Teleostomi, Euteleostomi, Sarcopterygii, Dipnotetrapodomorpha, Tetrapoda, Amniota, Mammalia, Theria, Eutheria, Boreoeutheria, Laurasiatheria, Eulipotyphla, Erinaceidae, Erinaceinae, *Erinaceus*, *Erinaceus concolor*.

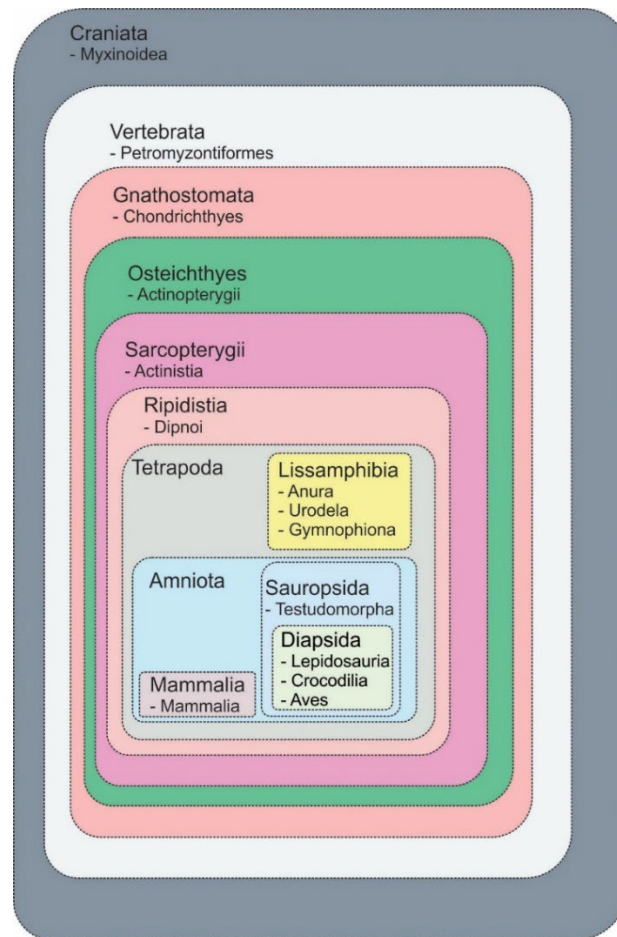
Zadnja dva taksona sta rodovno in vrstno ime beloprsega ježa, ki se pišeta ležeče.



Slika 13.18 Oblikovanje skupin v kladističnem konceptu, skupine so označene z rumenimi poligoni: monofiletska skupina obsega evlucijskega prednika in vse njegove potomce; parafiletska skupina obsega evlucijskega prednika, vendar ne vseh njegovih potomcev (izključevanje posamezne skupine); polifiletska skupina obsega nekaj izbranih taksonov, in ne vključuje skupnega prednika vseh taksonov niti vseh potomcev.



Slika 14.19 Klasifikacija recentnih vretenčarjev po tradicionalnem in kladističnem pristopu. Na desni strani slike (črno besedilo) je predstavljen tradicionalni pristop na levi (rdeče besedilo) pa kladistični pristop k sistematiki vretenčarjev. Z rdečimi prečniki so predstavljeni evlucijski dogodki, ki pomenijo sinapomorfije, skupne izpeljane znake oz. ločnice med naravnimi skupinami. Npr. za monofiletsko skupino Vertebrata je diagnostičen dogodek razvoj hrbtenice, ki je sestavljena iz vretenc, in vsi taksoni znotraj te skupine so vretenčarji. Za monofiletsko skupino štirinožce Tetrapoda je diagnostična prisotnost dveh parov nog in jo sestavljajo predstavniki ((Lissamphibia + (Anura + Urodela + Gymnophiona)) + (Amniota + (Sauropsida + (Testudomorpha)) + (Diapsida + (Lepidosauria + Crocodilia + Aves)) + (Mammalia + (Mammalia))). V tradicionalni klasifikaciji prepoznavamo sedem razredov vretenčarjev Vertebrata, ki so preživeli do danes (skupina rib oklepnice Placodermi je izumrla in ni predstavljena na drevesu), to so: brezčeljustnice Agnatha, ribe hrustančnice Chondrichthyes, ribe kostnice Osteichthyes, dvoživke Amphibia, plazilci Reptilia, ptiči Aves in sesalci Mammalia. V kladističnem pristopu na taksonomskem nivoju razredov razlikujemo osem monofiletskih kladov vretenčarjev: vretenčarji Vertebrata, čeljustnice Gnathostomata, kostnice Osteichthyes, mesnatoplavutarice Sarcopterygii, rfidisti Rhipidistia, tetrapodi Tetrapoda, amnioti Amniota in sesalci Mammalia. Primerjaj s Sliko 13.20. Na vrhu slike je okvirna časovna skala z milijoni let v preteklost za orientacijo o času evlucijskega dogodka in starosti posamezne monofiletske skupine. Slika je prirejena po Liem in sod. (2001). (Opomba. Od izida učbenika so se dogajale spremembe v taksonomiji vretenčarjev, aktualen kladogram je že nekoliko posodobljen).



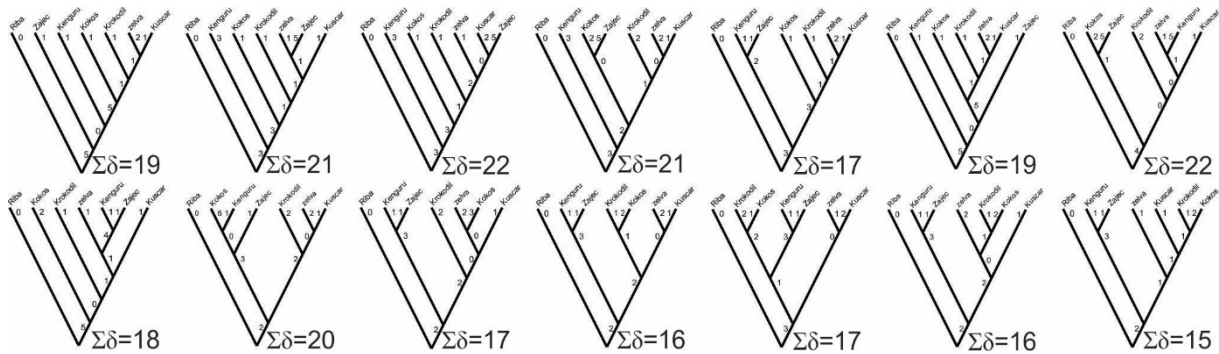
Slika 13.20 Kladistični pristop k filogeniji vretenčarjev. Monofiletske skupine so ugnezdene v hierarhične kategorije. Diagnostična lastnost je sinapomorfni znak. Primerjaj s Sliko 13.19.

13.2.1 Metoda največje varčnosti

Metoda največje varčnosti MP (ang. Maximum Parsimony Method) je pristop, kjer imajo enostavnejše rekonstrukcije prednost pred kompleksnejšimi. Metoda zgradi vsa ali večino možnih filogenetskih dreves. Drevesa, ki razložijo informacijo iz matrike podatkov z manjšim številom evolucijskih dogodkov, imajo prednost pred drevesi, ki za razlago matrike potrebujejo več dogodkov. Topologija **najbolj varčnega drevesa** (ang. Most Parsimonious Tree MPT) je taka, da je vsota vseh substitucij najmanjša ali je število vseh diskretnih fenetskih sprememb najmanjše oz. je skupna dolžina vej najmanjša. Vsaka posamezna substitucija ali sprememba v fenetskih znakih prinese eno enoto k dolžini veje. Dolžina drevesa je vsota dolžin vseh vej (Maximum parsimony 2023).

Gradnja in izbira optimalnega drevesa vključuje izračun dreves vseh možnih topologij in izračun števila evolucijskih dogodkov (sprememb), dejansko dolžine vej (Slika 13.21). Drevo z najkrajšo skupno dolžino vej oz. najmanjšim številom sprememb je optimalna rešitev oz. najbolj varčno drevo.

Prikaz filogenetske rekonstrukcije po metodi največje varčnosti na fenetskih znakih je predstavljen v Primeru 13.8.



Slika 13.21 Naključno izbranih 14 filogenetskih dreves izmed 945 dreves s sedmimi taksoni, ki jih je zgradil izčrpani algoritem po metodi največje varčnosti. Ob vsaki veji je navedena njena dolžina oz. število sprememb znakov. Ob vsakem drevesu je podana skupna dolžina vej $\Sigma\delta$. Najkrajša skupna dolžina vej je $\Sigma\delta = 15$ in to je najbolj varčno filogenetsko drevo za gradnjo katerega je potrebnih najmanjše število sprememb.

Osnovni pogoji in kriteriji za analizo varčnosti. Vhodni podatki so organizirani v pravokotni matriki OTU x znaki. Morfološki in drugi fenetski znaki imajo binarno vrednost 0/1 ali so v več kategoričnih stanjih, molekularni znaki imajo več stanj: A, C, T in G. Cilj analize je poiskati drevo s takšno topologijo, ki razloži stanje znakov obravnavanih taksonov z najmanjšim številom sprememb znakov. V vsakem koraku izvajanja algoritma se znak lahko spremeni. Evolucijskim dogodkom lahko definiramo zaporedje spreminjanja ali dovolimo poljubno zaporedje sprememb, prav tako lahko dovolimo reverzibilne spremembe ali pa ne. Znake lahko obtežimo ali pustimo v osnovni obliki. Glede na tip podatkov ter zaporedje spreminjanja in reverzibilnost sprememb poznamo več metodoloških variant izbire drevesa največje varčnosti: Fitcheva, Wagnerjeva, Dollojeva, Camin-Sokalova in popolna varčnost.

Koreninjenje drevesa izvedemo z vključitvijo zunanje skupine (ang. Outgroup). To je takson, ki je blizu obravnavani skupini organizmov, ni pa član monofiletske skupine. Z drugimi besedami, zunanja skupina nima sinapomorfnihih znakov in zato lahko predstavlja izhodiščno stanje, s katerim se primerjajo vsi ostali taksoni. Obenem ustvarimo polarnost (usmerjenost) znakov v kladistični analizi (Outgroup 2023).

Algoritmi za iskanje optimalnega drevesa. V metodi največje varčnosti algoritem zgradi vsa možna drevesa in nato izmed množice dreves izbere tisto ali tista, ki izkazujejo topologijo z najmanj sprememb znakov. Število dreves raste eksponentno s številom taksonov, pri vključitvi 10 taksonov število možnih dreves naraste na 34 mil. koreninjenih

oz. 2 mil. nekoreninjenih dreves. Razvitih je več algoritmov za iskanje optimalnega drevesa: izčrpn pristop, pristop z razvejitvijo in omejevanjem ter hevristični pristop. **Izčrpn algoritem** (ang. Exhaustive) zgradi in oceni vsa možna drevesa in izbere drevo z najkrajšo skupno dolžino vej. Metoda je omejena na približno 12 taksonov, ki lahko oceni do 600 milijonov dreves. **Algoritem z razvejitvijo in omejevanjem** (ang. Branch-and-bound) je ustrezna rešitev pri iskanju drevesa z več kot 11 taksoni. Algoritem gradi samo posamezne odseke drevesa. Optimalno zgrajenih odsekov ne spreminja več, in s tem bistveno zmanjša skupno število dreves. **Hevrističnih algoritmov** (ang. Heuristic) je več: **zamenjava najbližjega soseda** NNI (ang. Nearest Neighbour Interchange), **obrezovanje in cepljenje** SPR (ang. Subtree Pruning and Regrafting) in **razrezovanje in ponovno združevanje** TBR (ang. Tree Bisection and Reconnection).



PRIMER 13.8

Filogenetska rekonstrukcija amniotov –
metoda največje varčnosti

Naš namen je podoben kot pri prejšnjih pristopih v Primerih 13.6 in 13.7, predstaviti filogenijo amniotov. V tem primeru želimo filogenijo predstaviti v obliki kladograma. Vhodni podatki so organizirani v pravokotni matriki 7 OTU x 14 znakov (Preglednica 13.14). Drevo smo izdelali po dveh algoritmih z izčrpnim in z razveji in omeji, oboje po metodi največje varčnosti.

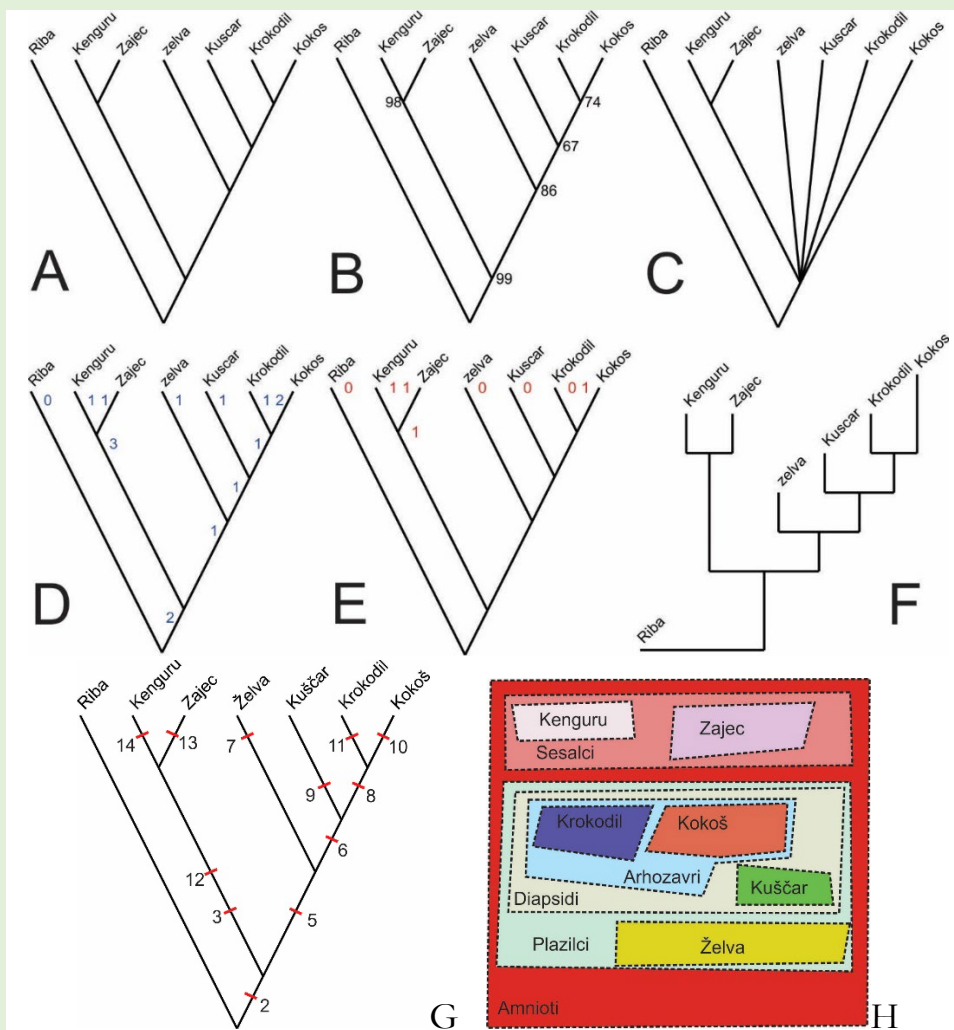
1. V izčrpnem algoritmu je bilo v ocenjevanje vključenih 945 dreves. Za optimalno rešitev je algoritem izbral samo eno drevo z dolžino vej 15 (Slika 13.22.D). Dreves z dolžino vej 16 je bilo 7 in z dolžino 17 jih je bilo 36. Na kladogramu so prikazane dolžine posamezne veje (Slika 13.22.D) in za ponazoritev smo označili taksoni, ki izkazujejo homeotermijo (Slika 13.22.E). Filogenetski odnosi med taksoni so predstavljeni tudi v obliki fenograma (Slika 13.22.F), v katerem dolžina veje predstavlja število morfoloških sprememb.

2. V algoritmu razvejaj in omeji smo definirali ocenjevanje dreves s skupno dolžino do 16. V ocenjevanje je bilo vključenih 186 dreves, samo 7 dreves je ustrezalo kriteriju dolžine drevesa do 16. Ocenjevanje podpore posamezni razvejitvi smo izvedli z metodo bootstrap s 1.000 samovzorčenji. Najvišjo podporo z 99 % ima razvejitev med sesalci in plazilci. Prav tako visoko podporo 98 % ima tudi razvejitev med kengurujem in zajcem, vse preostale razvejitve imajo podporo višjo od 50 % vendar manj od 90 % (Slika 13.22.B). Drevo, zgrajeno po strogem soglasju, izkazuje politomijo, iz katere izhaja pet vej, vsaka skupina plazilcev, ptičev in sesalcev s svojo vejo (Slika 13.22.C). Drevo večinskega soglasja vrne drevo brez politomij (Slika 13.22.A).

Newick oblika zapisa za drevo, zgrajeno po metodi večinskega soglasja in predstavljeno na Sliki 13.22.A ima obliko: utree= (Riba,((Kenguru,Zajec),(Želva,(Kuščar,(Krokodil, Kokoš))));

Filogenetsko analizo smo izvedli s programom PAST 4 (Hammer in sod. 2001).

Drevo, zgrajeno po metodi večinskega soglasja, izkazuje filogenetske odnose med šestimi taksoni amniotov v skladu z današnjim poznavanjem poteka evolucije. Z drugimi besedami, izbrani morfološki znaki omogočajo dobro filogenetsko rekonstrukcijo, ki je skladna z genetskimi ugotovitvami.



Slika 13.22 Drevesa s filogenijo amniotov in z ribo kot zunanjo skupino, narejena z metodo največje varčnosti. A Drevo po metodi večinskega soglasja, topologija drevesa podaja hipotezo o evlucijskih odnosih med amnioti. B Drevo, zgrajeno po algoritmu razvejaj in omeji, ob posameznem razvejišču je podana podpora tej topologiji. C Drevo po metodi strogega soglasja, razvejitev z nezadostno podporo je ostalo nerazrešeno in izkazuje politomijo. D Drevo, zgrajeno po izčrpnem algoritmu, ob vejah so navedene dolžine vej, skupna dolžina vej je 15.

E Drevo, zgrajeno po izčrpnem algoritmu, ob vejah so označeni taksoni, ki izkazujejo homeotermijo, to je znak, ki se neodvisno pojavi pri sesalcih in ptičih. F. Drevo, v katerem so dolžine vej korespondenčne številu sprememb. G. Kladogram z označenimi sinapomorfijami: 2-Amnion, 3-Dlaka, 5-Ena slušna kost, 6-Diapsidna lobanja, 7-Oklep, 8-Okno v mandibuli, 9-Kloaka, 10-Perje, 11-Panizzijeva odprtina, 12-Euviviparni, 13-Placenta, 14-Marzupij. H. Kladističen prikaz s hierarhijo ugnезdenih monofiletskih skupin.



Zaključki

- Filofenetika je veda o evolucijskih odnosih znotraj skupin in med njimi na podlagi morfoloških (fenetskih) znakov.
- Evolucijske odnose analiziramo z metodami tradicionalne morfometrije in geometrijske morfometrije, z namenom izdelave filogenetskega drevesa.
- Morfološki znaki vsebujejo filogenetsko in fenetsko informacijo. Fenetska informacija je povezana z modifikacijami v velikosti in oblike telesa kot odziv na biotske in abiotske dejavnike. Vir variabilnosti so tudi spolni dimorfizem, starostna variabilnost in alometrija. Dodatni vidiki variabilnosti zakrivajo filogenetsko informacijo v morfoloških znakih.
- V fenetskem pristopu analiziramo podobnost med organizmi, iz podobnosti sklepamo na evolucijske odnose med njimi. Rezultat filogenetske rekonstrukcije prikažemo s fenetskim drevesom, ki ponazarja morfološko podobnost med predstavniki evolucijske linije.
- Morfološki znaki so kvalitativni (prisoten/odsoten ali v več stanjih) in kvantitativni (diskretni in zvezni). Podatke organiziramo v pravokotno matriko, kjer so taksoni v vrsticah in spremenljivke v stolpcih.
- Fenetsko drevo gradimo po algoritmičnem pristopu, kjer drevo izdelamo na morfoloških razdaljah, po metodi združevanja sosedov ali metodi UPGMA.
- V kladističnem pristopu prepoznavamo naravne skupine organizmov, to je monofiletska skupina imenovana klad.
- Klad sestavljajo zadnji skupni prednik in vsi njegovi evolucijski potomci. Prepoznamo jih po tem, da si delijo skupno lastnost, ki izvira iz zadnjega skupnega prednika, pri sodobnih vrstah je lahko preoblikovana – sinapomorfni znak.
- Sinapomorfni morfološki znakom moramo potrditi homologijo. To pomeni, da znak izvira iz zadnjega skupnega prednika, izhaja iz istega zarodnega tkiva in ima enotno organizacijo. Oblika in funkcija morfološkega znaka lahko variirata od zelo podobne do zelo različne.
- Odnosi med monofiletskimi skupinami v kladističnem pristopu so prikazani s kladogramom. Z njim je vzpostavljena hierarhija monofiletskih skupin, hierarhično podrejene skupine so ugnezdene v višje skupine, za vsako skupino je definirana sinapomorfna lastnost.
- Kladogram gradimo po načelu varčnosti. Zgradimo ga z iskanjem optimalnega drevesa, to je z metodo največje varčnosti. Metoda zgradi vsa možna drevesa in na podlagi kriterija najmanjšega števila evolucijskih sprememb izberemo optimalno drevo.

Reference

- Apomorphy and synapomorphy. 2023. Wikipedia. Pridobljeno 28. 1. 2023, iz https://en.wikipedia.org/wiki/Apomorphy_and_synapomorphy.
- Bookstein F.L. 1991. Morphometric tools for landmark data. Cambridge University Press. Cambridge.
- Cladistic. 2023. Wikipedia. Pridobljeno 28. 1. 2023, iz <https://en.wikipedia.org/wiki/Cladistics>.
- Dray S., Dufour A.-B., Thioulouse J. 2020. ade4. Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences. R package version 2.5-6. <https://cran.r-project.org/web/packages/ade4/index.html>.
- Evolution. 2023. Wikipedia. Pridobljeno 28. 1. 2023, iz <https://en.wikipedia.org/wiki/Evolution>.
- Forey P. 2023. Cladistics for Palaeontologists. The Natural History Museum. London. Pridobljeno 28. 1. 2023, iz <https://www.palass.org/publications/newsletter/cladistics-palaeontologists>.
- Freeman S., Herron C.J. 2004. Evolutionary Analysis. Pearson Education Inc. Upper Saddle River, NJ.
- Goslee S.C., Urban D.L. 2007. The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, 22(7): 1–19. DOI:10.18637/jss.v022.i07.
- Hall G.B. 2011. Phylogenetic Trees Made Easy. A How-to Manual. Sinauer Associates. Sunderland.
- Hammer Ø., Harper D.A.T., Ryan P.D. 2001. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica*, 4(1): 9pp.
- Homology. 2023. Wikipedia. Pridobljeno 28. 1. 2023, iz [https://en.wikipedia.org/wiki/Homology_\(biology\)](https://en.wikipedia.org/wiki/Homology_(biology)).
- Kassambara A., Mundt F. 2020. factoextra: Extract and Visualize the Results of Multivariate Data Analyses_. R package version 1.0.7, <<https://CRAN.R-project.org/package=factoextra>>.
- Klenovšek T. 2014. Priročnik za uporabo geometrijske morfometrije v biologiji. Fakulteta za naravoslovje in matematiko. Maribor.
- Klingenberg C.P. 2011. MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources*, 11: 353–357.
- Legendre P., Legendre L. 2012. Numerical Ecology. Elsevier. Amsterdam.
- Lemey P., Salemi M., Vandamme A.-M. (eds.). 2009. The Phylogenetic Handbook: a Practical Approach to Phylogenetic Analysis and Hypothesis Testing. Cambridge University Press. Cambridge.
- Maximum parsimony. 2023. Wikipedia. Pridobljeno 28. 1. 2023, iz [https://en.wikipedia.org/wiki/Maximum_parsimony_\(phylogenetics\)](https://en.wikipedia.org/wiki/Maximum_parsimony_(phylogenetics)).
- Monophyly. 2023. Wikipedia. Pridobljeno 28. 1. 2023, iz <https://en.wikipedia.org/wiki/Monophyly>.
- Nei M., Kumar S. 2000. Molecular Evolution and Phylogenetics. Oxford University Press. New York.
- Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., Wagner H. 2019. vegan: Community Ecology Package. R package version 2.5-6. <https://CRAN.R-project.org/package=vegan>.
- Outgroup. 2023. Wikipedia. Pridobljeno 28. 1. 2023, iz [https://en.wikipedia.org/wiki/Outgroup_\(cladistics\)](https://en.wikipedia.org/wiki/Outgroup_(cladistics)).
- Phenetics. 2023. Wikipedia. Pridobljeno 28. 1. 2023, iz <https://en.wikipedia.org/wiki/Phenetics>.
- Phylogenesis. 2023. Wikipedia. Pridobljeno 28. 1. 2023, iz <https://en.wikipedia.org/wiki/Phylogenesis>.
- Phylogenetic tree. 2023. Wikipedia. Pridobljeno 28. 1. 2023, iz https://en.wikipedia.org/wiki/Phylogenetic_tree.
- Phylogenetics. 2023. Wikipedia. Pridobljeno 28. 1. 2023, iz <https://en.wikipedia.org/wiki/Phylogenetics>.
- R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Sneath A.H.P., Sokal R.R. 1973. Numerical Taxonomy. The Principles and Practice of Numerical Classification. W.H. Freeman and Company. San Francisco.

14. poglavje



FILOGENETIKA

Povzetek

14. poglavje: Filogenetika

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

Filogenetika je veda o evoluciji živih bitij in o evlucijskih odnosih med skupinami in znotraj njih. Filogenetika je tudi metodološki pristop k filogenetskim rekonstrukcijam, s katerim analiziramo zaporedja nukleotidov ali zaporedja aminokislin. Variabilnost v zaporedju in frekvenci nukleotidov oz. aminokislin razlagamo s statističnim modeliranjem in metodami verjetnosti. Evlucijske odnose analiziramo z metodami računske filogenetike in filogenetskega sklepanja, z namenom izdelave filogenetskega drevesa. Drevo je graf, s katerim ponazorimo verjeten potek evolucije.

Filogenetske analize na dednini temeljijo na zaporedju nukleotidov in primerjavah med nukleotidnimi zaporedji. Algoritmi, s katerimi rekonstruiramo verjeten potek evolucije, so razviti glede na zaporedje nukleotidnih baz, njih frekvenc in glede na stopnjo substitucije. Algoritmi omogočajo izračun genetskih razdalj glede na število razlik v zaporedju primerjanih nukleotidov npr. Jukes-Cantorjeva razdalja. Filogenetske odnose med taksoni predstavimo s filogenetskim drevesom. Gradimo ga v algoritmičnem pristopu, kjer drevo izdelamo na podlagi genetskih razdalj z metodo združevanja sosedov ali z metodo UPGMA. Metoda največjega verjetja temelji na izbiri substitucijske modela, s katerim izračunamo verjetnosti za potek substitucij. V analizi spreminjamo število parametrov in vrednosti parametrov, ki najbolj razložijo opazovana zaporedja. Prav tako na verjetnosti temelji Bayesovo sklepanje, s katerim rekonstruiramo potek evolucije. Bayesov pristop temelji na posteriorni verjetnosti, kjer najprej zastavimo priori verjetnost in jo primerjamo z empiričnimi meritvami.

Topologijo drevesa testiramo s samovzorčenjem, v katerem z metodo bootstrap preverjamo podporo razvejitvam. Če razvejitve ne izkazujejo zadostne podpore, zgradimo drevo soglasja, v katerem nezadostno podprte razvejitve ostanejo nerazrešene in izkazujejo politomijo.

Glede na informacijo, ki je zajeta v filogenetskem drevesu, razlikujemo med različnimi tipi dreves: radialno drevo ni koreninjeno in je aditivno, dolžine vej ponazarjajo število mutacij; koreninjeno in ultrametrično drevo predstavlja hierarhični potek divergentnih cepitev v nove taksone in ga lahko uskladimo s časovno skalo; kladogram temelji na oblikovanju monofiletskih skupin in ima označene sinapomorfije.

S testiranjem filogenetskega signala primerjamo filogenetske rekonstrukcije na morfoloških znakih z rekonstrukcijami na nukleotidnih zaporedjih.

Nukleotidna zaporedja so dostopna v genskih bankah npr. NCBI – National Center for Biotechnology Information, ENA – European Nucleotide Archive, EMBL – European Bioinformatics Institute, idr.

Filogenetika (sin. filogenija, ang. Phylogenetics) je veda o evoluciji živih bitij in njihovih evolucijskih odnosih znotraj taksonomskih skupin in med njimi. Evolucijske odnose analiziramo z metodami **računske filogenetike** (ang. Computational Phylogenetics) in **filogenetskega sklepanja** (ang. Phylogenetic Inference), katerega rezultat je **filogenetska rekonstrukcija** (ang. Phylogenetic Reconstruction). Po navadi jo prikažemo z izrisom **filogenetskega drevesa** (sin. genealoško drevo, ang. Phylogenetic Tree) (Computational phylogenetics 2023, Phylogenetic Inference 2022, Phylogenesis 2023, Phylogenetics 2023, Phylogenetic Tree 2023).

Filogenetika je tudi metodološki pristop k filogenetskim rekonstrukcijam, v katerem analiziramo variabilnost v zaporedju nukleotidov ali zaporedju aminokislin. Variabilnost v zaporedju in frekvenci nukleotidov oz. aminokislin razlagamo s statističnim modeliranjem in metodami verjetnosti (Phylogenetics 2023). Napredek molekularnih metod in bioinformatike v preučevanju nukleotidnih zaporedij je odprl nove perspektive v taksonomiji in še bolj v izdelavi in vrednotenju filogenij oz. filogenetskih rekonstrukcij. Tako je v sedanjem času to že samostojno raziskovalno področje z imenom **filogenomika** (ang. Phylogenomics).

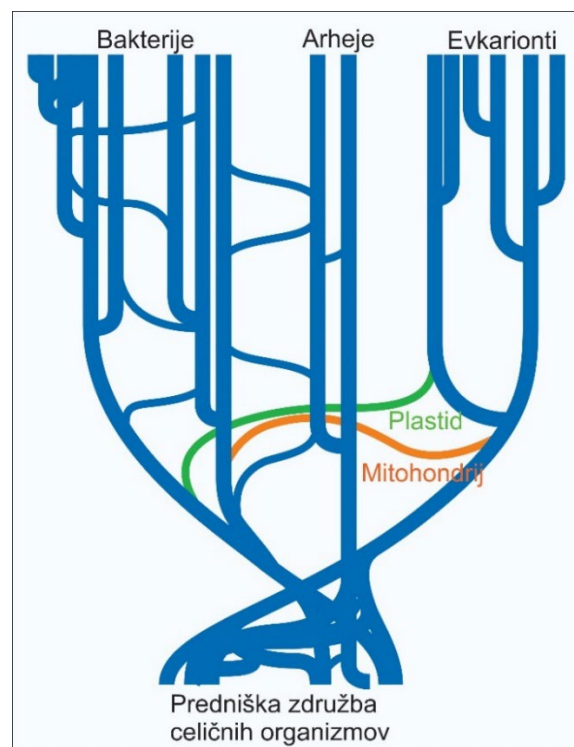
Filogenija (ang. Phylogeny) je evolucijska zgodovina vrste ali skupine organizmov (npr. filogenija človečnjakov ali filogenija placentálnih sesalcev). Filogenetska drevesa, s katerimi so prikazani evolucijski odnosi med taksoni (filogenije), predstavljajo novo izhodišče in začetni niz podatkov za vrednotenje biodiverzitete. Pristop na podlagi filogenetskih odnosov med člani življenjske združbe oz. taksoni, ki sestavljajo skupino organizmov, je samostojen vidik biodiverzitete; imenovan **filogenetska pestrost** (ang. Phylogenetic Diversity). Vsebine filogenetske pestrosti so obravnavane v samostojnem Poglavju 6.

Filogenetsko drevo je graf v obliki drevesa (filogram), ki ponazarja evolucijske odnose med taksoni. Listi na filogenetskem drevesu so konkretni osebki (dejansko njihova dednina), ki so predstavniki evolucijske linije. Razvejitve so evolucijski dogodki in predstavljajo kladogenezo. Veje povezujejo korenino, razvejitve in liste med seboj in predstavljajo filetske linije v evolucijskem procesu. Filogram koreninimo po metodi zunanje skupine (Phylogenetic Tree 2023). Podrobnejša predstavitev filogenetskih dreves je v Poglavju 14.5.

Drevo življenja (ang. Tree of Life) je razširjen arhetip, skupen številnim religijam, mitološkim zgodbam in pripovedkam različnih kultur. Poenostavimo lahko, da drevo življenja predstavlja povezavo med vsemi živimi bitji. Zgodnji biološki pogledi na drevo življenja so izhajali iz opisovanja organizmov po njihovem fizičnem videzu in metabolizmu. Moderne molekularne metode in razvoj genomike sta odprla nov vpogled v razumevanje poteka filogenije živih bitij in evolucijske odnose med njimi. **Zadnji univerzalni skupni prednik LUCA** (ang. Last Universal Common Ancestor) je

konstrukt – hipotetični organizem, iz katerega izvajamo vsa živa bitja, ki sestavljajo biodiverzitetno na Zemlji. To so vsi organizmi in vsi so zgrajeni iz celic oz. so enocelični. Izvor virusov ni znan, vendar imajo isto genetsko kodo kot celični organizmi. LUCA ni prvo živo bitje na Zemlji, ampak zadnja oblika prednika, iz katerega so izšle vse današnje oblike življenja. Domnevno se je razvil na oceanskem dnu ob izvirih vroče vode v prisotnosti magme pred okrog 4 milijardami let. Sodobne raziskave podpirajo hipotezo o dveh začetniških domenah, to sta bili bakterije (Bacteria) in arheje (Archaea). Evkarionti (Eucaryota) so se domnevno razvili iz arhej. Filogeneza ni potekala samo z dihonomno delitvijo, enakovredno evlucijsko vlogo sta imela **horizontalni prenos genov HGT** (ang. Horizontal Gene Transfer) in **endosimbioza** (ang. Endosymbiosis). Sodobno drevo življenja ni več dihonomno razraslo, temveč oblikuje mrežo medsebojno povezanih vej (Slika 14.1) (LUCA 2023).

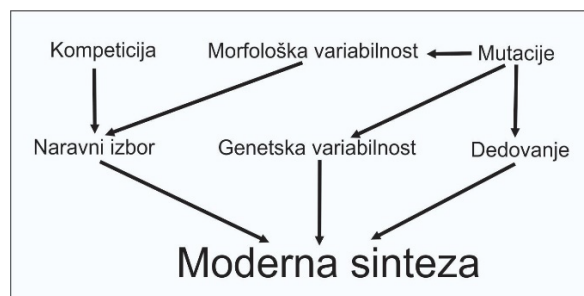
Drevo življenja oz. The Tree of Life Web Project <http://tolweb.org/tree.html> je zbirka informacij in znanj o biodiverziteti, ki jo sestavlja skupina prostovoljcev. Njihov cilj je vzpostavljati zbirke podatkov s fotografijami, besedili, filogenetskimi odnosi in drugimi informacijami za vse žive in izumrle organizme. Povezave med spletnimi stranmi sledijo vzorcu topologije filogenetskega razvoja.



Slika 14.1 Drevo življenja se začneja s predniško združbo celičnih organizmov, vključno z LUCA. Dihotomne razvejitve vej ponazarjajo speciacijo, prečne povezave ponazarjajo horizontalni prenos genov med filetskimi linijami. Barvni povezavi predstavljata endosimbiozo. Z vključevanjem prokariotov razlagamo izvor mitohondrijev v celicah evkariontov in kloroplastov pri rastlinah (povzeto in prirejeno iz LUCA 2023). Drevo sta leta 2005 predlagala Smets F. B. in Barkay T. 2005. Horizontal gene transfer: perspectives at a crossroads of scientific disciplines. *Nature Reviews Microbiology*, 3 (9): 675–678.

Evolucijski proces. Poenostavimo lahko, da so mutacije vir genetske in morfološke variabilnosti in se dedujejo. Variabilnost organizmov in kompeticija za naravne dobrine sta vzrok za naravni izbor. Vse navedene lastnosti so povezane v razlago evolucijskega procesa imenovanega **moderna sinteza** (ang. Modern Synthesis) (Slika 14.2) (Modern Synthesis 2023). **Mutacija** je trajna sprememba v zaporedju nukleotidov v genomu organizma (Mutation 2023). S citološkega vidika mutacije nastajajo v somatskih, spolnih in zarodnih celicah. V kontekstu filogenetskih rekonstrukcij nas zanimajo predvsem mutacije v spolnih in zarodnih celicah, ki se dedujejo in vplivajo na evolucijski proces. Mutacije lahko delimo na genske mutacije, kromosomske mutacije, genomske mutacije in enonukleotidni polimorfizem. Kromosomske in genomske mutacije so t.i. »makromutacije«. V Poglavju 14.1 je nanizanih nekaj parametrov genskih mutacij (sin. mikromutacije, točkovne mutacije). Nekatere od mutacij so v danih ekoloških razmerah lahko ugodne in za njihove nosilce (fenotipe) pomenijo adaptacijo, ki poveča reprodukcijo ali preživetje. Nekatere druge mutacije so lahko za njihove nosilce neugodne in zmanjšajo možnosti za preživetje ter vodijo k izumrtju. Posebna oblika mutacij je takšna, da le-te ne vplivajo na fenotip organizma in so nevtralne.

Vsa današnja živa bitja so šla skozi proces evolucije. Preživele so tiste populacije, pri katerih so mutacije pomenile ustrezen odziv oz. prilagoditve na abiotske in biotske razmere. Drugače povedano, vsa današnja živa bitja so prestala preverbo selekcije, nenehno spreminjanje jim je omogočilo preživetje v spreminjajočih se habitatih, ta proces se imenuje **naravni izbor** (sin. naravno odbiranje, naravno izbiranje, ang. Natural Selection) (Evolution 2023, Natural selection 2023).



Slika 14.2 Združevanje bioloških fenomenov in idej o evolucijskem procesu
(povzeto in prirejeno po Modern Synthesis 2023).

Posebej moramo poudariti, da proces spreminjanja od enostavnejših h kompleksnejšim oblikam življenja ni usmerjen. Kljub temu da so spremembe v genetskem zapisu majhne, se v času kopičijo, kar vodi k počasnemu spreminjanju genotipa in posledično fenotipa. Filogenijo lahko razumemo kot nenehno spreminjanje filetske linije v prostoru in času. Ključnega pomena za ohranitev genetske informacije v času in prostoru sta **razmnoževanje** in **genski pretok** (ang. Gene Flow), ki zagotavljata povezavo med evolucijsko starejšimi (predniki) in evolucijsko mlajšimi (potomci) naravnimi skupinami,

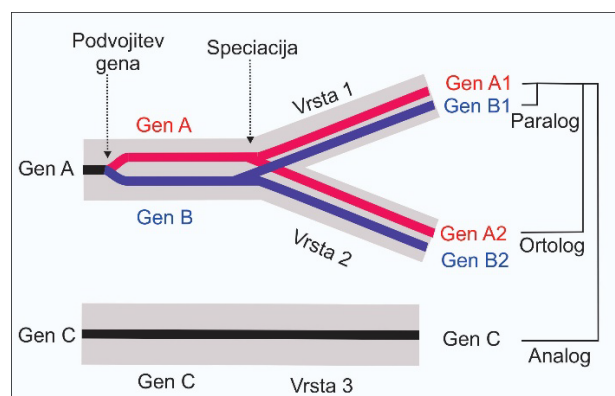
vendar ne le razmnoževanje, temveč razmnoževanje z adaptacijami in naravni izbor. Pri tem razmnoževanje zagotavlja kontinuirano povezavo med skupinami v času in prostoru ter prenos genetske informacije iz prednikov na potomce. Mutacije in naravni izbor skupaj z morfološko variabilnostjo omogočajo pestrost življenjskih oblik, in s tem biodiverzitetu.

Speciacija (ang. Speciation) je evolucijski proces, v katerem se populacija razvije v novo vrsto organizma. Nova vrsta nastane s cepitvijo populacije v dihodontnem ali z njenim preoblikovanjem v anageneznem procesu. Reprodukativno ločena populacija se lahko neodvisno spreminja v prostoru in času. Ko posamezna populacija doseže stopnjo reprodukativne izolacije ali nekompatibilnosti z drugimi populacijami, govorimo o novi vrsti. Nastanek nove vrste na filogenetskem drevesu prikažemo kot hierarhičen odnos med prednikom in potomcem, kjer se predniška veja dihodontno razveja v dve veji potomcev, kar imenujemo **kladogeneza**. Slednje istočasno pomeni, da sta oba novonastala potomca med seboj v odnosu potomec-potomec. V relaciji do skupnega prednika sta oba v odnosu prednik-potomec. V filogenetskih rekonstrukcijah ob dihodontiji uporabljamo tudi politomijo. V drevesu s **politomijo** dejansko zaporedje nastanka potomcev ni znano, zato predvidevamo istočasni nastanek treh ali več evolucijskih potomcev, drevo s politomijo ni razrešeno. V primeru **anageneze** gre za evolucijsko zamenjavo ene vrste z drugo znotraj iste filetske linije in brez razvejitve. Evolucijska sprememba poteka v času in znotraj iste filetske linije. Njihove populacije iz časovno različnih obdobj prepoznamo kot različne vrste. Starejša skupina je do mlajše skupine v odnosu prednik-potomec. Na filogenetskem drevesu proces anageneze predstavljajo zunanje veje (Speciation 2023).

Sistematska biologija. Razvoj in napredek na področju filogenetike in kladistike je spodbudno vplival na razvoj celotne sistematske biologije (npr. sistematske zoologije vretenčarjev). Sistematiko danes razumemo kot krovni termin, ki vključuje tri discipline. Prva obsega **opisovanje vrst** (ang. Species Description) in **prepoznavanje vrst** (ang. Identification). Druga disciplina obsega **binarno poimenovanje vrst** (ang. Binomial Nomenclature) in **taksonomijo** (ang. Taxonomy), kar pomeni razvrščanje vrst v taksonomske kategorije (primerjaj s Poglavjem 12 Numerična taksonomija). Tretja disciplina zajema evolucijske odnose med taksoni, to je **filogenetika**. Sistematiki se danes bolj kot z razvrščanjem organizmov ukvarjamo z biodiverzitetu, torej s proučevanjem pestrosti živega sveta. Poslanstvo sistematike vključuje prepoznavanje naravnih oz. monofiletskih skupin. Razlage o evolucijskih odnosih znotraj naravnih skupin in med njimi iščemo s filogenetskimi raziskavami. Filogenetski koncept razmišljanja nas je pripeljal do spoznanja, da živa bitja niso razporejena v lestvico vse naprednejših razvojnih oblik. Organizmov ne razvrščamo na »višje« in »nižje«, temveč govorimo o pestrosti življenjskih oblik (Systematics 2023).

Filogenetska rekonstrukcija je hipoteza o evolucijskih odnosih med sorodniki glede na njihov izvor iz zadnjega skupnega prednika. Rekonstrukcije izvajamo v dveh ključnih korakih: 1. prvi korak je identifikacija in mapiranje **homolognih genov** (sin. molekularni markerji, molekularni znaki, nukleotidno zaporedje) in 2. drugi korak je rekonstrukcija evolucijske zgodovine teh homolognih genov in njena predstavitev v obliki drevesa. Z metodami računske filogenetike in filogenetskega sklepanja izvajamo filogenetske rekonstrukcije na dveh tipih znakov – na molekularnih in na morfoloških. V tem poglavju so predstavljeni filogenetski pristopi na molekularnih znakih, pristop na morfoloških znakih je predstavljen v Poglavju 13.

Homologni geni (sin. homologne sekvence, ang. Sequence Homology) imajo skupni evolucijski izvor iz zadnjega skupnega prednika. Razlikujemo tri mehanizme, v katerih nastajajo homologni geni, to so: 1. speciacija, v kateri nastajajo ortologni geni (sin. ortologi), 2. podvajanje genov, v kateri nastajajo paralogni geni (sin. paralogi) in 3. horizontalni prenos genov, v katerem nastajajo ksenologni geni (sin. ksenologi). Ortologna zaporedja izvirajo iz istega zaporedja zadnjega skupnega prednika, nastala so s speciacijo in pri različnih vrstah opravljajo enako funkcijo. Pri vrstah, ki sta nastali iz skupnega prednika, se ista gena imenujeta ortologa (Slika 14.3). Paralogna zaporedja se nanašajo na nastanek podvojenega gena v zadnjem skupnem predniku obravnavanih vrst. Podvojeni gen se potem v procesu speciacije prenese v dve novonastali vrsti in pri različnih vrstah lahko opravlja različno funkcijo. Pri vrstah, ki sta podedovali kopiji istega gena, se kopiji imenujeta paraloga (Slika 14.3). Za filogenetske rekonstrukcije so pomembni predvsem ortologi (Sequence homology 2023).



Slika 14.3 Predniški gen A se je podvojil in nastala sta paraloga (gen A in B). S speciacijo pri dveh potomcih nastaneta ortologa (gen A1 in A2). Obe novonastali vrsti vsebujeta tudi paraloga (gen A1 in B1 ter gen A2 in B2). V ločeni filetski liniji ima gen C enako funkcijo, vendar ima drugačen evolucijski izvor, zato je analog. Prirejeno po Sequence homology (2023).

Filogenetika vs. fenetika. Morfološki znaki vsebujejo fenetsko informacijo, to je informacija o telesni organizaciji, obliki in delovanju organizma. Fenotip je opredelitev morfološkega izgleda telesne organizacije in delovanja osebka. Raziskovalnemu področju, kjer ugotavljamo evolucijsko sorodnost na podlagi morfoloških znakov, torej na fenetski

informaciji, pravimo filofenetika (ang. Phylophenetics), predstavljena je v Poglavju 13. V zadnjem obdobju prevladuje filogenetika, to so rekonstrukcije na molekularnih znakih na zaporedjih nukleotidov ali aminokislin. Vse genetske informacije, ki jih vsebuje osebek, so genotip.

Filogenetika vs. kladistika. Kladistični pristop temelji na prepoznavanju zadnjega skupnega prednika in definiranju izvornega stanja izbranega znaka. Znak je morfološki ali molekularni in razložiti moramo njegovo preoblikovanja v stanje, kot ga poznamo pri sodobnih organizmih (sinapomorfijo in homologijo) (glej Poglavje 13.2). Vsi organizmi, ki si delijo ta znak v izvorni ali preoblikovani obliki, so člani te skupine (monofiletska skupina). Skupine so hierarhično oblikovane po načelu ugnezdenosti, vsaka hierarhično nižja skupina je podskupina nadrejene skupine (Slika 13.20). Zato kladograma ne razlagamo kot filogenetsko drevo, temveč kot hierarhijo ugnezdenih skupin. Filogenetski pristop temelji na statistični analizi in analizi verjetnosti, s katerima razlagamo potek evolucije.

V procesu filogenetskega sklepanja in gradnje filogenetskih dreves izhajamo iz načela **varčnosti** (ang. Parsimony) ali načela **verjetja** (ang. Likelihood). Topologijo drevesa gradimo na način, ki vključuje bodisi najmanjše število evlucijskih sprememb (varčnost) bodisi največjo verjetnost porazdelitve lastnosti (verjetje). Raziskovalna praksa v zadnjem desetletju daje absolutno prednost filogenetskim rekonstrukcijam na zaporedju nukleotidov in jih privzemamo za najverjetnejšo hipotezo o poteku evolucije. Resna omejitev nukleotidnega pristopa je pri analiziranju fosilnih ostankov, sploh ko imamo opravka s fosili brez možnosti izolacije fosilne DNK (ang. Ancient DNA, aDNA). V primerih fosilnega materiala, ki nima ohranjene dednine, je pristop z analizo morfoloških znakov edini pristop za iskanja filogenetske informacije.

Na prehodu v 21. tisočletje spremljamo izreden napredek v razvoju tehnike in analitike na dednini oz. na zaporedju nukleotidov v molekulah DNK (deoksiribonukleinska kislina, ang. Deoxyribonucleic acid DNA). Na nukleotidnih zaporedjih temelječe filogenetske analize in z njimi povezano filogenetsko pestrost lahko razdelimo na sedem osnovnih korakov: 1. izbira taksonov in izbira homologne genomske regije, npr. mitohondrijski gen za protein citokrom *b* (MT-CYB) v dihalni verigi pri vrsti *Apodemus flavicollis*; 2. izbira oz. prepoznavanje in pridobivanje nizov homolognih nukleotidnih zaporedij; 3. poravnava zaporedij; 4. izbira substitucijskega modela in izračun genetske razdalje med nukleotidnimi zaporedji; 5. izračun in gradnja ter vrednotenje filogenetskih dreves; 6. vrednotenje filogenetskega signala morfoloških struktur in 7. vrednotenje filogenetske pestrosti. Začetnih korakov 1, 2 in 3 ne obravnavamo; v nadaljevanju je predstavljen analitski del korakov 4, 5 in 6, korak 7 je obravnavan v Poglavju 6 Filogenetska pestrost.

Vsebina tega poglavja je omejena na osnovne pojme iz dinamike mutacij, nujnih za razumevanje substitucijskih modelov in gradnjo filogenetskih dreves. Področje molekularne evolucije je hitro razvijajoča se disciplina, predvsem v luči izjemnega napredka molekularne biologije. Nova odkritja odpirajo nove razlage in konkuriranje med hipotezami, kar se odraža v živahni izmenjavi stališč. Prav tako v tej knjigi ne obravnavamo vsebin molekularne biologije, genetike, evolucijske genetike in mehanizmov evolucije. Poudarek je namenjen razumevanju, kako informacijo iz nukleotidnih zaporedij preoblikujemo v filogenetsko drevo. Slednje predstavlja osnovno informacijo za analizo filogenetske pestrosti, ki je posebej naslovljena v Poglavju 6.

Temeljna literatura, iz katere smo črpali vsebino poglavja, je: Anisimova (2019), Felsenstein (1985), Freeman in Herron (2004), Hall (2011), Lemey in sod. (2009), Liem in sod. (2001), Nei in Kumar (2000), Swofford in Sullivan (2009) in Tamura in sod. (2021).

Izračuni in grafični izdelki so bili narejeni s programskimi orodji: BEAST (Drummond in sod. 2018b), BEAUti (Drummond in sod. 2018a), FigTree (Rambaut 2018), MEGA 11 (Tamura in sod. 2021), TreeAnnotator (Rambaut in Drummond 2018) ter v R okolju (R Core Team 2022) in s paketi: "ade4" (Dray in sod. 2020), »ecodist« (Goslee in Urban 2007), »phangorn« (Schliep 2011), "Picante" (Kembel in sod. 2010) in "vegan" (Oksanen in sod. 2019).

Vsebina poglavja je razdeljena na:

- Substitucijski modeli in genetske razdalje.
- Genske banke.
- Največje verjetje.
- Bayesovo sklepanje.
- Filogenetska drevesa.
- Filogenetski signal.

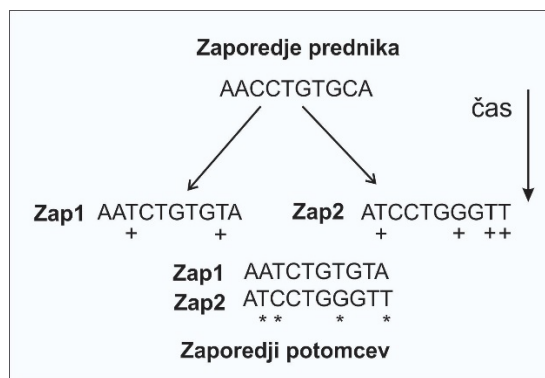
14.1 Substitucijski modeli in genetske razdalje

Multiple poravnana nukleotidna zaporedja enake dolžine lahko vključimo v model, s katerim so opisane evolucijske spremembe v zaporedju nukleotidov. **Nukleotidno zaporedje** (sin. sekvenca, ang. Sequence) je zaporedje nukleotidov v molekuli ali odseku molekule nukleinske kisline. Evolucijski model omogoča bodisi izračun genetskih razdalj med geni oz. med taksoni bodisi izračun filogenetskega drevesa z metodami verjetnosti. **Genetska razdalja** (sin. evolucijska razdalja, ang. Genetic Distance) je vrednost genetske razlike med dvema homolognima zaporedjema, merimo jo med populacijami znotraj ene

vrste ali med vrstami. Dejansko vrednotimo neskladje v nukleotidih po nukleotidnih mestih med dvema zaporedjema.

V tem poglavju so predstavljeni substitucijski modeli in na njih temelječi postopki izračuna in gradnja drevesa po metodah verjetnosti, to sta metoda največjega verjetja in metoda Bayesovega sklepanja BI (ang. Bayesian Inference) (Slika 14.11). Genetske razdalje lahko vključimo v gradnjo filogenetskih dreves po metodi UPGMA ali NJ (Slika 14.11), ta dva postopka gradnje drevesa sta predstavljena v Poglavju 13 (Primeri 13.6 in 13.7).

Zaporedje nukleotidov v molekuli DNK se spreminja zaradi delovanja več evolucijskih sil. Posledice so vidne po določenem času kot sprememba zaporedja ali frekvence nukleotidov (Slika 14.4), kar imenujemo mutacija.



Slika 14.4 Dve nukleotidni zaporedji (Zap1 in Zap2), ki izhajata iz istega skupnega prednika, sta se v času spremenili. V filogenetski rekonstrukciji s pomočjo substitucijskih modelov razlagamo, kakšno bi lahko bilo hipotetično zaporedje nukleotidov pri predniku. S plusi so označena mesta razlikovanja med zaporedji potomcev glede na prednika. Z zvezdicami so označene razlike med obema zaporedjema potomcev.

Genetske mutacije so **substitucija** (zamenjava enega ali nekaj nukleotidov), **insercija** (podaljšanje zaporedja za enega ali nekaj nukleotidov, ki se vrinejo) in **delecija** (skrajšanje zaporedja za enega ali nekaj nukleotidov). V nadaljevanju je nanizanih samo nekaj lastnosti substitucij.

Substitucija kot **usmerjen proces**: enojna substitucija ($A \leftarrow \mathbf{A} \rightarrow C$), večkratna substitucija ($A \leftarrow \mathbf{A} \rightarrow C \rightarrow T$), naključna substitucija ($G \leftarrow \mathbf{A} \rightarrow C$), paralelna substitucija ($C \leftarrow \mathbf{A} \rightarrow C$), konvergentna substitucija ($T \leftarrow \mathbf{A} \rightarrow C \rightarrow A$) in povratna substitucija ($A \leftarrow \mathbf{A} \rightarrow C \rightarrow A$); (v krepkem tisku je izhodiščno stanje, puščica pa predstavlja spremembo iz začetnega v novo stanje).

Dinamika substitucij: enoparametrična substitucija (nukleotidi imajo isto verjetnost zamenjave – Jukes-Cantorjev model), dvoparametrična substitucija (tranzicije so pogostejše kot transverzije).

Kodirajoče substitucije: sinonimna substitucija (sin. tiha substitucija) ne spremeni kodiranja aminokislina, nesinonimna substitucija spremeni kodiranje aminokislina. Sinonimna substitucija nima posledic, zato nosilec ni podvržen selekciji in se mutacije hitreje kopičijo ter se ne odstranijo. Nesinonimna substitucija pomeni posledico za organizem, ki je preverjana v selekcijskem izboru. Nesinonimne so redke, ker maloštevilne prestopajo selekcijski izbor.

Hitrost substitucij V je definirana s številom zamenjav v času in jo zapišemo s formulo:

$$V = \frac{K}{2t},$$

kjer je K povprečno število substitucij na leto in t je čas ločitve dveh zaporedij.

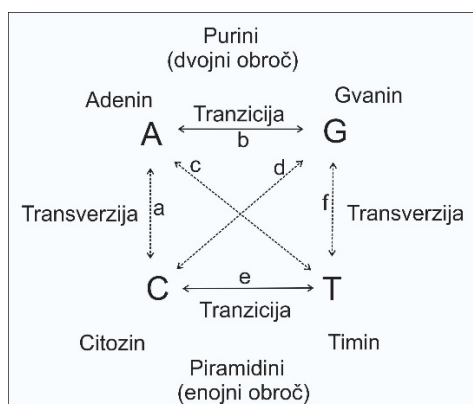
Kodirajoči odseki nukleotidnega zaporedja imajo manj mutacij v primerjavi z nekodirajočimi odseki.

Stopnja metabolizma in generacijski čas: stopnja mutacij je višja pri organizmih z višjo stopnjo metabolizma in pri organizmih s krajšim generacijskim časom.

S stališča filogeneze nas zanimajo trije procesi. Prvič, kako močno se razlikujeta dve homologni zaporedji, ki sta izšli iz zadnjega skupnega prednika in se razvijali neodvisno ena od druge (Slika 14.4). Razliko med dvema zaporedjema ovrednotimo z genetsko razdaljo. V nadaljevanju so predstavljene najpogosteje uporabljene cenilke za izračun genetske razdalje. Drugič, če poznamo oceno števila mutacij v definiranem času in privzamemo, da je njihova frekvenca v času konstantna ali če poznamo njeno dinamiko, potem lahko ocenimo čas razhajanja dveh zaporedij od njunega skupnega prednika. Na tem temeljita določanje evolucijske starosti taksonov in **molekularna ura** (ang. Molecular Clock). Mi se s tem področjem ne bomo ukvarjali. Tretjič, poravnana zaporedja vpeljemo v izgradnjo substitucijskega modela, s katerim razložimo variabilnost v podatkih. Model poišče takšno število in kombinacijo parametrov, ki najbolj zvišajo verjetje za razlago poteka evolucije. Parametri modela omogočajo gradnjo filogenetskega drevesa. Postopek izračuna filogenetskega drevesa po metodi največjega verjetja je predstavljen v Primeru 14.3 in po metodi Bayesovekag sklepanja v Primeru 14.4.

Živalsko molekulo DNK gradijo štirje tipi nukleotidov, razlikujejo se glede na dušikovo bazo, ki je lahko: adenin – A, gvanin – G, citozin – C ali timin – T. Citozin in timin sta zgrajena samo iz enega obroča in sta uvrščena med pirimidine; adenin in gvanin sta zgrajena iz dveh obročev in sta uvrščena med purine. Glede na substitucije med nukleotidi razlikujemo med tranzicijo in transverzijo. Tranzicija je sprememba v nukleotidih molekule DNK, ki nastane s substitucijsko zamenjavo znotraj dveh purinskih nukleotidov ($A \leftrightarrow G$) ali znotraj dveh pirimidinskih ($T \leftrightarrow C$). Transverzija je sprememba, ki nastane s

substitucijsko zamenjavo purinskih s pirimidinskimi nukleotidi ali obratno ($A \leftrightarrow T$, $A \leftrightarrow C$, $T \leftrightarrow G$, $C \leftrightarrow G$) (Slika 14.5). Tranzicije so praviloma pogostejše, njihova frekvenca je okrog 66,6 %, kot transverzijske, ki so zastopane v okrog 33,3 %.



Slika 14.5 Odnos med tranzicijskimi (b in e) in transverzijijskimi substitucijami (a , c , d in f) med nukleotidi A, G, C in T.

Posamezni substitucijski model na različne načine upošteva stopnjo substitucije med nukleotidi in frekvence posameznih nukleotidnih baz (Preglednici 14.1). Predstavitev substitucijskih modelov smo povzeli po Tamura in sod. (2021). Oznake posameznih količin v izračunu genetskih razdalj:

d – tranzicije in transverzije, število različnih nukleotidnih mest.

s – samo tranzicije, število različnih nukleotidnih mest zaradi tranzicij.

v – samo transverzije, število različnih nukleotidnih mest zaradi transverzij.

$R = s/v$ – razmerje med tranzicijami in transverzijami.

L – število mest, število primerjav nukleotidnih mest.

p – delež različnih mest glede na vsa mesta: $p = \frac{d}{L}$.

P – delež mest, ki izkazuje tranzicijske spremembe.

Q – delež mest, ki izkazuje transverzijske spremembe.

Število različnih nukleotidov (ang. No. of Differences (Nucleotide)) je cenilka, s katero podamo število mest, v katerih se primerjana zaporedja razlikujeta.

Nukleotidna p -razdalja (ang. p -distance (Nucleotide)) je delež (p) različnih mest glede na vsa primerjana mesta:

$$p_{xy} = \frac{d}{L}.$$

Jukes-Cantorjeva razdalja JC oz. Jukes-Cantorjev model (ang. Jukes-Cantor Model) napoveduje enako stopnjo zamenjav za vse pare nukleotidov A, T, C in G. Metrika za izračun Jukes-Cantorjeve razdalje:

$$d_{xy} = -\frac{3}{4} \ln \left(1 - \frac{4}{3} p \right).$$

Kimura 2 razdalja K2 oz. Kimura 2-parametrični model (ang. Kimura 2-parameter Model) upošteva stopnjo tranzicijskih in transverzijijskih substitucij in predvideva enake frekvence nukleotidov. Metrika za izračun količin pri Kimura 2 razdalji:

$$d_{xy} = -\frac{1}{2} \ln(1 - 2P - Q) - \frac{1}{4} \ln(1 - 2Q),$$

$$s = -\frac{1}{2} \ln(1 - 2P - Q) + \frac{1}{4} \ln(1 - 2Q),$$

$$v = -\frac{1}{2} \ln(1 - 2Q),$$

$$R = \frac{s}{v}.$$

Tajima-Nei razdalja TN oz. Tajima-Nei model (ang. Tajima-Nei Model) je prilagoditev izračuna genetske razdalje za primere, ko relativna frekvenca nukleotidov značilno odstopa od 0,25. V takih primerih Tajima-Neijev model daje boljše rezultate kot Jukes-Cantorjev model. Tajima-Neijevo razdaljo računamo po formuli:

$$d_{xy} = -b \ln \left(1 - \frac{p}{b} \right),$$

kjer b izračunamo s formulo:

$$b = \frac{1}{2} \left[1 - \sum_{i=1}^4 g_i^2 + \frac{p^2}{c} \right],$$

in c izračunamo s formulo:

$$c = \sum_{i=1}^3 \sum_{j=i+1}^4 \frac{x_{ij}^2}{2g_i g_j},$$

kjer je x_{ij} relativna frekvenca nukleotidnega para i in j , g_i so nukleotidne frekvence.

Tamura 3 razdalja T3 oz. Tamura 3-parametrični model (ang. Tamura 3-parameter Model) upošteva razlike v stopnji tranzicije in transverzije, prav tako upošteva pristranost G + C. Tamura 3 razdaljo in metrike računamo po formulah:

$$d_{xy} = -w_g \ln \left(1 - \frac{p}{w_g} - Q \right) - \frac{1}{2} (1 - w_g) \ln(1 - 2Q),$$

$$s = -w_g \ln \left(1 - \frac{P}{w_g} - Q \right) + \frac{1}{2} w_g \ln(1 - 2Q),$$

$$v = -\frac{1}{2} \ln(1 - 2Q),$$

$$R = \frac{s}{v},$$

kjer je w_g :

$$w_g = 2G(1 - G),$$

in je G :

$$G = g_C + g_G.$$

Tamura-Nei razdalja TN oz. Tamura-Nei model (ang. Tamura-Nei Model) upošteva razlike v stopnji substitucije med nukleotidi in razlike v frekvencah nukleotidov. Model razlikuje med deležem tranzicijskih substitucij med purini in deležem transverzijaških substitucij med pirimidini. Tamura-Nei razdaljo in metrike računamo po formulah:

$$d_{xy} = -k_1 \ln(w_1) - k_2 \ln(w_2) - k_3 \ln(w_3),$$

$$s = -k_1 \ln(w_1) - k_2 \ln(w_2) - (k_3 - 2g_R g_Y) \ln(w_3),$$

$$v = -2g_R g_Y \ln(w_3),$$

$$R = \frac{s}{v},$$

kjer so:

$$k_1 = \frac{2g_A g_G}{g_R},$$

$$k_2 = \frac{2g_T g_C}{g_Y},$$

$$k_3 = 2 \left(g_R g_Y - \frac{g_A g_G g_Y}{g_R} - \frac{g_T g_C g_R}{g_Y} \right),$$

$$w_1 = 1 - \frac{P_1}{k_1} - \frac{Q}{2g_Y},$$

$$w_2 = 1 - \frac{P_2}{k_2} - \frac{Q}{2g_Y},$$

$$w_3 = 1 - \frac{Q}{2g_R g_Y},$$

in sta P_1 in P_2 deleža tranzicijskih razlik med nukleotidi A in G ter med T in C. Oznake g_A , g_C , g_G in g_T so frekvence nukleotidov A, C, G in T; $g_R = g_A + g_G$, $g_Y = g_T + g_C$.

Generalizirani časovno reverzibilni model GTR (ang. Generalised Time Reversible) je najbolj splošen, nevtralen, neodvisen in časovno reverzibilni model.

Postopek izračuna genetske razdalje od poravnanih nukleotidov do vrednosti Jukes-Cantorjeve razdalje je predstavljen v Primeru 14.1.

Preglednica 14.1 Pregled osnovnih modelov za razlago nukleotidne variabilnosti in za izračun genetskih razdalj. Posamezni modeli na različne načine upoštevajo stopnje substitucije in razmerja med deleži nukleotidnih baz (A, C, G, T); *a*, *c*, *d* in *f* so transverzije, *b* in *e* sta tranziciji (primerjaj s Sliko 14.5)

| Model | Oznaka | Stopnja substitucije | Relativna frekvenca baz |
|---|-----------|---------------------------------|--------------------------------|
| Jukes-Cantor | JC69 (JC) | $a = b = c = d = e = f$ | $g_A = g_C = g_G = g_T = 0,25$ |
| Kimura 2 | K80 (K2P) | $(a = c = d = f) \neq (b = e)$ | $g_A = g_C = g_G = g_T = 0,25$ |
| Tajima-Nei | | $a = b = c = d = e = f$ | Frekvence g_i so proste. |
| Tamura 3 | T92 | $(a = c = d = f) \neq (b = e)$ | Frekvence g_i so proste. |
| Tamura-Nei | TN93 | $(a = c = d = f) \neq b \neq e$ | Frekvence g_i so proste. |
| Generalizirani časovno-reverzibilni model | GTR (REV) | Stopnja je prosta. | Frekvence g_i so proste. |

Omejitve na razdaljah temelječih metod. Izračun genetskih razdalj, ki temeljijo na razlikah v nukleotidih, ima več omejitev:

- Substitucije so neodvisne ne glede na nukleotidno mesto.
- Če se stopnja evlucijskih sprememb razlikuje med mesti, moramo podatke korigirati.
- Stopnja substitucij je konstantna v času in med primerjanima zaporedjema.
- Sestava nukleotidov je uravnotežena.
- Verjetnost substitucij je konstantna glede na položaj v nukleotidnem zaporedju in se ne spreminja v času.

Izbira ustreznega modela substitucije. Izbira ustreznega DNK substitucijskega modela je ključna za izračun filogenetske rekonstrukcije po metodah največjega verjetja ML in Bayesovega sklepanja BI (glej poglavje 14.3). Prav tako metodi UPGMA in NJ za izdelavo filogenetskih dreves temeljita na genetskih razdaljah (Slika 14.11), ki jih računamo po substitucijskih modelih. Posamezen substitucijski model drugače upošteva: 1. stopnjo substitucije med posameznimi pari nukleotidov, vključno z razmerjem med tranzicijskimi in transverzijijskimi substitucijami, in 2. deleže oz. frekvenco posameznih nukleotidov v zaporedju.

S **hierarhičnim testom razmerja verjetij** hLRT (ang. Hierarchical Likelihood Ratio Test) substitucijske modele razvrstimo glede na delež informacije, ki jo povzamejo iz podatkov (Tamura in sod. 2021). Na podlagi vrednosti testa skupaj s podporo Akaikovega informacijskega kriterija AIC, korigiranega AICc in Bayesovega informacijskega kriterija BIC algoritem razporedi modele po deležu razložene informacije. Vrednost AIC interpretiramo kot izgubo informacije z uporabo konkretnega modela, s katerim razlagamo informacijo iz podatkovnega niza (razlago AIC in BIC glej v poglavju 17.2). Model ali modeli z najmanj izgube informacije imajo prednost pred drugimi modeli. S **testom razmerja verjetij** LRT (ang. Likelihood Ratio Test) preverjamo ujemanje med dvema modeloma (Tamura in sod. 2021). Ob primerjanju relativno kompleksnejšega

modela s preprostejšim in ko oba modela povzemata enak ali primerljiv delež informacij iz osnovnih podatkov, ima preprostejši model prednost pred kompleksnejšim.

OTU

V filogenetski praksi za skupine organizmov neopredeljenega taksonomskega ranga uporabljamo kratico OTU (ang. Operational Taxonomic Unit). Z njo ne navajamo konkretnega taksonomskega nivoja (npr. vrsta, rod, družina ... ali subspecifični nivo: podvrsta, populacija, osebek), ampak je le-ta razviden iz konteksta.



PRIMER 14.1

Filogenetski odnosi med prvaki Primates –
izračun genetske razdalje

Izračun genetske razdalje po Jukes-Cantorjevem modelu in gradnjo filogenetskega drevesa demonstriramo na zelo kratkem zaporedju 16 nukleotidov štirih vrst prvakov: človeku, šimpanzu, gorili in orangutanu (Preglednica 14.2).

Preglednica 14.2 Nukleotidno zaporedje 16 nukleotidov za štiri vrste prvakov

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|-----------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|
| človek | A | C | T | T | G | A | C | C | C | T | T | A | C | G | A | T |
| orangutan | A | G | C | T | G | G | C | C | C | T | G | A | T | T | A | C |
| šimpanz | A | G | T | T | G | A | C | C | A | T | T | A | C | G | A | T |
| gorila | A | G | C | T | G | G | T | C | C | T | G | A | T | G | A | C |

Primerjave med zaporedji izvedemo za vsak par taksonov posebej in za vsak par nukleotidov posebej. V Preglednici 14.3 je predstavljena primerjava med nukleotidnim zaporedjem človeka in orangutana. Dejansko nas je zanimalo število substituiranih (zamenjanih) parov, teh je sedem. Od tega so bile štiri spremembe tranzicijske in tri transverzijijske. Jukes-Cantorjeva razdalja med človekom in orangutanom ima vrednost 0,657 in jo izračunamo s spodnjo enačbo, povprečna razdalja med vsemi pari je $JC = 0,48$.

$$D_{\text{človek-orangutan}} = -\frac{3}{4} \ln \left(1 - \frac{4}{3} \frac{7}{16} \right) = 0,6566$$

Razdalje med vsemi pari so predstavljene v Preglednici 14.4. Razdalje smo vključili v izgradnjo filogenetskega drevesa po metodi združevanja sosedov (Slika 14.6). Poravnavo zaporedij smo izpeljali z algoritmom MUSCLE, izračun genetskih razdalj po Jukes-Cantorjevem modelu in izgradnjo drevesa z metodo združevanja sosedov, vse postopke smo izvedli s programom MEGA 11 (Tamura in sod. 2021).

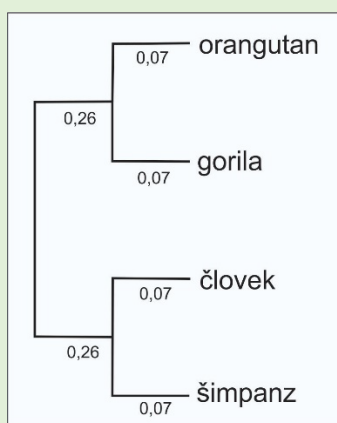
Algoritem je združil par orangutan-gorila in par šimpanz-človek (Slika 14.6).

Preglednica 14.3 Postopek primerjanja dveh nukleotidnih zaporedij med človekom in orangutanom. Primerjamo vsak posamezni par homolognih nukleotidov. Preštejemo substituirane (različne) pare, označeni so z modro barvo. Substituirani pari dobijo vrednost ena (1). Posamezni substituirani par je klasificiran kot tranzicijska ali transverzija mutacija

| OTU | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | sum |
|-----------------------------|------|-------|-------|------|------|-------|------|------|------|------|-------|------|-------|-------|------|-------|-----|
| človek | A | C | T | T | G | A | C | C | C | T | T | A | C | G | A | T | |
| orangutan | A | G | C | T | G | G | C | C | C | T | G | A | T | T | A | C | 16 |
| Ni razlike med nukleotidoma | TRUE | FALSE | FALSE | TRUE | TRUE | FALSE | TRUE | TRUE | TRUE | TRUE | FALSE | TRUE | FALSE | FALSE | TRUE | FALSE | |
| Substitucija | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 7 |
| Tranzicija | | 0 | 1 | | | 1 | | | | | 0 | | 1 | 0 | | 1 | 4 |
| Transverzija | | 1 | 0 | | | 0 | | | | | 1 | | 0 | 1 | | 0 | 3 |

Preglednica 14.4 Jukes-Cantorjeva razdalja med posameznimi vrstami

| | človek | orangutan | šimpanz | gorila |
|-----------|--------|-----------|---------|--------|
| človek | 0 | | | |
| orangutan | 0,657 | 0 | | |
| šimpanz | 0,137 | 0,657 | 0 | |
| gorila | 0,657 | 0,137 | 0,657 | 0 |



Slika 14.6 Filogenetsko drevo evolucijskih odnosov med štirimi vrstami prvakov. Filogenetsko drevo je zgrajeno po metodi združevanja sosedov. Ob vejah je podana dolžina veje.



14.2 Genske banke

Nukleotidna zaporedja pridobivamo z izolacijo iz tkiv vzorčenih organizmov. Objavljena zaporedja so dostopna v **genskih bankah** (ang. Gen Bank) npr. NCBI – National Center for Biotechnology Information, ENA – European Nucleotide Archive, EMBL – European Bioinformatics Institute, idr.

Genske banke in druge inštitucije ustvarjajo knjižnice nukleotidnih in aminokislinskih zapisov. Vsako zaporedje je povezano z informacijo o organizmu, iz katerega izvira, o avtorju zaporedja, o znanstveni publikaciji, kjer je bilo objavljeno, ter datumom in krajem nabiranja vzorca. Pomembno je tudi, da ima vsako zaporedje edinstveno kodo, ki se imenuje pristopna številka (ang. Accession Number) in omogoča sledljivost.

Pridobivanje podatkov iz GeneBank NCBI, ki je dostopna na povezavi <https://www.ncbi.nlm.nih.gov/genbank/>, poteka z iskalnikom, v katerega vpišemo ključne besede, sestavljene iz gena in organizma npr.: *cytochrome b* in *Apodemus*. Rezultat iskanja je pregled zaporedij, vnesenih v bazo (lahko jih je več sto), z možnostjo pregledovanja po več argumentih (objava, avtor, lokaliteta ...). Ko se odločimo za zaporedje, izberemo FASTA zapis nukleotidnega zaporedja, ga prekopiramo v beležnico in shranimo kot .txt datoteko. V beležnici enega za drugim nanizamo zaporedja posameznih OTU. To so izhodni podatki in so namenjeni za nadaljnjo obdelavo.

Za poravnavo zaporedij ter izračun in izpis dendrograma je več ponudnikov in programov za različno zahtevne uporabnike. Osnovni nabor je dostopen na povezavi <https://www.ncbi.nlm.nih.gov/genbank/>. Po oddaji podatkov počakamo na odgovor po elektronski pošti, s priloženo povezavo do rezultatov, ki obsegajo poravnana zaporedja in dendrogram. Glavni koraki v iskanju in pripravi zaporedij je predstavljen v Primeru 14.2.



PRIMER 14.2

Protokol priprave nukleotidnih podatkov

- Na spletni strani genske banke izvedemo poizvedbo za nukleotidno zaporedje tarčnega gena (npr. cytochrom b, MT-CYB) za izbrano vrsto (npr. pet vrst rodu *Apodemus*): <https://www.ncbi.nlm.nih.gov/guide/data-software/>.
- V obliki preglednice podamo osnovne lastnosti vzorcev: OTU, referenco, število baznih parov, dostopno kodo.
- FASTA obliko zapisa prenesemo (copy/paste) v beležnico.
- Nukleotidna zaporedja več organizmov naj bodo primerljive dolžine.

Oblika zapisa: naslovna vrstica posameznega vzorca se začne z » > » in je namenjena sledljivosti; sledi nukleotidno zaporedje:

```
>AJ511914.1 Apodemus sylvaticus mitochondrial partial mt-cyb gene for cytochrome
b, tissue library JRM-130
ATTCGAAAAACACACCCACTACTAAAAATTATCAATCACTCTTTCATCGACCTACCAGCTCCATCTAACATTTTCATCATG
ATGAAACTTCGGCTCATTGCTAGGAGTCTGCCTGATAATCCAAATCCTCACAGGCTTATTTCTAGCAATACACTACACAT
CAGACACAATAACAGCATTTTCTTCAGTAACCCATATCTGTTCGAGACGTA.....
>AJ605670.1 Apodemus flavicollis mitochondrial partial mt-cyb gene for cytochrome
b, tissue library JRM-238
AATGACCACATTCGAAAAACACACCCCTACTAAAAATTATTAACCACTCCTTCATTGACTTACCAGCCCCATCCAATAT
TTCATCATGATGAAACTTTGGCTCCCTTTTAGGAATCTGCCTAATAATCCAGATCCTTACAGGCCTATTTCTAGCAATAC
ACTACACATCAGATAACAATAACAGCATTTTCTTCAGTAACTCATATTTGC.....
```

- Opcija – vse zbrane vzorce prenesemo (copy/paste) v pogovorno okno programa MAFFT 7 in oddamo v analizo: <https://mafft.cbrc.jp/alignment/server/>. Rezultate vrnejo po elektronski pošti na naslov, ki ga vnesemo v pogovorno okno.
- Opcija – vse zbrane vzorce prenesemo (copy/paste) v pogovorno okno programa MEGA 11 in v njem izvedemo poravnavo nukleotidnih zaporedij, poiščemo ustrezni substitucijski model, izračunamo genetske razdalje in zgradimo drevo.



14.3 Metoda največjega verjetja

Metoda največjega verjetja ML (ang. Maximum Likelihood Method) gradi topologijo drevesa z uporabo statističnih postopkov, v katerih modelira število parametrov in njihove vrednosti, da doseže največje verjetje. Z drugimi besedami, med nukleotidnimi zaporedji algoritem poišče takšno kombinacijo v številu in vrednostih parametrov, ki maksimirajo verjetje.

Metoda največjega verjetja ima naslednje elemente in korake:

- **Podatki** – multiplo poravnana in enako dolga zaporedja nukleotidov (ang. Multiple Sequence Alignment MSA).
- **Parametri** – nabor cenilk (sin. parametrov), s katerimi razlagamo variabilnost podatkov, npr.: frekvenca posameznih nukleotidov, stopnja substitucije med posameznimi pari nukleotidov, različna cena za posamezni par substitucije nukleotid – nukleotid, topologija drevesa in dolžina vej ter druge značilnosti (nanizane so v uvodu v to poglavje).
- **Model** – substitucijski model je simulacija evolucijskega dogajanja, ki ga razlagamo.

- **Verjetje** – je vrednost verjetja za posamezni znak. Verjetje posameznega znaka je vsota vseh verjetnosti za stanje predniškega znaka, glede na opazovano kombinacijo nukleotidov pri potomcih.
- **Skupno verjetje** – je vrednost verjetja za vse znake. Skupno verjetje je produkt verjetij posameznih znakov.
- **Največje verjetje** – je vrednost verjetja, pri kateri imajo vključeni parametri vrednosti, s katerimi najbolje razlagajo variabilnost v podatkih.

Primer analize po metodi največjega verjetja je demonstriran v Primeru 14.3.

TERMINOLOŠKO

Verjetnost je mera dogodka, ki je izražena z vrednostjo od 0 do 1. Npr. pri popisovanju sov je verjetnost za prisotnosti kozače $p = 0,5$ kar pomeni, da na 10 opazovanj pričakujemo 5 primerov prisotnosti te vrste sove.

Verjetje je vrednost funkcije verjetja za vrednosti slučajne spremenljivke v vzorcu. Z drugimi besedami, funkcija verjetja je produkt posameznih verjetnosti. Npr. verjetnost za opazovanje kozače v Prekmurju je $p = 0,01$, na Štajerskem $p = 0,05$, na Gorenjskem $p = 0,2$, na Notranjskem $p = 0,6$, na Dolenjskem $p = 0,8$ in na Primorskem je $p = 0,3$. Izračun vrne vrednost verjetja, da bodo vsi popisovalci videli kozačo $L = 0,000014$ oz. $\ln L = -11,148$. Produkt posameznih verjetnosti je verjetje, da bi vseh 6 popisovalcev vsak na svoji lokaciji potrdili to vrsto sove. To je malo verjetno.



PRIMER 14.3

Filogenetski odnosi med prvaki – Metoda največjega verjetja

Zastavili smo si nalogo, izdelati filogenetsko drevo, s katerim razložimo evolucijske odnose med štirimi vrstami prvakov po metodi največjega verjetja ML. Postopek smo izvedli v naslednjih korakih:

1. Oblikujemo multiplo poravnana zaporedja nukleotidov.
2. Oblikujemo nabor parametrov (sin. cenilk).
3. Izberemo substitucijski model.
4. Izračunamo verjetnosti za stanje predniških znakov glede na opazovano kombinacijo nukleotidov pri potomcih.
5. Izračunamo verjetje za posamezni znak, ki je vsota vseh verjetnosti iz koraka 4.
6. Izberemo (naključne) začetne vrednosti parametrov.

7. Ponavljamo računanje skupnega verjetja, ki je produkt verjetij posameznega znaka iz koraka 5, ob spreminjanju vrednosti parametrov.
8. Maksimalna vrednost največjega verjetja je končni rezultat, ki vsebuje vrednosti parametrov iz koraka 3.

Analizo največjega verjetja smo izvedli v programu MEGA 11 (Tamura in sod. 2021). Celoten postopek obsega korake in procedure:

1. Oblikujemo multiplo poravnana zaporedja nukleotidov

Podatkovni niz je identičen kot v Primeru 14.1 Preglednica 14.2. Tukaj bomo predstavili izračun verjetnosti samo za mesto 3, ki vsebuje nukleotide:

| | ... | 3 | ... |
|-----------|-----|---|-----|
| človek | | T | |
| orangutan | | C | |
| šimpanz | | T | |
| gorila | | C | |

2. Oblikujemo nabor cenilk (sin. parametrov)

S cenilkami oz. parametri bomo razlagali variabilnost v osnovnih podatkih. Vrednosti parametrov, ki imajo maksimirano verjetje, so naš iskani rezultat:

- Topologija drevesa in dolžina vej.
- Frekvence posameznih nukleotidov: $\pi_T, \pi_C, \pi_A, \pi_G$
- Stopnja substitucije oz. substitucijska verjetnost za zamenjavo med posameznimi pari nukleotid – nukleotid: $P_{AA}, P_{AC}, P_{AT}, P_{AG} \dots P_{GG}$.

3. Izberemo substitucijski model

S testom razmerja verjetij v programu MEGA 11 smo iskali najustreznejši model za naše podatke. Algoritem je kandidatne modele razvrstil glede na vrednost AIC_C in BIC (Preglednica 14.5). Najnižjo vrednost AIC_C ima najenostavnejši model (Jukes-Cantorjev model), ki predvideva izenačene frekvence nukleotidov in enako stopnjo substitucij med vsemi pari nukleotidov.

Preglednica 14.5 Razvrstitev petih najbolje ocenjenih substitucijskih modelov glede na vrednost AIC_c in BIC

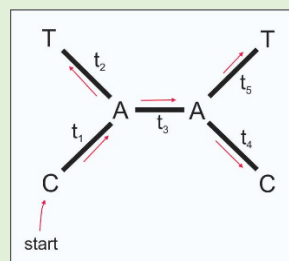
| Model | Št. parametrov | BIC | AIC_c | $\ln L$ |
|-------|----------------|--------|---------|---------|
| JC | 5 | 133,36 | 123,60 | -56,29 |
| K2 | 6 | 136,10 | 124,62 | -55,58 |
| JC+I | 6 | 137,52 | 126,04 | -56,29 |
| JC+G | 6 | 137,53 | 126,05 | -56,29 |
| K2+I | 7 | 140,26 | 127,15 | -55,58 |

4. Izračunamo verjetnosti za stanje predniškega znaka glede na opazovano kombinacijo nukleotidov pri potomcih

Prikazali bomo postopek samo za mesto 3, ki ima zaporedje T, C, T, C.

Izračunamo verjetnost Pr predniškega znaka za vse štiri nukleotide. Parametri, ki jih upoštevamo, so frekvence posameznih nukleotidov π_i in verjetnost substitucij P_{ij} v času t_i . Postopek začnemo s prvim modelom, ki predvideva starševsko kombinacijo A-A. To je samo prvi od 16 členov izračuna verjetja za mesto 3, ki so v celoti prikazani v naslednjem koraku. Izračun pričnemo na poljubnem mestu in ga nadaljujemo proti drugemu koncu, kot je prikazano na Sliki 14.7. Verjetnost zapišemo v obliki enačbe:

$$Pr = \pi_C P_{CA}(t_1) P_{AT}(t_2) P_{AA}(t_3) P_{AC}(t_4) P_{AT}(t_5).$$



Slika 14.7 Razporeditev štirih nukleotidov potomcev, ki so na zunanjih razvejitvah, za prvi hipotetični starševski par A-A na notranjih razvejitvah. Puščice usmerjajo sestavljanje enačbe za izračun verjetnosti takšnega poteka substitucij.

5. Izračunamo verjetje za posamezni znak, ki je vsota vseh verjetnosti iz koraka 4

Relacija med starševskimi nukleotidi in potomci ima 16 možnih kombinacij, kot je predstavljeno na Sliki 14.8 in zapisano v obliki enačbe. Verjetje za mesto 3 L_3 je vsota vseh šestnajstih verjetnosti.

$$L_3 = P_{AA} + P_{AC} + P_{AT} + P_{AG} + P_{CA} + P_{CC} + P_{CT} + P_{CG} + P_{TA} + P_{TC} + P_{TT} + P_{TG} + P_{GA} + P_{GC} + P_{GT} + P_{GG}$$

$$\begin{aligned}
L_3 = & \Pr \left(\begin{array}{c} T & & T \\ & \diagdown & / \\ & A & - A \\ & / & \diagdown \\ C & & C \end{array} \right) + \Pr \left(\begin{array}{c} T & & T \\ & \diagdown & / \\ & A & - C \\ & / & \diagdown \\ C & & C \end{array} \right) + \Pr \left(\begin{array}{c} T & & T \\ & \diagdown & / \\ & T & - T \\ & / & \diagdown \\ C & & C \end{array} \right) + \Pr \left(\begin{array}{c} T & & T \\ & \diagdown & / \\ & A & - G \\ & / & \diagdown \\ C & & C \end{array} \right) + \\
& + \Pr \left(\begin{array}{c} T & & T \\ & \diagdown & / \\ & C & - A \\ & / & \diagdown \\ C & & C \end{array} \right) + \Pr \left(\begin{array}{c} T & & T \\ & \diagdown & / \\ & C & - C \\ & / & \diagdown \\ C & & C \end{array} \right) + \Pr \left(\begin{array}{c} T & & T \\ & \diagdown & / \\ & C & - T \\ & / & \diagdown \\ C & & C \end{array} \right) + \Pr \left(\begin{array}{c} T & & T \\ & \diagdown & / \\ & C & - G \\ & / & \diagdown \\ C & & C \end{array} \right) + \\
& + \Pr \left(\begin{array}{c} T & & T \\ & \diagdown & / \\ & T & - A \\ & / & \diagdown \\ C & & C \end{array} \right) + \Pr \left(\begin{array}{c} T & & T \\ & \diagdown & / \\ & T & - C \\ & / & \diagdown \\ C & & C \end{array} \right) + \Pr \left(\begin{array}{c} T & & T \\ & \diagdown & / \\ & T & - T \\ & / & \diagdown \\ C & & C \end{array} \right) + \Pr \left(\begin{array}{c} T & & T \\ & \diagdown & / \\ & T & - G \\ & / & \diagdown \\ C & & C \end{array} \right) + \\
& + \Pr \left(\begin{array}{c} T & & T \\ & \diagdown & / \\ & A & - A \\ & / & \diagdown \\ C & & C \end{array} \right) + \Pr \left(\begin{array}{c} T & & T \\ & \diagdown & / \\ & G & - C \\ & / & \diagdown \\ C & & C \end{array} \right) + \Pr \left(\begin{array}{c} T & & T \\ & \diagdown & / \\ & C & - T \\ & / & \diagdown \\ C & & C \end{array} \right) + \Pr \left(\begin{array}{c} T & & T \\ & \diagdown & / \\ & G & - G \\ & / & \diagdown \\ C & & C \end{array} \right)
\end{aligned}$$

Slika 14.8 Izračun verjetja za znak 3, L_3 je vsota vseh 16 verjetnosti Pr in razporeditev vseh 16 kombinacij starševskih nukleotidov, ki sta na notranjih razvejitev.

6. Izberemo (naključne) začetne vrednosti parametrov

Izbiro začetnih vrednosti izvede algoritem.

7. Računanje skupnega verjetja, ki je vsota verjetij posameznega znaka iz koraka 5, in ponavljanje ob spreminjanju vrednosti parametrov

Skupno verjetje L izračunamo kot produkt verjetij vseh posameznih mest L_j z enačbo:

$$L = L_1 L_2 L_3 L_4 L_5 L_6 L_7 L_8 L_9 L_{10} L_{11} L_{12} L_{13} L_{14} L_{15} L_{16} = \prod_{j=1}^N L_j.$$

Posamezna verjetja imajo nizko vrednost, manjšo od 1, množenje nizkih vrednosti vrne še nižje vrednosti. Da vrednosti niso ekstremno majhne, je izračun skupnega verjetja transformiran z naravnim logaritmom in ima obliko:

$$\begin{aligned}
\ln(L) = & \ln(L_1) + \ln(L_2) + \ln(L_3) + \ln(L_4) + \ln(L_5) + \ln(L_6) + \ln(L_7) + \ln(L_8) + \\
& \ln(L_9) + \ln(L_{10}) + \ln(L_{11}) + \ln(L_{12}) + \ln(L_{13}) + \ln(L_{14}) + \ln(L_{15}) + \ln(L_{16}) = \\
& \sum_{j=1}^N \ln(L_j).
\end{aligned}$$

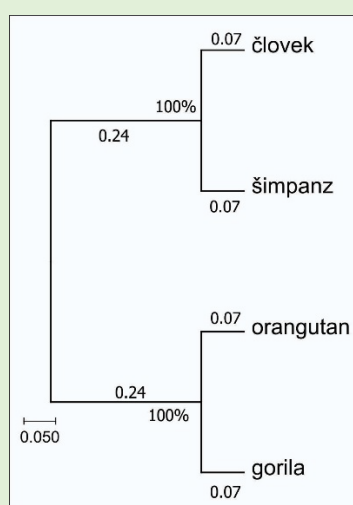
Ker so vrednosti verjetja logaritmirane, jih sedaj seštevamo. In ker so surova števila v razponu nad 0 in manj kot 1 ($0 < L < 1$), je logaritem vrednosti negativno število.

8. Končni rezultat, ki vsebuje vrednosti parametrov iz koraka 3, je maksimalna vrednost največjega verjetja

Algoritem ponavlja korak 7 in ob vsaki ponovitvi spremeni vrednosti parametrov. Parametri se zvišujejo v smeri, ki večajo skupno verjetje. Postopek je končan, ko je doseženo maksimalno verjetje, torej takrat, ko nadaljnje spreminjanje parametrov vrača nižje vrednosti verjetja.

Rezultat filogenetske rekonstrukcije po metodi največjega verjetja so vrednosti petih parametrov: 1. Substitucijski model je Jukes-Cantor. 2. Topologija drevesa je predstavljena na Sliki 14.9. 3. Dolžina vej je predstavljena na Slika 14.9. 4. Frekvence nukleotidov so izenačene in imajo vrednost: $\pi_T = 0,25$, $\pi_C = 0,25$, $\pi_A = 0,25$, $\pi_G = 0,25$. 5. Stopnja substitucijskih zamenjav je izenačena in ima vrednost: $r_{AT} = 0,083$, $r_{AC} = 0,083$, $r_{AG} = 0,083$, $r_{TA} = 0,083$, $r_{TC} = 0,083$, $r_{TG} = 0,083$, $r_{CT} = 0,083$, $r_{CG} = 0,083$, $r_{GA} = 0,083$, $r_{GC} = 0,083$.

Najvišja vrednost največjega verjetja je dosežena pri $\ln L = -59,14$. Filogenetsko drevo med štirimi prvaki poveže para človek-šimpanz in orangutan-gorila (Slika 14.9). Bootstrap testiranje s 500 ponovitvami samovzorčenja izkazuje 100 % podporo obema razvejitvama.



Slika 14.9 Filogenetsko drevo med štirimi vrstami prvakov. Narejeno po metodi največjega verjetja in z Jukes-Cantorjevim modelom. Vrednost ob veji podaja dolžino veje. Vrednost ob razvejitvi podaja bootstrap podporo tej razvejitvi. Merilna skala podaja število substitucij na označeni razdalji.



14.4 Bayesovo sklepanje

Bayesov pristop razlaga verjetnost s stopnjo našega prepričanja v nek dogodek. Stopnja prepričanja temelji bodisi na predhodnem znanju in meritvah bodisi na osebnem prepričanju o dogodku. V tem se Bayesov pristop razlikuje od t. i. frekvencističnega pristopa. Statistična definicija verjetnosti pravi, da je verjetnost dogodka A v danem poskusu število P , pri katerem se stabilizira relativna frekvenca dogodka A v velikem številu ponovitev tega poskusa. To pomeni, če poskus ponovimo m krat in se A zgodi n krat, je relativna frekvenca: n/m in to je ocena za verjetnost P .

14.4.1 Bayesov izrek

Bayesov pristop za izračun verjetnosti uporabi pogojno verjetnost. To je opis verjetnosti dogodka na podlagi predhodnih informacij ali prepričanja o dogodku ali o pogojih, povezanih z dogodkom. Na slednjem je narejen **Bayesov izrek** (ang. Bayes' Theorem) in utemeljeno **Bayesovo sklepanje** (ang. Bayesian Inference), s katerim ocenimo parametre porazdelitve verjetnosti oz. parametre modela (Bayes' theorem 2023, Bayesian inference 2023).

Pogojna verjetnost (ang. Conditional Probability) opisuje verjetnost dogodka z upoštevanjem drugih dogodkov in jo zapišemo s formulo:

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

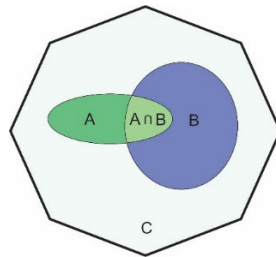
kjer so:

$P(A|B)$ je pogojna verjetnost, da se zgodi dogodek A, če se je zgodil dogodek B.

$P(A \cap B)$ je verjetnost produkta dogodkov A in B.

$P(B)$ je verjetnost, da se zgodi dogodek B.

Pogojno verjetnost razumemo, da je verjetnost dogodka A ob pogoju, da se je zgodil dogodek B, enaka kvocientu med števcem (verjetnostjo produkta dogodkov A in B) in imenovalcem (verjetnostjo dogodka B) (Conditional probability 2023). Dogodek A in B ter njun produkt $A \cap B$ je predstavljen z Vennovim diagramom na Sliki 14.10.



Slika 14.10 Vennov diagram dogodkov; množica A predstavlja dogodek A in množica B predstavlja dogodek B; presečišče obeh množic predstavlja produkt dogodkov A in B; prostor C predstavlja še vse druge možne dogodke.

Skupna verjetnost (ang. Joint Probability) dogodkov A in B oz. verjetnost produkta $A \cap B$ zapišemo s formulami:

$$P(A \cap B) = P(A|B)P(B) \text{ in}$$

$$P(A \cap B) = P(B|A)P(A)$$

(Joint probability distribution 2023).

Iz pogojne verjetnosti in skupne verjetnosti izpeljemo **Bayesov izrek**:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

(Bayes' theorem 2023). Zgornjo enačbo razlagamo na način:

$P(B|A)$ je posteriorna verjetnost (sin. končna verjetnost, ang. Posterior Probability).

$P(A|B)$ je matematično verjetje (sin. matematično pričakovanje, ang. Likelihood) (postopek izračuna verjetja je predstavljen v Primeru 14.3).

$P(B)$ je priorna verjetnost (sin. apriorna verjetnost, vnaprejšnja verjetnost, ang. Prior Probability).

$P(A)$ je robna verjetnost (ang. Marginal Probability).

V filogenetskih raziskavah je dogodek B hipoteza (filogenetsko drevo), ki jo preverjamo, in dogodek A so meritve (poravnana nukleotidna zaporedja) oz. naše znanje o dogodku (npr. stopnja substitucije med nukleotidi, frekvenca nukleotidov). Filogenetsko obliko Bayesovega izreka zapišemo:

$$P(\text{Drevo} + \text{Parametri} \mid \text{Zaporedje}) = \frac{P(\text{Zaporedje} \mid \text{Drevo} + \text{Parametri}) P(\text{Drevo} + \text{Parametri})}{P(\text{Zaporedje})}$$

V skladu z zgornjo razlago splošne enačbe posteriorne verjetnosti in njeno aplikacijo v filogenetski rekonstrukciji posamezne člene enačbe razlagamo na način:

$P(\text{Drevo} + \text{Parametri} \mid \text{Zaporedje})$ – posteriorna verjetnost porazdelitve vseh možnih vrednosti parametrov substitucijskega modela.

$P(\text{Zaporedje} \mid \text{Drevo} + \text{Parametri})$ – verjetje prileganja med poravnanimi nukleotidnimi zaporedji in parametri modela.

$P(\text{Drevo} + \text{Parametri})$ – priorna verjetnost, ki jo tvorijo vrednosti parametrov substitucijskega modela (stopnja substitucije in deleži nukleotidov) in topologija drevesa (navadno privzamemo, da imajo vse topologije enako verjetnost).

$P(\text{Zaporedje})$ – robna verjetnost vključuje porazdelitev zaporedij v času oz. pri prednikih in jo zapišemo v obliki enačbe:

$$P(\text{Zaporedje}) = P(\text{Zaporedje}, t_0) + P(\text{Zaporedje}, t_1) + \dots + P(\text{Zaporedje}, t_n),$$

pri tem je $n + 1$ število vseh možnih dreves. Enačbo preuredimo v posteriorno verjetnost:

$$P(\text{Zaporedje}) = P(\text{Zaporedje} \mid t_0) P(t_0) + P(\text{Zaporedje} \mid t_1) P(t_1) + \dots + P(\text{Zaporedje} \mid t_n) P(t_n),$$

pri tem je vrednost t enaka: $t = 1/n + 1$.

Sedaj imamo vse člene enačbe, vendar zaradi prevelikega števila vseh možnih kombinacij izračun konkretnega dreves ni mogoč.

Izračun zastavimo na način, da parametre modela (deleži nukleotidov, stopnja substitucije ...) združimo v enoten vektor θ (theta) in namesto posteriorne verjetnosti računamo posteriorno porazdelitev f . Preurejena enačba za računanje posteriorne porazdelitve ima obliko:

$$f(\theta|A) = \frac{f(A|\theta)f(\theta)}{\int f(\theta)f(A|\theta) d\theta}$$

(Posterior probability 2023). Slednja pomeni, da je posteriorna porazdelitev sorazmerna produktu priorne porazdelitve in verjetja. Z drugimi besedami, računamo funkcijo verjetnosti porazdelitve ob danih vrednosti parametrov. To pomeni, da ne obstaja ena sama rešitev oz. rešitev ni eno samo filogenetsko drevo. Rešitev oz. hipotez o filogenetskem drevesu je več in iščemo interval s kombinacijo rešitev, pri kateri je verjetnost najvišja (Slika 14.11). Ta problem rešujemo z metodo markovskih verig.

14.4.2 Monte Carlo z markovskimi verigami

Metoda Monte Carlo z markovskimi verigami MCMC (ang. Monte Carlo Markov Chain) je stohastična metoda za reševanje računsko zahtevnih primerov, v kateri je sedanja vrednost parametra odvisna le od njegove predhodne vrednosti (Markov chain Monte Carlo 2023).

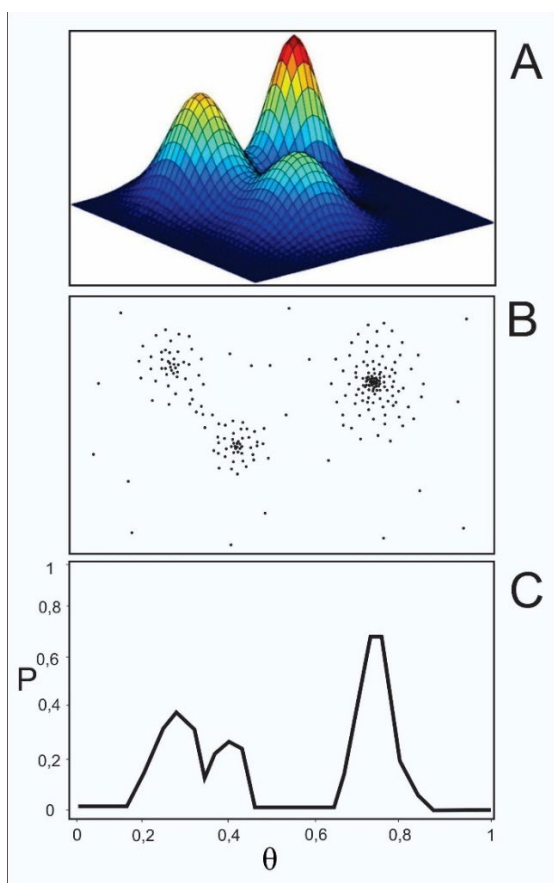
Začetni pogoji v MCMC:

1. **Prostor parametrov** zajema vse parametre in vse njihove možne vrednosti (na sliki 14.10 so to vsa tri polja A, B in C). Vsaka sprememba vrednosti kateregakoli parametra pomeni novo točko i v prostoru.
2. Za vsako točko x v prostoru parametrov, torej za vsako posamezno vrednost parametra, lahko izračunamo **matematično verjetje**: $L_i = P(\text{Podatki} \mid \text{Vrednost parametra})$.
3. Za vsako točko x v prostoru parametrov, torej za vsako posamezno vrednost parametra poznamo **priorno verjetnost**: $P_i = P(\text{Vrednost parametra})$.
4. Za vsako točko x v prostoru parametrov lahko izračunamo **posteriorno verjetnost**, to je produkt med priorno verjetnostjo in verjetjem: $P_x = P L$.

MCMC ima definirane priorne verjetnosti, to je prostor parametrov, pravimo mu **pokrajina verjetnosti** (ang. Probability Landscape) (Slika 14.11.A). To pomeni, da ima vsaka točka v pokrajini znano vrednost vnaprejšnje verjetnosti. Markovski algoritem

naključno izbere točko v pokrajini in vrednost priorne verjetnosti pomnoži z matematičnim verjetjem za izbrano točko. Rezultat je vrednost posteriorne verjetnosti P_x . Iz izhodiščne točke algoritem izbere naključno sosednjo točko in za novo točko izračuna posteriorno verjetnost P_y . V primeru, da ima nova točka enako ali višjo vrednost P ($P_y \geq P_x$), se algoritem premakne nanjo, sicer pa ne. Sledi izbira naslednje sosednje točke. Algoritem se premika po pokrajini priorne verjetnosti in jo primerja z empiričnim podatki, rezultat primerjav je posteriorna verjetnost. Za vsako točko, na katero se algoritem premakne, zabeleži vrednosti vseh parametrov. Na tak način določa točke v prostoru (Slika 14.11.B), razporeditev točk vzdolž izbranega gradienta oz. parametra prikažemo z grafom gostote porazdelitve (Slika 14.11.C).

Primerjanje vrednosti P_i med sedanjo in prihodnjo točko se ponavlja (lahko več milijon krat) in pravimo, da algoritem **vzorči porazdelitev** po pokrajini verjetnosti. Algoritem napreduje po pokrajini, dokler se viša vrednost posteriorne verjetnosti P . V žargonu pravimo, da je pokrajina verjetnosti hribovita in ko algoritem pride na vrh hriba, doseže maksimum in se ustavi. Težava, ki se pri tem pojavi, je zaustavitev na lokalnem vrhu. Dejansko ne vemo, ali je algoritem našel najboljšo možno rešitev ali je obtičal na lokalni vzpetini. Optimalna rešitev je takšna kombinacija parametrov, ki najbolje opiše filogenetsko rekonstrukcijo, to informacijo pridobimo z vzorčenjem celotne pokrajine verjetnosti. Za rešitev obtičanega algoritma na lokalnem maksimumu je razvitih nekaj metod, v filogeniji najpogosteje uporabimo Metropolis vzorčenje.



Slika 14.11 Shematski prikaz Bayesovega sklepanja. A Pokrajina verjetnosti oz. prostor parametrov, po kateri MCMC vzorči priorno verjetnost. Pokrajina ima tri vrhove, od tega dva lokalna. B Točke na pokrajini verjetnosti, na katere se je algoritem premaknil in izvedel vzorčenje, večja gostota točk pomeni zgostitev višjih posteriornih vrednosti. C Porazdelitev gostote posteriorne verjetnosti, na x osi je gradient parametrov θ in na y osi je posteriorna verjetnost P . Optimalne vrednosti parametrov so znotraj razpona vrednosti parametrov θ od 0,7 do 0,8.

14.4.3 Metropolis Monte Carlo z markovskimi verigami

V filogenetsko rekonstrukcijo na poravnanih nukleotidnih zaporedjih po navadi vključimo več taksonov in več parametrov. Množica podatkov z vrednostmi parametrov pomeni več milijonov možnih kombinacij, med katerimi moramo poiskati najugodnejše. Numerično iskanje najugodnejše kombinacije parametrov v multidimenzionalnem prostoru poteka z metodo markovskih verig. Zanj smo v predhodnem poglavju zapisali, da lahko obtiči na lokalnem vrhu. Rešitev navedenega problema je v hkratnem zagonu več markovskih verig, metoda s simultanimi verigami se imenuje **Metropolis Monte Carlo z markovskimi verigami MCMCMC** (ang. Metropolis-coupled Monte Carlo Markov Chain). Algoritem deluje na način, da ena MCMC izvaja vzorčenje v normalnem postopku (opisano v predhodnem poglavju 14.4.2), to je t.i. »hladna veriga«. Preostale verige se po pokrajini pomikajo s pospešeno hitrostjo (hitrost algoritma lahko definiramo), in to so t.i. »vroče verige«. Na vsake toliko korakov se vroča veriga zamenja s hladno verigo. Po tem algoritmu hladna veriga lahko obiše več oddaljenih točk in izvede več vzorčenj. Obenem je to pristop, da hladna veriga ne obtiči na lokalnem vrhu in da katera od verig obiše tudi najvišji vrh v pokrajini verjetnosti (Markov chain Monte Carlo 2023, Metropolis–Hastings algorithm 2023).

Dodatni ukrep za preprečitev zastoja markovske verige na lokalnem vrhu je t.i. »zažiganje« začetnih točk. Z drugimi besedami, zavrzemo prvih nekaj sto ali tisoč meritev. To pomeni, da bo približno 1.000 zažganih vzorčenj odpravilo vpliv začetnih meritev oz. začetnega lokalnega maksimuma. Nadaljnje izbrane vrednosti parametrov ohranimo.

Rezultat Bayesovega sklepanja je razpon vrednosti parametrov, znotraj katerega je najvišja verjetnost za topologijo filogenetskega drevesa in za dolžino njegovih vej.

Primer Bayesovega sklepanja s filogenetsko rekonstrukcijo med belonogimi mišmi je predstavljen v Primeru 14.5.

14.5 Filogenetska drevesa

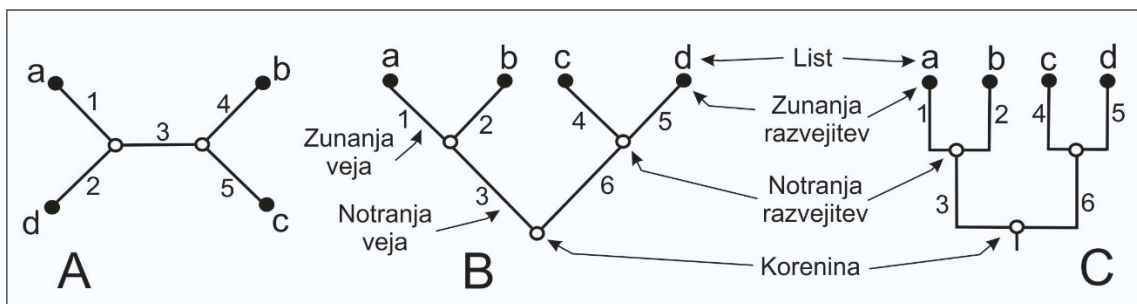
Filogenetsko drevo je graf v obliki drevesa (Sliki 14.12 in 14.15), s katerim ponazorimo hipotezo o evolucijskih odnosih med predniki in njihovimi potomci, ki so privedli do razvoja danes živečih ali že izumrlih taksonov. Topologija drevesa je vzorec razvejanosti in predstavlja rekonstrukcijo evolucijskih odnosov. V **koreninjenem** (ang. Rooted Tree) filogenetskem drevesu **korenino** predstavlja zadnji skupni prednik obravnavanih taksonov (Slika 14.15.B). V primeru, kadar zadnji skupni prednik ni znan, zgradimo **nekoreninjeno drevo** (ang. Unrooted Tree) (Slika 14.15.A). Evolucijski dogodki predstavljajo **notranje razvejitve** (ang. Internal Nodes) in pomenijo divergenco hipotetičnega predniškega taksona (kladogeneza), ko iz njega nastaneta dva nova taksona

– evolucijska potomca. **Zunanje razvejitev** (ang. External Nodes, sin. Terminal Nodes, Tips) predstavljajo dejanske taksoni, ki so predmet obravnave. Razvejitev so med seboj povezana z **vejami** (ang. Branch, sin. Edge). Notranje veje povezujejo notranje razvejitev, zunanja razvejitev je povezano z zunanjo vejo. **Listi** (ang. Leaf) na vrhu filogenetskega drevesa so imena predstavnikov obravnavanih taksonov. (Opomba. Nekateri avtorji enačijo zunanje razvejitev in liste). Število notranjih razvejitev in število vej je povezano s številom zunanjih razvejitev oz. s številom listov N in se razlikuje med nekoreninjenim in koreninjenim drevesom. Vsako posamezno vejo lahko označimo, kot je predstavljeno na Slikah 14.12.A, B in C in za vsako izračunamo dolžino vej, kar je predstavljeno pri posamezni metodi gradnje drevesa. Za nekoreninjeno drevo velja:

$$\begin{aligned}\text{Št. vej} &= 2N - 3 \text{ in} \\ \text{Št. notranjih razvejitev} &= N - 2.\end{aligned}$$

V koreninjenem drevesu se število vej in razvejitev izračuna:

$$\begin{aligned}\text{Št. vej} &= 2N - 2 \text{ in} \\ \text{Št. notranjih razvejitev} &= N - 1.\end{aligned}$$



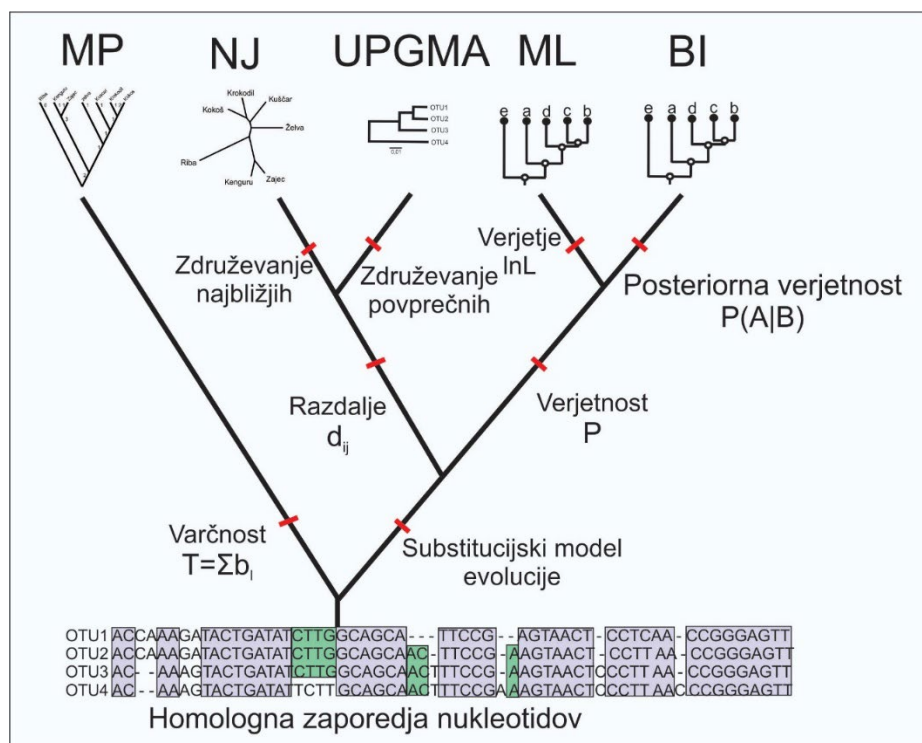
Slika 14.12 Filogenetsko drevo. Topologija drevesa prikazuje filogenetske odnose med štirimi listi, to so taksoni a , b , c , in d , ki se nahajajo na zunanjih razvejitvah in so označeni s polnimi krogi. Razvejitev so med seboj povezane z vejami, ki so označena s številko (1, 2, 3, 4, 5 in 6) in so predstavljene s polno črto. Veje, ki povezujejo notranje razvejitev, so notranje (veja 3 in 6), zunanje veje povezujejo zunanje in notranje razvejitev (veje 1, 2, 4 in 5). Notranja razvejitev predstavlja dihotomno razvejitev v dva taksona in je označeno s praznim krogom. Korenina je zadnji skupni prednik. A Nekoreninjeno drevo. B Koreninjeno drevo, v obliki obrnjene piramide. C Koreninjeno drevo, v obliki pravokotnika. Drevesi B in C sta vsebinsko identični.

Na voljo so različni postopki za gradnjo filogenetskih dreves. Predstavljamo nekaj najpogosteje uporabljenih pristopov v raziskavah filogenetskih odnosov. Alternativni pristopi h gradnji in k razumevanju dreves so nanizani v nadaljevanju.

Algoritmično vs. iskanje. Filogenetsko drevo lahko gradimo po dveh pristopih: prvi je **algoritmičen** (ang. Algorithmic Methods) in drugi je diskretni z **iskanjem drevesa** (ang. Tree-searching Methods). Vhodni podatki so bodisi morfološki znaki bodisi poravnana

nukleotidna zaporedja – na znakih temelječe metode. Znake oblikujemo v pravokotno matriko OTU x znaki.

Algoritmičen pristop se prične z izbiro substitucijskega modela, s katerim najbolje opišemo variabilnost v sestavi nukleotidnega zaporedja (glej poglavje 14.1). Število parametrov in njihove vrednosti vključimo v gradnjo drevesa. Algoritmičen pristop omogoča gradnjo drevesa po dveh osnovnih poteh, vsaka od njiju ima več različic; omenjamo samo dve osnovni metodi: 1. pot, ki temelji na razdaljah, in 2. pot, ki temelji na verjetnosti. Na **razdaljah temelječi pristop** v substitucijskem modelu izračuna genetske razdalje med vsemi pari OTU. Razdalje so oblikovane v simetrično matriko, v gradnjo drevesa jih vpeljemo z **metodo združevanja sosedov NJ** ali z **UPGMA metodo** (metodi sta predstavljeni v Poglavju 13). Na **verjetnosti temelječi pristop** uporabi substitucijski model za izračun verjetnosti posameznih razvejitev in celotne topologije drevesa. Ta pristop vključuje **metoda največjega verjetja** in **metoda Bayesovega sklepanja** (Slika 14.13).



Slika 14.13 Klasifikacija metod filogenetske rekonstrukcije na nukleotidnih zaporedjih. Vhodni podatki so homologna, poravnana in enako dolga zaporedja nukleotidov, izhodni rezultat je filogenetska rekonstrukcija v obliki dendrograma. Podatkovni niz so homologna in poravnana zaporedja nukleotidov (A-adenin, C-citozin, G-gvanin, T-timin) štirih hipotetičnih organizmov (OTU1, OTU2, OTU3 in OTU4). Z modro barvo so označena monomorfna mesta, z zeleno barvo polimorfna mesta in s črtico so označene vrzeli (posledica insercij ali delecij). Razlikujemo dva osnovna pristopa h gradnji drevesa. Eden je varčnost, kjer iščemo najkrajše drevo, to je metoda največje varčnosti MP. Drugi pristop je s pomočjo substitucijskega modela evolucije. Model omogoča bodisi izračun genetskih razdalj, na katerih temeljita metodi združevanja sosedov NJ in metoda neuteženih povprečij UPGMA, bodisi izračun verjetnosti, na katerih temeljita metoda največjega verjetja ML in Bayesovo sklepanje BI.

Pristop z iskanjem drevesa zgradi množico dreves (potencialno vsa možna drevesa), optimalno drevo izbere na podlagi kriterija varčnosti. V pristopu iskanja optimalnega drevesa je razvita **metoda največje varčnosti MP**.

Postopki izračuna po metodah UPGMA, NJ in MP so predstavljeni v Poglavlju 13. Metoda ML je predstavljena v poglavju 14.3 ter s Primerom 14.4 in metoda BI v poglavju 14.4 ter s Primerom 14.5.

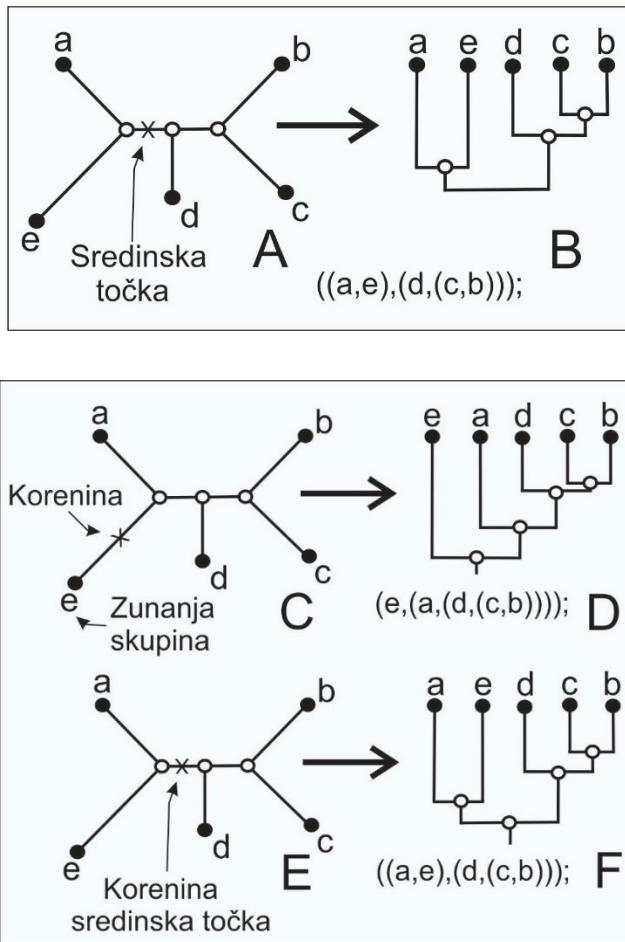
Koreninjeno vs. nekoreninjeno drevo. Drevo brez hipotetičnega zadnjega skupnega prednika je **nekoreninjeno** (ang. Unrooted) in je lahko neusmerjeno ali usmerjeno. Drevo lahko **koreninimo** (ang. Rooted) na dva načina. Med obravnavane taksonse vključimo **zunanjo skupino** (ang. Outgroup), ki si z notranjimi taksoni ne deli skupnih lastnosti in predstavlja korenino (Slika 14.14.C). Iz korenine, utemeljene na zunanji skupini, izhajata ena zunanja veja k zunanji razvejitvi z listom (zunanja skupina) in ena notranja veja k notranji razvejitvi. Iz preostalih notranjih razvejitev izhajajo veje k naslednjim notranjim ali zunanjim razvejitevam (Slika 14.14.D). Nekoreninjeno filogenetsko drevo lahko koreninimo z metodo **koreninjenja v sredinski točki** (ang. Midpoint Rooting). V neusmerjenem drevesu izračunamo razdalje med vsemi zunanjimi razvejitevami, poiščemo par z najdaljšo razdaljo in v sredini določimo korenino (Slika 14.14.E). Graf preuredimo s korenino na eni strani in njej nasproti so listi (Slika 14.14.F). Korenina podaja smer evolucijskega dogajanja in drevo je usmerjeno, tok genetskih informacij pa teče od korenine v smeri proti listom.

Topologijo drevesa lahko predstavimo grafično v obliki dendrograma (Slika 14.15). Druga možnost je predstavitev topologije drevesa v obliki **Newick zapisa** (ang. Newick Format). Odnosi med taksoni so predstavljeni s pomočjo oklepajev in so ločeni z vejico, zapis se konča s podpičjem. Sestrski taksona sta znotraj oklepaju, npr:

$$\text{Cluster} = (a,((f,g),(b,(c,(d,e)))));$$

Med taksoni lahko navedemo tudi dolžino vej na način, da za oznako taksona vstavimo dvopičje in dodamo dolžino veje (brez presledka), takšen zapis ima potem obliko:

$$\text{Cluster} = (a:0.525,((g:0.707,f:0.707):0.692,(b:0.804,((e:0.931,d:0.801):0.214,c:0.786):0.188):0.196):0.525):0;$$



Slika 14.14 Nekoreninjeno in koreninjeno drevo. A – nekoreninjeno in neusmerjeno drevo; z X je označena sredinska točka med najbolj oddaljenima listoma *e* in *b*; B – nekoreninjeno in usmerjeno drevo in Newick oblika zapisa; C – nekoreninjeno in neusmerjeno drevo, z X je označena veja za koreninjenje po metodi zunanje skupine, takson *e* predstavlja zunanjo skupino; D – koreninjeno in usmerjeno drevo po metodi zunanje skupine in Newick oblika zapisa; E – nekoreninjeno in neusmerjeno drevo, z X je označeno mesto za koreninjenje po metodi sredinske točke, ki je na sredini med najbolj oddaljenima listoma *e* in *b*; F – koreninjeno in usmerjeno drevo po metodi sredinske točke in Newick oblika zapisa. Razlika med koreninjenim in nekoreninjenim usmerjenim drevesoma je v številu notranjih razvejitvev, nekoreninjeno drevo ima 3, koreninjeno ima 4 notranje razvejitve. Notranje razvejitve so označena s praznimi krogi, zunanje razvejitve z listi so označena s polnimi krogi.

Ultrametrično vs. aditivno drevo. Drevo, pri katerem so dolžine vej od korenine do listov enaka za vse taksoni, je **ultrametrično** (ang. Ultrametric) (Slika 14.14.B). Ultrametrično drevo lahko uskladimo s časovno skalo, in s tem kalibriramo molekularno uro. Drevo, pri katerem se dolžine vej od korenine do listov razlikujejo med taksoni, je **aditivno** (ang. Additive) (Slika 14.14.D). V aditivnem drevesu dolžina vej ponazarja število mutacij in omogoča primerjave v frekvencah mutacij za posamezni takson. Oba tipa dreves sta zgrajena hierarhično in razlagata hipotezo o zaporedju cepitev filogenetskih linij.

Usmerjeno vs. neusmerjeno. Koreninjeno drevo, pri katerem korenina izhaja iz zadnjega skupnega prednika in predstavlja hierarhijo evlucijskih dogodkov in dihlotomnih cepitev prednikov na potomce, je **usmerjeno** (npr. filogram) (Slika 14.14.B, D). V primeru ultrametričnosti lahko drevesu priredimo časovno skalo. Nekoreninjeno drevo, pri katerem hierarhija dogodkov ni znana, je **ni usmerjeno** (npr. radialno drevo) (Slika 14.14.A). Izhodiščno notranjo razvejitvev neusmerjenega drevesa izberemo poljubno na enem ali drugem koncu drevesa. Iz njega izhajajo tri veje k potomcem, od teh dve zunanji veji k listom in ena notranja veja k sosednji notranji razvejitvi. Vsa preostale notranje razvejitve se od veje s prednikom nadalje razvejejo v dve veji potomcev, bodisi k naslednjim notranjim razvejitvam ali k listom (Slika 14.14.A). Izhodiščno razvejitvev nekoreninjenega, vendar usmerjenega drevesa določimo na način, da poiščemo lista, med katerima je dolžina vej najdaljša, v sredini te razdalje določimo točko, ki postane izhodišče. Graf preuredimo

z izhodiščem na eni strani in njemu nasproti so listi, s tem ponazorimo smer toka genetskih informacij in drevo usmerimo (Slika 14.14.D).

Vrste dreves. S filogenetskim drevesom prikažemo hipotezo o poteku evolucijskega dogajanja. Glede na računske postopke in vsebino, ki jo želimo z drevesom izpostaviti, razlikujemo več vrst dreves.

Dendrogram (ang. Dendrogram) je skupno ime za različne oblike grafov v obliki drevesa. Drevo lahko vsebuje korenino in se končuje z listi, to so taksoni (Slika 14.15). Topologija drevesa predstavlja hipotezo o poteku divergentnih cepitev in hierarhijo evolucijskih dogodkov, in s tem hierarhijo med evolucijskimi predniki in potomci. Dendrogram, zgrajen kot ultrametrično koreninjeno in usmerjeno drevo, lahko uskladimo s časovnim potekom dogodkov in ocenimo starost filogenetskih linij. Dolžina vej od korenine do lista je enaka za vse taksone in vgrajena je hipoteza o enaki stopnji mutacij za vse taksone (Slika 14.15.B).

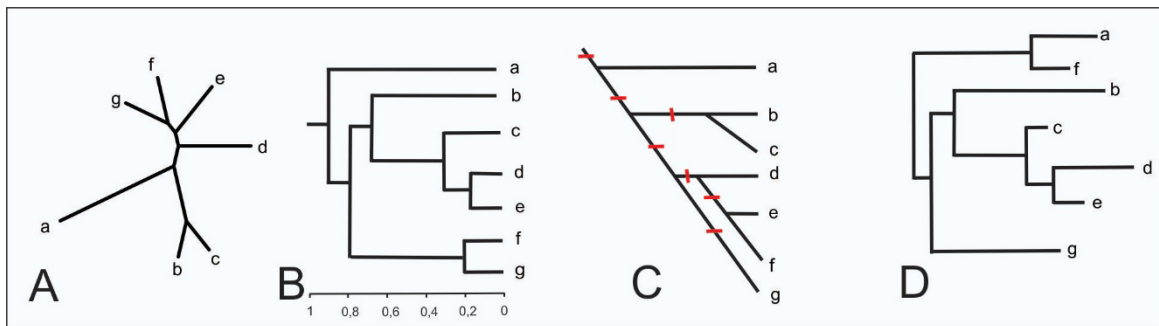
Radialno drevo (ang. Radial Tree) je aditivno in nekoreninjeno drevo, ki nima definirane smeri evolucijskega dogajanja, saj hierarhija evolucijskih dogodkov ni znana. Drevo je zgrajeno iz notranjih vej, ki povezujejo notranje razvejitev, in zunanjih vej, ki izhajajo iz notranjih razvejitev, in se končajo z listom – taksonom (Slika 14.15.A). Notranje veje praviloma dodajamo v obliki krožnice, iz katere izhajajo zunanje veje, zato radialno. Dolžina posamezne veje ponazarja število mutacij; topologija drevesa z razvejanostjo ponazarja sorodnost oz. podobnost med taksoni.

Filogram (ang. Phylogram) je ultrametrično koreninjeno drevo (Slika 14.15.B). Dolžina od korenine do lista je enaka za vse taksone, vgrajena je hipoteza o enaki stopnji mutacij med taksoni. Topologija drevesa predstavlja hipotezo o poteku divergentnih cepitev in hierarhijo med evolucijskimi predniki in potomci. Filogramu lahko priredimo informacijo o časovnem poteku dogodkov, in s tem kalibriramo molekularno uro.

Fenogram (ang. Phenogram) je drevo narejeno na fenetskih podatkih, ki ponazarja fenotipsko podobnost med taksoni. Grafična predstavitev fenograma je lahko v obliki ultrametričnega ali aditivnega drevesa, redkeje kot radialno drevo. Filogenetska informacija fenograma je praviloma močno omejena. Prisotnost filogenetske informacije v morfoloških znakih preverimo s testiranjem filogenetskega signala (glej Poglavlje 14.7).

Kladogram (ang. Cladogram) je drevo, s katerim predstavimo odnose med monofiletskimi skupinami. Na drevesu so označene sinapomorfije, ki so kriterij za oblikovanje skupin. Divergentne cepitve ne predstavljajo zadnjega skupnega prednika, ampak skupine, ki si delijo sinapomorfije po kladističnem pristopu (Slika 14.15.C).

Ni enoznačnega priporočila, katera metoda ali vrsta drevesa je primerna za analizo konkretnih podatkov. Izbira drevesa in postopek gradnje sta odvisna od namena in cilja, ki ga zasledujemo.



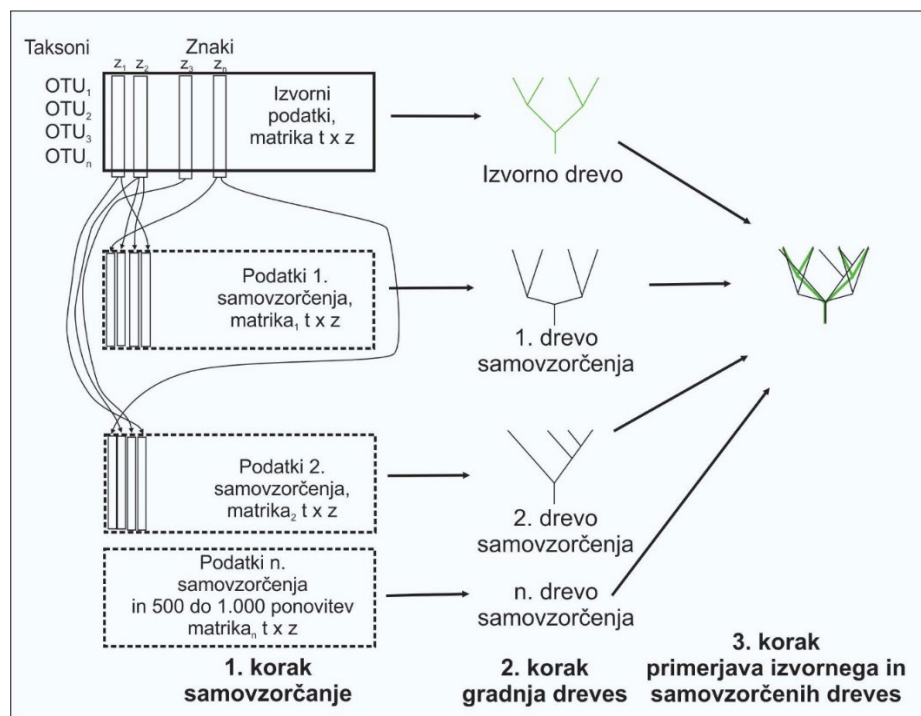
Slika 14.15 Filogenetska drevesa. A Radialno drevo, nekoreninjeno, aditivno drevo narejeno po metodi združevanja sosedov. B Filogram, koreninjeno ultrametrično drevo, narejeno po metodi UPGMA. C Kladogram z označenimi sinapomorfijami (rdeči prečniki), narejeno po metodi maksimalne varčnosti. D Nekoreninjeno aditivno drevo narejeno po metodi največjega verjetja.

14.6 Vrednotenje dreves

Kako dobro filogenetsko drevo povzame informacijo o medsebojnih odnosih iz izvorne matrike, ocenimo z izračunom podpore posameznim razvejitvam. Za oceno uspešnosti filogenetske rekonstrukcije poznamo različne metode, tukaj predstavljamo dve metodi po načelu samovzorčenja, to sta metoda bootstrap in metoda jackknife. Razvejitve z nezadostno podporo ostanejo nerazrešena in izkazujejo politomijo.

Bootstrap (ang. Bootstrap, sin. kljukčeva metoda) je metoda za oceno podpore razvejitvam v filogenetskem drevesu. Izvorno drevo primerjamo z drevesi, zgrajenimi z naključno razvrstitvijo znakov v postopku samovzorčenja. V prvi fazi algoritem iz izvorne matrike naključno izbere znak, ki postane prvi znak v samovzorčeni matriki. Sledi naslednja naključna izbira znaka iz celotnega nabora in ta znak postane naslednji znak v samovzorčeni matriki. Postopek naključnega izbiranja znakov se ponavlja, dokler število znakov ni enako izvorni matriki. Ker je izbor znakov naključen, je en znak lahko izbran večkrat, kakšen drug pa nikoli. V naslednji fazi algoritem iz posamezne samovzorčene matrike zgradi drevo po isti metodi, kot ga ocenjujemo. Postopek samovzorčenja in gradnje drevesa se ponovi od 500 do 1.000 krat. V tretji fazi ocenjevanja algoritem za vsako posamezno razvejitev v izvornem drevesu izračuna delež pojavljanja v samovzorčenih drevesih (Slika 14.16). Rezultat podpore posamezni razvejitvi je podan v deležu; 100 % podpora izbrani razvejitvi pomeni, da so vsa samovzorčena drevesa imela identično zgrajeno razvejitev. Takšna razvejitev izkazuje maksimalno verjetnost.

Jackknife (ang. Jackknife, sin. metoda pipca) je alternativna metoda za oceno podpore razvejitvam v filogenetskem drevesu. Izvorno drevo primerjamo z drevesi, zgrajenimi z naključno razvrstitvijo znakov v postopku samovzorčenja. V prvi fazi algoritem iz izvorne matrike naključno izbere polovico znakov. V drugi fazi algoritem iz posamezne samovzorčene matrike (s polovičnim naborom znakov) zgradi drevo po isti metodi, kot ga ocenjujemo. Postopek samovzorčenja in gradnje drevesa se ponovi od 500 do 1.000 krat. V tretji fazi ocenjevanja algoritem za vsako posamezno razvejitev v izvornem drevesu izračuna, v kolikšnem deležu se pojavi v samovzorčenih drevesih. Rezultat podpore posamezni razvejitvi je podan v deležu. 100 % podpora izbrani razvejitvi pomeni, da so vsa samovzorčena drevesa imela identično zgrajeno razvejitev. Takšna razvejitev izkazuje maksimalno verjetnost. Metodi bootstrap in jackknife se razlikujeta samo v prvem koraku v načinu samovzorčenja znakov.



Slika 14.16 Samovzorčenje pri oceni podpore filogenetske rekonstrukcije po metodi bootstrap (prirejeno po Felsenstein 1985). Postopek poteka v treh korakih: 1. v prvem koraku algoritem generira matrike po metodi samovzorčenja. Velikost matrik je $t \times z$ (t – takson, z – znak). 2. v drugem koraku algoritem iz vsake samovzorčene matrike zgradi drevo. 3. v tretjem koraku algoritem za vsako izvorno razvejitev posebej izračuna, koliko samovzorčenih dreves je imelo enako zgrajeno to razvejitev.

Drevo soglasja. Posamezna metoda filogenetske rekonstrukcije (npr. metoda maksimalne varčnosti) lahko vrne več dreves z enako dolžino oz. z enako varčnostjo. Raziskovalec mora med enakovrednimi rešitvami izbrati optimalno drevo. Pri tem si pomagamo z oblikovanjem **drevesa soglasja** (ang. Consensus Tree). Drevo soglasja ohrani razvejitev z največjo podporo. V **metodi strogega soglasja** (ang. Strict Consensus Tree) so ohranjene samo razvejitve s 100 % podporo, preostale razvejitve ostanejo nerazrešene in izkazujejo politomijo. Manj konservativne metode oblikujejo drevo soglasja po **pravilu**

večine (ang. Majority-rule Consensus Tree), v katerem določimo prag za ohranitev razvejitve (npr. 50 % podpora). Razvejitve, ki ne dosežejo praga, ostanejo nerazrešene in izkazujejo politomijo.

Bayesova statistika

Thomas Bayes (1701–1761) je bil angleški statistik, filozof in duhovnik. Formuliral je izrek, ki po njem nosi ime Bayesov izrek, s katerim je podal rešitev problema inverzne verjetnosti. Na podlagi predhodne verjetnosti in pogojne verjetnosti je pokazal, da lahko izračunamo posteriorno verjetnost.

Praktična uporaba izreka in računanja posteriorne verjetnosti se je začela v petdesetih letih dvajsetega stoletja z razvojem računalniške tehnologije.

Bayesov pristop postaja vse bolj uveljavljena statistična metodologija v znanstvenoraziskovalnem delu. Uveljavil se je predvsem pri reševanju kompleksnih in računsko zelo zahtevnih operacijah. Tudi pri iskanju filogenetskih odnosov in razlagi evolucijskega dogajanja na podlagi nukleotidnega zaporedja.



PRIMER 14.4

Filogenetski odnosi med belonogimi mišmi *Apodemus* – metoda največjega verjetja

Iz genske banke NCBI GeneBank (<https://www.ncbi.nlm.nih.gov/>) smo pridobili nukleotidna zaporedja za gen citokrom *b* MT-CYB, za šest vrst belonogih miši: *Apodemus agrarius*, *Apodemus epimelas*, *Apodemus flavicollis*, *Apodemus sylvaticus*, *Apodemus microps*, *Apodemus alpicola* (Preglednica 14.6). Nukleotidna zaporedja smo prenesli v program MEGA 11 in jih poravnali z rutino MUSCLE, po poravnavi in rezanju začetnih in končnih elementov je ostalo 1.009 nukleotidnih mest. Opisna statistika z lastnostmi poravnavi zaporedij je predstavljena v Preglednici 14.7. Nukleotidi niso enakomerno zastopani, saj med purini prevladuje adenin (povprečna frekvenca 31,1 %), med pirimidini pa timin (povprečna frekvenca 29,4 %) (Preglednica 14.7). Primerjave parov mest med zaporedji izkazuje prevladujočo frekvenco homonomnih parov (AA = 284, TT = 246, CC = 225 in GG = 117), druge kombinacije parov so bistveno redkeje zastopane (Preglednica 14.7).

Evolucijske odnose med belonogimi mišmi smo predstavili po dveh algoritmičnih metodah, najprej z genetskimi razdaljami in nato tudi s substitucijskim modelom. V prvem metodološkem pristopu smo poravnana zaporedja vključili v izračun Jukes-Cantorjeve razdalje (Preglednica 14.9), povprečna JC = 0,15. V izračun in izdelavo topologije drevesa po metodi združevanja sosedov smo vključili JC razdalje (Slika 14.17.A).

Preglednica 14.6 Informacije o pridobljenih nukleotidnih zaporedjih za gen MT-CYB za šest vrst rodu *Apodemus*

| Vrsta | Haplotip | Pristopna številka | Dolžina bp | Vir |
|-----------------------------|---------------|--------------------|------------|--|
| <i>Apodemus epimelas</i> | Hap29 | KR0031081 | 1076 | Darvish J., Mohammadi Z., Ghorbani F., Mahmoudi A., Dubey S. Phylogenetic relationships of <i>Apodemus</i> Kaup, 1829 (Rodentia: Muridae) species in the eastern Mediterranean inferred from mitochondrial DNA, with emphasis on Iranian species. J. Mammal. Evol. 22, 583-595 (2015). |
| <i>Apodemus agrarius</i> | Fuyuan-Aa-145 | KJ857286 | 1144 | Wang C.Q., Gao J.H., Li M., Guo W.P., Lu M.Q., Wang W., Hu M.X., Li M.H., Yang J., Liang H.J., Tian X.F., Holmes E.C., Zhang Y.Z. Co-circulation of Hantaan, Kenkeme, and Khabarovsk Hantaviruses in Bolshoy Ussuriysky Island, China. Virus Res. 191, 51-58 (2014). |
| <i>Apodemus flavicollis</i> | JRM-219 | AJ605631 | 1144 | Michaux J.R., Libois R., Paradis E., Filippucci M.G. Phylogeographic history of the yellow-necked fieldmouse (<i>Apodemus flavicollis</i>) in Europe and in the Near and Middle East. Mol. Phylogenet Evol 32 (3), 788-798 (2004). |
| <i>Apodemus sylvaticus</i> | AB033695 | AB033695.1 | 1140 | Suzuki H., Tsuchiya K., Takezaki N. A molecular phylogenetic framework for the Ryukyu endemic rodents <i>Tokudaia osimensis</i> and <i>Diplobrix legata</i> . Mol. Phylogenet Evol 15 (1), 15-24 (2000). |
| <i>Apodemus microps</i> | | AF159393.1 | 1144 | Martin Y., Gerlach G., Schlotterer C., Meyer A. Molecular phylogeny of European muroid rodents based on complete cytochrome b sequences. Mol Phylogenet Evol 16 (1), 37-47 (2000). |
| <i>Apodemus alpicola</i> | HS596 | AB032854.1 | 1140 | Serizawa K., Suzuki H., Tsuchiya K. A phylogenetic view on species radiation in <i>Apodemus</i> inferred from variation of nuclear and mitochondrial genes. Biochem Genet 38 (1-2), 27-40 (2000). |

V drugem metodološkem pristopu smo najprej s testom razmerja verjetij razvrstili substitucijske modele glede na izgubo informacije. Model z najmanjšo izgubo je bil GTR+I, prvih pet najboljših modelov je predstavljenih v Preglednici 14.8. V izračun in izdelavo drevesa največjega verjetja smo vključili parametre iz substitucijskega modela GTR+I (Slika 14.17.B).

Oba metodološka postopka smo izvedli v programu MEGA 11 (Tamura in sod. 2021).

Preglednica 14.7 Deleži nukleotidov pri mitohondrijskem genu citokrom *b* pri šestih vrstah rodu *Apodemus*; skupno število poravnanih nukleotidnih mest in povprečne vrednosti

| Oznaka | Vrsta / Nukleotid | T | C | A | G | Total |
|--------|-----------------------------|------|------|------|------|-------|
| APOEPI | <i>Apodemus epimelas</i> | 29,5 | 26,9 | 31,6 | 12,0 | 1009 |
| APOAGR | <i>Apodemus agrarius</i> | 29,4 | 27,7 | 29,8 | 13,1 | 1009 |
| APOSYL | <i>Apodemus sylvaticus</i> | 30,7 | 26,3 | 30,6 | 12,4 | 1009 |
| APOFLA | <i>Apodemus flavicollis</i> | 29,1 | 27,3 | 31,2 | 12,4 | 1009 |
| APOMIC | <i>Apodemus microps</i> | 28,8 | 27,2 | 31,8 | 12,2 | 1009 |
| APOALP | <i>Apodemus alpicola</i> | 28,7 | 27,5 | 31,2 | 12,6 | 1009 |
| | Povprečje | 29,4 | 27,1 | 31,1 | 12,4 | 1009 |

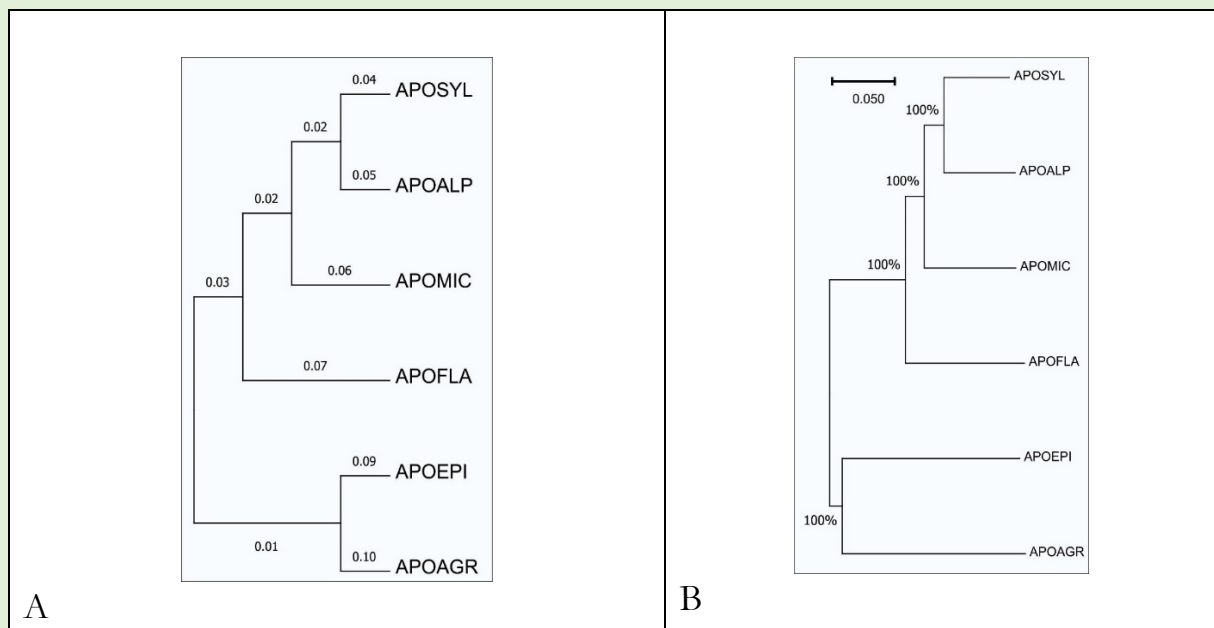
Preglednica 14.7 Frekvence parov in metrika mutacij med šestimi zaporedji belonogih miši *Apodemus*. Povprečja so podana za vseh šest zaporedij in 1.009 mest; ii = identični pari, si = tranzicijski pari, sv = transverzijijski pari, R = si/sv

| Metrika | Povprečje |
|---------|-----------|
| ii | 872 |
| si | 90 |
| sv | 47 |
| R | 1,92 |
| TT | 246 |
| TC | 35 |
| TA | 12 |
| TG | 0 |
| CT | 40 |
| CC | 225 |
| CA | 8 |
| CG | 1 |
| AT | 12 |
| AC | 12 |
| AA | 284 |
| AG | 7 |
| GT | 1 |
| GC | 0 |
| GA | 8 |
| GG | 117 |

Preglednica 14.8 Razporeditev substitucijskih modelov glede na najnižjo vrednost verjetja lnL v prileganju modelov k podatkom; število parametrov v modelu ter vrednosti BIC in AICc

| Model | Parametri | BIC | AICc | lnL |
|----------|-----------|--------|--------|---------|
| GTR+I | 18 | 6481,5 | 6360,9 | -3162,4 |
| GTR+G+I | 19 | 6484,8 | 6357,4 | -3159,7 |
| TN93+G | 15 | 6500,7 | 6400,1 | -3185,0 |
| TN93+G+I | 16 | 6509,3 | 6402,1 | -3185,0 |
| TN93+I | 15 | 6513,7 | 6413,2 | -3191,6 |

Izris filogenetskih dreves po dveh različnih postopkih (na podlagi genetskih razdalj in substitucijskega modela) vrne identično topologijo (Slika 14.17). Obe drevesi razvrstita vrste belonogih miši v dve skupini, ki ustrezata trem podrodovom. Prvo skupino sestavljata podrod *Karstomys* z *Apodemus epimelas* in podrod *Apodemus* z *Apodemus agrarius*. Drugo skupino sestavljajo vrste podrodu *Sylvaemus* s štirimi vrstami: *Apodemus sylvaticus*, *Apodemus alpicola*, *Apodemus microps* in *Apodemus flavicollis*.



Slika 14.17 Evolucijski odnosi med šestimi vrstami rodu *Apodemus*. **A** Filogenetsko drevo, zgrajeno na Jukes-Cantorjevih razdaljah po metodi združevanja sosedov, vrednost ob veji podaja njeno dolžino. **B** Filogenetsko drevo, zgrajeno po modelu GTR+I in z metodo največjega verjetja. Merilna skala je v enotah števila substitucij. Vrednosti ob razvejitvah podajajo podporo izkazano z bootstrap testiranjem s 500 ponovitvami.

Preglednica 14.9 Jukes-Cantorjeve razdalje med šestimi vrstami rodu *Apodemus* (razlaga oznak za vrste je v Preglednici 14.7)

| | APOEPI | APOAGR | APOSYL | APOFLA | APOMIC | APOALP |
|--------|--------|--------|--------|--------|--------|--------|
| APOEPI | 0 | | | | | |
| APOAGR | 0,1754 | 0 | | | | |
| APOSYL | 0,1807 | 0,1782 | 0 | | | |
| APOFLA | 0,1667 | 0,1754 | 0,1260 | 0 | | |
| APOMIC | 0,1667 | 0,1830 | 0,1354 | 0,1062 | 0 | |
| APOALP | 0,1742 | 0,1606 | 0,1295 | 0,0882 | 0,1085 | 0 |



PRIMER 14.5

Filogenetski odnosi med belonogimi mišmi *Apodemus* – Bayesovo sklepanje

Bayesovo sklepanje temelji na posteriorni verjetnosti in je alternativa metodi največjega verjetja, predstavljeni v Primeru 14.4. Tudi v tem primeru smo analizirali filogenetske odnose med šestimi vrstami belonogih miši rodu *Apodemus*. Pričakovani rezultat je topologija filogenetskega drevesa in dolžina njegovih vej. Izračun Bayesovega sklepanja poteka v več korakih, ki jih izvajamo z različnimi programi. Zato so v nadaljevanju predstavljeni koraki, programi s katerimi so izvedeni, in oblika računalniškega zapisa.

1. Korak: zaporedja, program MEGA in NEXUS oblika

Poravnana nukleotidna zaporedja mitohondrijskega gena MT-CYB so identična kot v metodi največjega verjetja in navedena v Preglednici 14.6. Zaporedja smo prenesli v program MEGA 11 (Tamura in sod. 2021) in jih poravnali z rutino MUSCLE. Po poravnavi ter rezanju začetnih in končnih elementov je ostalo 1.009 nukleotidnih mest, zapis smo shranili v NEXUS oblika (Slika 14.18).

```
#NEXUS
[ TITLE Phylogenetic Analysis]
BEGIN TAXA;
  DIMENSIONS NTAX=6;
  TAXLABELS
  APOEPI
  APOAGR
  APOFLA
  APOSYL
  APOMIC
  APOALP
;
END;
BEGIN CHARACTERS;
  DIMENSIONS NCHAR=1009;
  FORMAT MISSING=? GAP=- MATCHCHAR=-. datatype=nucleotide;
MATRIX
APOEPI AAAACACATCCATTAATAAAAAATTATTAATCACTC/ /CCTAACTGGAATTGGCGGCAACC
APOAGR AAAACTCACCCCTATTAAAAATGATAACCATTC/ /CCCTACCTGAATGGCGGCAACC
APOFLA AAAACACACCCCACTACTAAAAATTAATCACTC/ /TCATCATGAATGGAGGCAACC
APOSYL AAAACACACCCCACTACTAAAAATTAATCACTC/ /CCTAACTGGAATTGGAGGCAACC
APOMIC AAAACACACCCCTATTAAAAATTAATCACTC/ /CCTAACTGGAATTGGAGGCAACC
APOALP AAAACACACCCCACTACTAAAAATTAATCACTC/ /TCATCACTGAATGGAGGCAACC
;
END;
```

Slika 14.18 NEXUS oblika zapisa z 1.009 nukleotidnimi zaporedji za 6 vrst; zapis sestavlja blok z imeni taksonov (oznake vrst so razložene v Preglednici 14.7) in blok s podatki (navedeni so samo začetni in končni nukleotidi).

2. Korak: substitucijski model in MCMC, program BEAUti in XML oblika

NEXUS zapis smo uvozili v program BEAUti (Bayesian Evolutionary Analysis Utility, v1.10.4) (Drummond in sod. 2018a). V tem koraku izberemo parametre filogenetskega modela: 1. substitucijski model je GTR+I (model smo izbrali že v Primeru 14.4 Preglednica 14.8). 2. strukturne lastnosti evolucijskega drevesa, izbrali smo koreninjeno drevo po Yule-vev enostavnem ničelnem modelu speciacije. 3. vrednosti markovske verige Monte Carlo MCMC algoritma (Markov Chain Monte Carlo), izbrali smo 10.000.000 vzorčenj z metodo Metropolis in zažig prvih 1.000 vzorčenj. Parametre modela smo shranili v datoteko oblike .XML.

3. Korak: vzorčenje pokrajine verjetnosti, program BEAST ter oblike log, ops in trees

Parametre modela smo uvozili v program BEAST (Bayesian Evolutionary Analysis Sampling Trees, v1.10.4) (Drummond in sod. 2018b). Program izvede vzorčenje pokrajine verjetnosti in izračuna parametre z največjo posteriorno verjetnostjo. Rezultati analize so shranjeni v treh datotekah oblik .log, .ops in .trees.

Datoteka .log vsebuje navedbe vseh parametrov, vključenih v analizo, in njihove vrednosti za vsako točko vzorčenja (Preglednica 14.10).

Datoteka .ops vsebuje informacije o poteku operacije.

Datoteka .trees vsebuje NEXUS obliko zapisa za vsa drevesa, drevesa so navedena v Newick obliki. Slednja datoteka z informacijo o topologiji dreves in dolžino vej je ključni rezultat analize; datoteko peljemo naprej v vizualizacijo drevesa.

4. Korak: izbor drevesa, program TreeAnnotator in NEXUS oblika

Datoteko z drevesi (.trees) smo uvozili v program TreeAnnotator, v1.10.4 (Rambaut in Drummond 2018). Program povzame informacije o drevesih, izdelanih v predhodnem koraku, in izdelava »tarčno drevo« (ang. »Target Tree«). Rezultat je zapisan v NEXUS datoteki, ki vsebuje topologijo drevesa in dolžine vej za izbrano drevo z najvišjo posteriorno verjetnostjo. Zapis drevesa je v Newick obliki in ima vrednost: `tree_STATE_0 = ((APOAGR:22.4, APOEPI:22.45):116.88, ((APOMIC:29.52, (APOSYL:2.51, APOALP:2.51):27.01):73.69, APOFLA:103.22):36.11);`

Preglednica 14.10 Vrednosti parametrov območja z najvišjo posteriorno verjetnostjo za vsako tisočo meritev izmed deset tisoč najverjetnejših

| state | joint | prior | likelihood | gtr,rates,rateAC | gtr,rates,rateAG | gtr,rates,rateAT | gtr,rates,rateCG | gtr,rates,rateCT | gtr,rates,rateGT | frequencies1 | frequencies2 | frequencies3 | frequencies4 | pInv |
|-------|-------|-------|------------|------------------|------------------|------------------|------------------|------------------|------------------|--------------|--------------|--------------|--------------|------|
| 0 | -4976 | -876 | -4101 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 0,25 | 0,25 | 0,25 | 0,25 | 0,50 |
| 1000 | -3520 | -7 | -3513 | 1,00 | 1,00 | 0,98 | 1,00 | 1,01 | 1,01 | 0,24 | 0,26 | 0,25 | 0,25 | 0,65 |
| 2000 | -3427 | -3 | -3425 | 1,00 | 1,01 | 0,98 | 1,00 | 1,01 | 1,01 | 0,25 | 0,28 | 0,22 | 0,25 | 0,59 |
| 3000 | -3401 | -2 | -3399 | 0,99 | 1,02 | 0,98 | 1,00 | 1,01 | 1,00 | 0,26 | 0,27 | 0,20 | 0,27 | 0,66 |
| 4000 | -3400 | -3 | -3397 | 0,99 | 1,01 | 0,99 | 1,03 | 1,00 | 0,99 | 0,25 | 0,27 | 0,20 | 0,28 | 0,65 |
| 5000 | -3375 | -2 | -3374 | 0,99 | 1,00 | 0,98 | 1,03 | 1,02 | 0,98 | 0,25 | 0,26 | 0,18 | 0,31 | 0,63 |
| 8000 | -3369 | -1 | -3368 | 0,99 | 1,03 | 0,98 | 1,02 | 1,02 | 0,95 | 0,24 | 0,28 | 0,16 | 0,32 | 0,62 |
| 6000 | -3369 | -3 | -3365 | 0,99 | 1,03 | 0,98 | 1,02 | 1,02 | 0,96 | 0,25 | 0,27 | 0,17 | 0,32 | 0,64 |
| 7000 | -3367 | -3 | -3364 | 0,98 | 1,02 | 0,98 | 1,03 | 1,02 | 0,96 | 0,24 | 0,27 | 0,17 | 0,32 | 0,61 |
| 9000 | -3350 | -2 | -3348 | 0,99 | 1,03 | 0,99 | 1,02 | 1,02 | 0,95 | 0,27 | 0,28 | 0,14 | 0,32 | 0,62 |

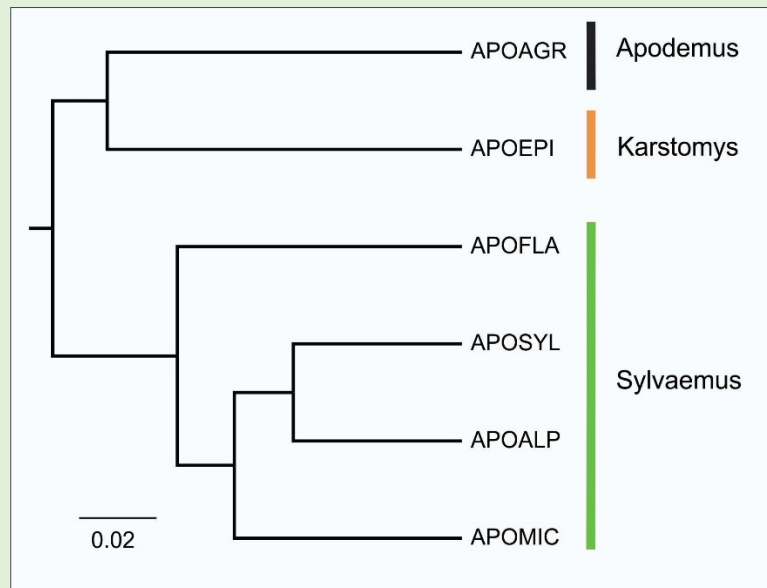
5. Korak: izris drevesa, program FigTree

Zadnji korak je vizualizacija filogenetskega drevesa, ki smo ga izvedli s programom FigTree (Tree Figure Drawing Tool, v 1.4.4) (Rambaut 2018). Datoteko .trees iz predhodnega koraka uvozimo v program FigTree. Program omogoča različne grafične učinke in tehnično ureditev drevesa. Končno verzijo drevesa izvozimo kot sliko v obliki .jpg zapisa ali kot .pdf datoteko.

Interpretacija drevesa

Metoda Bayesovega sklepanja zgradi drevo, v katerem so vrste združene v dve skupini. Prva skupina združuje vrste *A. agrarius* in *A. epimelas*, med katerima je globoka divergenca in pripadata podrodovoma *Apodemus* in *Karstomys*. V drugi skupini so združene vrste podrodu *Sylvaemus* (*A. flavicollis*, *A. sylvaticus*, *A. alpicola* in *A. microps*) (Slika 14.19).

Na istem nizu podatkov, na zaporedju 1.009 nukleotidov za gen MT-CYB, smo izvedli filogenetske rekonstrukcije po treh metodah: združevanja sosedov in največjega verjetja (predstavljeno na Sliki 14.17A in B) ter Bayesovega sklepanja (Slika 14.19). Vse tri metode vrnejo identično topologijo drevesa. V dolžini vej med vrstama APOAGR in APOEPI se razlikuje drevo, narejeno po metodi združevanja sosedov. Obe metodi, ki temeljita na verjetnosti, zgradita identični drevesi.



Slika 14.19 Evolucijski odnosi med šestimi vrstami rodu *Apodemus*. Filogenetsko drevo, zgrajeno po substitucijskem modelu GTR+I in z metodo Bayesovega sklepanja. Merilna skala je v enotah števila substitucij. (Razlaga oznake vrst je v Preglednici 14.7.)



14.7 Filogenetski signal

V makroekoloških študijah včasih preverjamo hipoteze, ali fenetski znaki vsebujejo filogenetsko informacijo. Pri vrednotenju izhajamo iz domneve, da so si vrste, ki si delijo bližje prednike, bolj podobne kot vrste, ki imajo bolj oddaljene evolucijske prednike. Z drugimi besedami, če primerjamo dve filogenetsko sorodni vrsti zanju domnevamo, da sta si ti bolj podobni, kot dve naključno izbrani vrsti. Tesno sorodne vrste izkazujejo podobne morfološke lastnosti, ta fenomen imenujemo **filogenetski signal** (ang. Phylogenetic Signal).

Prisotnost filogenetskega signala v fenetskih znakih povezuje morfologijo vrst z njihovo evolucijo. Povezanost merimo s K statistiko, ki primerja opazovani signal v morfoloških znakih z modelom Brownovega gibanja v evoluciji.

Filogenetske odnose, rekonstruirane na zaporedjih nukleotidov, privzamemo kot dejanski potek evolucije. Odnose med vrstami, ki jih izkazujejo morfološki znaki, primerjamo z odnosi na genetskem materialu. Pri tem genetski rezultati predstavljajo ničelno stanje. Cenilko jakosti filogenetskega signala K izračunamo s formulo:

$$K = \frac{MSE0}{MSE},$$

kjer je MSE povprečna vrednost kvadrata napake (ang. Mean Squared Error), izračunane iz variančno-kovariančne matrike za filogenetsko drevo, narejeno na morfoloških znakih. MSE0 je povprečna vrednost napake, izračunane iz variančno-kovariančne matrike za filogenetsko drevo, narejeno na nukleotidnih zaporedjih (Blomberg in sod. 2003).

Vrednost cenilke K okrog ena ($K \approx 1$) ustreza Brownovemu gibanju, ki implicira prisotnost filogenetskega signala. Vrednost K blizu nič ($K \approx 0$) ustreza naključnemu konvergentnemu vzorcu evolucije, tak morfološki znak ne vsebuje filogenetskega signala. Naraščanje K nad vrednostjo ena ($K > 1$) ustreza močnemu filogenetskemu signalu in konservativnosti morfološkega znaka.

Izračun ocene o filogenetski informaciji v fenetskih, morfoloških lobanjskih znakih belonogih miši je predstavljen v Primeru 14.6.



PRIMER 14.6

Filogenetski signal v morfoloških znakih pri belonogih miših *Apodemus*

Prisotnost in jakost filogenetskega signala v morfoloških znakih pomeni višjo podobnost med evolucijsko sorodnimi vrstami in omogoča filogenetske rekonstrukcije na fenetskih znakih. Filogenetsko rekonstrukcijo na nukleotidih interpretiramo kot najboljšo možno razlago evolucijskega razvoja in evolucijskih odnosov med obravnavanimi taksoni. Pri razlagi filogenetskega signala preverjamo podobnost med fenetskimi odnosi na morfoloških znakih in topologija drevesa, narejeno na nukleotidnih zaporedjih.

V oceno filogenetskega signala smo vključili dva niza podatkov. Prvi niz je filogenetsko drevo, narejeno na nukleotidnih zaporedjih po metodi največjega verjetja (Primer 14.4). Newick format drevesa je imel zapis in dolžino vej: (((APOSYL: 0.04238277, APOALP: 0.04700266): 0.01610493, APOURA: 0.05718245): 0.01620074, APOFLA: 0.06851080, (APOEPI: 0.08513733, APOAGR: 0.09592254): 0.04579355);. Vključena zaporedja in postopek poravnave ter izdelave drevesa so opisani v Primeru 14.4; analize smo izvedli s programom MEGA 11 (Tamura in sod. 2021). Drugi niz podatkov so enorazsežne morfometrične meritve dolžin lobanje, podane v Preglednici 13.7.

Testiranje filogenetskega signala v morfometričnih znakih lobanje šestih vrst belonogih miših rodu *Apodemus* smo izvedli v programskem okolju R s paketom »picante« in z ukazom *multiPhylosignal()* (Kembel in sod. 2010).

Prisotnost filogenetskega signala smo preverjali pri 13 lobanjskih spremenljivkah. Pet spremenljivk je imelo vrednost K višjo od ena ($K > 1$). Najvišjo vrednost $K = 1,3$ je imel znak VL14 – interorbitalna širina, in edini dosega statistično značilen vpliv ($p < 0,05$). Zanj sklepamo, da vsebuje značilno filogenetsko informacijo (Preglednica 14.11).

Splošen zaključek je, da enorazsežne meritve lobanje vsebujejo zelo skromen filogenetski signal. V predstavitvi evlucijskih odnosov med belonogimi mišmi, kot jih izkazujejo lobanjski znaki, predstavljeni na Sliki 13.4, je vključene zelo malo filogenetske informacije. Zato drevo ne izkazujejo dejanskih filogenetskih odnosov, ampak je močno zaznamovano z modifikacijami in prilagoditvami v okviru ekološke niše.

Preglednica 14.11 Statistika K preverjanja filogenetskega signala pri belonogih miših v lobanjskih znakih. Znaki so predstavljeni na Sliki 13.4. Višja vrednost K pomeni močnejši signal, vrednost p je signifikantnost, dosegel jo je samo znak interorbitalna širina VL14

| Znak | K | p |
|------|-------|--------|
| VL1 | 0,971 | 0,2995 |
| VL2 | 1,080 | 0,1220 |
| VL3 | 0,862 | 0,5135 |
| VL4 | 1,124 | 0,1200 |
| VL7 | 0,741 | 0,7035 |
| VL8 | 1,001 | 0,2510 |
| VL9 | 1,019 | 0,1870 |
| VL13 | 0,924 | 0,3750 |
| VL14 | 1,297 | 0,0280 |
| VL16 | 0,997 | 0,2790 |
| VL17 | 0,999 | 0,2095 |
| VL18 | 0,971 | 0,3405 |
| VL19 | 0,903 | 0,3495 |

```
R > package = »picante«, »phangorn«
# Filogenetski signal. Uvoz drevesa in znakov.
drevoapo<-read.tree(file = "apodemusdrevo")
drevoapo #Filogenetsko drevo
traitApodemus #Fenetski znaki
library(phangorn) #Paket phangorn za koreninjenje drevesa
korendrevoApo<-midpoint(drevoapo)
korendrevoApo
plot(korendrevoApo)
library(picante) #Paket pikante za izračun K statistike
signalApo<-multiPhylosignal(traitApodemus, korendrevoApo)
signalApo
```



Zaključki

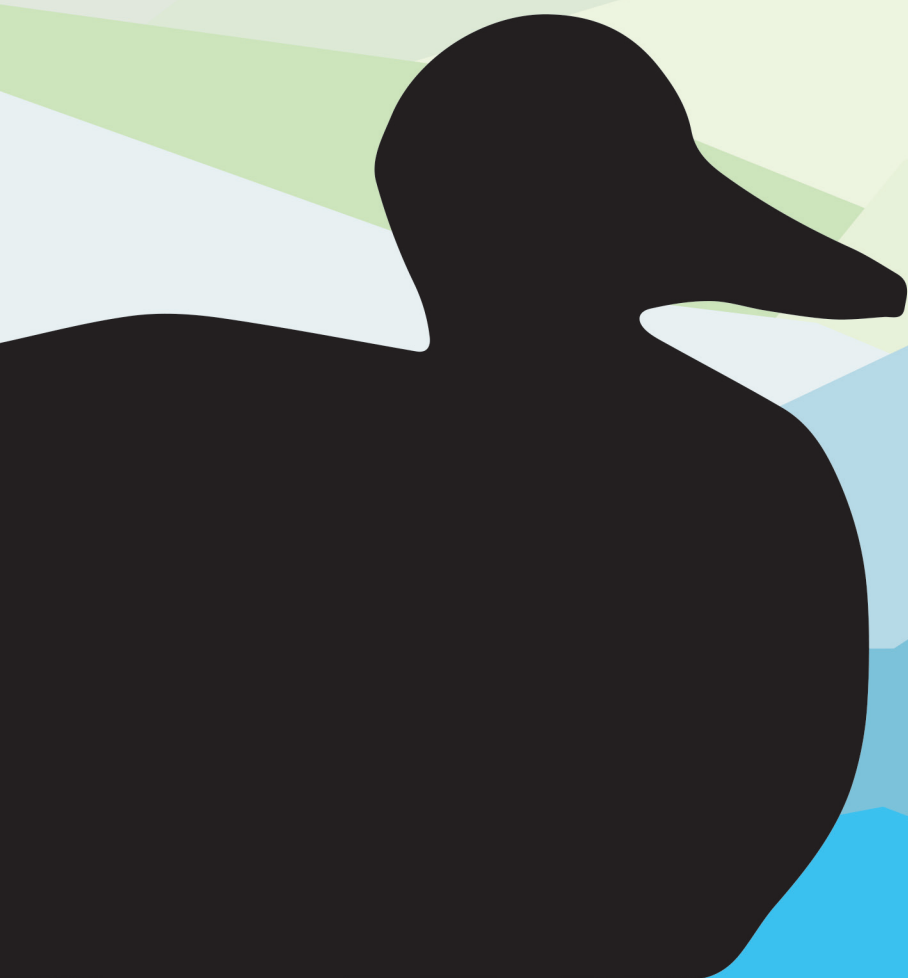
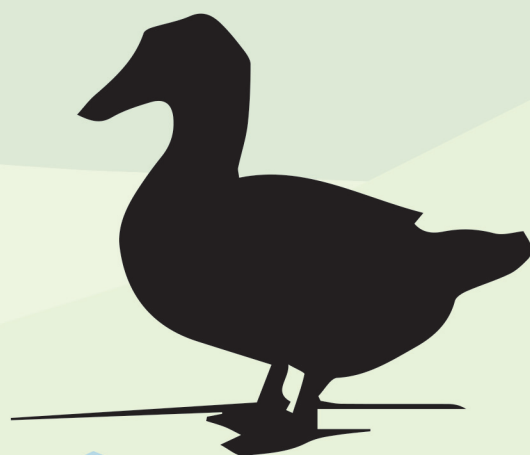
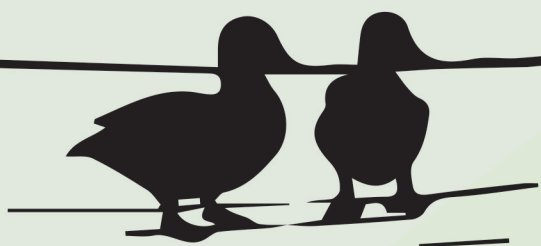
- Filogenetika je veda o evoluciji živih bitij in o evlucijskih odnosih znotraj skupin in med njimi.
- Evlucijske odnose analiziramo z metodami računske filogenetike in filogenetskega sklepanja, z namenom izdelave filogenetskega drevesa. Drevo je graf, s katerim ponazorimo najverjetnejši potek evolucije.
- V filogenetskih rekonstrukcijah analiziramo zaporedja nukleotidov v izbrani regiji DNK ali zaporedja aminokislin v izbranem proteinu.
- Analiza nukleotidnega zaporedja vključuje: izbiro in pridobivanje nizov zaporedij iz homologne genomske regije primerjanih vrst; poravnavo homolognih nukleotidov; oblikovanje substitucijskega modela, s katerim razložimo opazovani vzorec nukleotidnega zaporedja; izračun genetske razdalje med dvema nukleotidnima zaporedjema; izračun in gradnjo filogenetskega drevesa.
- Algoritmi za izračun genetskih razdalj (npr. Jukes-Cantorjeva razdalja) temeljijo na frekveni nukleotidnih baz v genomu in stopnji substitucije.
- Filogenetske odnose med vrstami predstavimo s filogenetskim drevesom. Drevo praviloma gradimo z algoritmičnim pristopom po metodah, ki temeljijo na genetskih razdaljah (metoda združevanja sosedov in metoda UPGMA) ali na verjetnosti (metoda največjega verjetja in metoda Bayesovega sklepanja).
- Glede na informacijo, ki je zajeta v filogenetskem drevesu, razlikujemo: radialno drevo je nekoreninjeno in aditivno, dolžine vej ponazarjajo število mutacij; filogram je koreninjeno in ultrametrično drevo, ki predstavlja hierarhični potek divergentnih cepitev v nove taksonse in ga lahko uskladimo s časovno skalo; kladogram oblikuje monofiletske skupine in ima označene sinapomorfije; fenogram je predstavitev fenetskih odnosov med taksoni.
- Filogenetske odnose med vrstami, kot jih izkazujejo morfološke lastnosti, primerjamo z odnosi, kot jih izkazujejo nukleotidi. Povezanost med morfologijo in nukleotidi ocenimo s testiranjem filogenetskega signala.
- Nukleotidna zaporedja so dostopna v genskih bankah npr. NCBI – National Center for Biotechnology Information, ENA – European Nucleotide Archive, EMBL – European Bioinformatics Institute.

Literatura

- Anisimova M. (ed.) 2019. *Evolutionary Genomics. Statistical and Computational Methods*. Second Edition. Swiss Institute of Bioinformatics. Lausanne, Switzerland.
- Bánki O., Roskov Y., Döring M., Ower G., Vandepitte L., Hobern D., Remsen D., Schalk P., DeWalt R.E., Keping M., Miller J., Orrell T., Aalbu R., Adlard R., Adriaenssens E.M., Aedo C., Aescht E., Akkari N., Alexander S., et al. 2022. Catalogue of Life Checklist (Annual Checklist 2022). Catalogue of Life. <https://doi.org/10.48580/dfq8>.
- Bayes' theorem. 2023. Wikipedia. Pridobljeno 1. 3. 2023, iz https://en.wikipedia.org/wiki/Bayes%27_theorem.
- Bayesian inference. 2023. Wikipedia. Pridobljeno 1. 3. 2023, iz https://en.wikipedia.org/wiki/Bayesian_inference.
- Blomberg S.P., Garland T., Ives A.R. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*, 57: 717–745.
- Catalogue of Life Checklist. 2022. Catalogue of LifeLeiden, NL. Pridobljeno dne 20. 11. 2022 iz: <https://www.checklistbank.org/dataset/9837>.
- Computational phylogenetics. 2023. Wikipedia. Pridobljeno 28. 1. 2023, iz https://en.wikipedia.org/wiki/Computational_phylogenetics.
- Conditional probability 2023. Wikipedia. Pridobljeno 1. 3. 2023, iz https://en.wikipedia.org/wiki/Conditional_probability.
- Dray S., Dufour A.-B., Thioulouse J. 2020. ade4. Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences. R package version 2.5-6. <https://cran.r-project.org/web/packages/ade4/index.html>.
- Drummond J.A., Rambaut A., Suchard A.M. 2018b. BEAST. Bayesian Evolutionary Analysis Sampling Trees, v1.10.4. <https://beast.community/programs>.
- Drummond J.A., Rambaut A., Suchard A.M., Xie W. 2018a. BEAUti. Bayesian Evolutionary Analysis Utility, v1.10.4. <https://beast.community/programs>.
- Evolution. 2023. Wikipedia. Pridobljeno 28. 1. 2023, iz <https://en.wikipedia.org/wiki/Evolution>.
- Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 39: 783–791.
- Freeman S., Herron C.J. 2004. *Evolutionary Analysis*. Pearson Education Inc. Upper Saddle River, NJ.
- Genetic drift. 2023. Wikipedia. Pridobljeno 28. 1. 2023, iz https://en.wikipedia.org/wiki/Genetic_drift.
- Goslee S.C., Urban D.L. 2007. The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, 22(7): 1–19. DOI:10.18637/jss.v022.i07.
- Hall G.B. 2011. *Phylogenetic Trees Made Easy. A How-to Manual*. Sinauer Associates. Sunderland.
- Joint probability distribution. 2023. Wikipedia. Pridobljeno 1. 3. 2023, iz https://en.wikipedia.org/wiki/Joint_probability_distribution.
- Kembel S.W., Cowan P.D., Helmus M.R., Cornwell W.K., Morlon H., Ackerly D.D., Blomberg S.P., Webb C.O. 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26: 1463–1464.
- Lemey P., Salemi M., Vandamme A.-M. (eds.). 2009. *The Phylogenetic Handbook: a Practical Approach to Phylogenetic Analysis and Hypothesis Testing*. Cambridge University Press. Cambridge.
- Liem K.F., Bemis W.E., Walker W.F.Jr., Grande L. 2001. *Functional Anatomy of the Vertebrates. An Evolutionary Perspective*. Harcourt College Publishers. Orlando.
- LUCA. 2023. Wikipedia. Pridobljeno 15. 2. 2023, iz https://en.wikipedia.org/wiki/Last_universal_common_ancestor.
- Markov chain Monte Carlo. 2023. Wikipedia. Pridobljeno 1. 3. 2023, iz https://en.wikipedia.org/wiki/Markov_chain_Monte_Carlo.
- Metropolis–Hastings algorithm. 2023. Wikipedia. Pridobljeno 1. 3. 2023, iz https://en.wikipedia.org/wiki/Metropolis%E2%80%93Hastings_algorithm.
- Modern synthesis. 2023. Wikipedia. Pridobljeno 15. 2. 2023, iz [https://en.wikipedia.org/wiki/Modern_synthesis_\(20th_century\)](https://en.wikipedia.org/wiki/Modern_synthesis_(20th_century)).
- Mutation. 2023. Wikipedia. Pridobljeno 15. 2. 2023, iz <https://en.wikipedia.org/wiki/Mutation>.
- Natural selection. 2023. Wikipedia. Pridobljeno 28. 1. 2023, iz https://en.wikipedia.org/wiki/Natural_selection.
- Nei M., Kumar S. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press. New York.
- Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., Wagner H. 2019. vegan: Community Ecology Package. R package version 2.5-6. <https://CRAN.R-project.org/package=vegan>.
- Phylogenesis. 2023. Wikipedia. Pridobljeno 28. 1. 2023, iz <https://en.wikipedia.org/wiki/Phylogenesis>.
- Phylogenetic Inference. 2022. Stanford Encyclopedia of Philosophy. Pridobljeno 28. 1. 2023, iz <https://plato.stanford.edu/entries/phylogenetic-inference/>.
- Phylogenetic tree. 2023. Wikipedia. Pridobljeno 28. 1. 2023, iz https://en.wikipedia.org/wiki/Phylogenetic_tree.

- Phylogenetics. 2023. Wikipedia. Pridobljeno 28. 1. 2023, iz <https://en.wikipedia.org/wiki/Phylogenetics>.
- Posterior probability. 2023. Wikipedia. Pridobljeno 1. 3. 2023, iz https://en.wikipedia.org/wiki/Posterior_probability.
- R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rambaut A. 2018. Tree Figure Drawing Tool, v 1.4.4. <https://beast.community/programs>.
- Rambaut A., Drummond J.A. 2018. TreeAnnotator, v1.10.4. <https://beast.community/programs>.
- Schliep K.P. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics*, 27(4) 592–593.
- Sequence homology. 2023. Wikipedia. Pridobljeno 15. 2. 2023, iz https://en.wikipedia.org/wiki/Sequence_homology.
- Speciation. 2023. Wikipedia. Pridobljeno 28. 1. 2023, iz <https://en.wikipedia.org/wiki/Speciation>.
- Swofford L.D., Sullivan J. 2009. Phylogeny inference based on parsimony and other methods using Paup*. V: Lemey P., Salemi M., Vandamme A.M. (eds.). *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*. Cambridge University Press. Cambridge.
- Systematics. 2023. Wikipedia. Pridobljeno 28. 1. 2023, iz <https://en.wikipedia.org/wiki/Systematics>.
- Tamura K., Stecher G., Kumar S. 2021. MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution* <https://doi.org/10.1093/molbev/msab120>.

15. poglavje



MONITORING

Povzetek

15. poglavje: Monitoring

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

Monitoring je načrtno in sistematično spremljanje in merjenje izbranih parametrov, pojavov ali procesov tekom daljšega časovnega obdobja, po opredeljeni metodologiji in z definiranim namenom.

Monitoring biodiverzitete je pri nas urejen v *Zakonu o varstvo okolja* ter v *Zakonu o ohranjanju narave*. Monitoring stanja okolja obsega spremljanje in nadzorovanje kakovosti tal, voda in zraka ter biotske raznovrstnosti. Monitoring ohranjenosti narave obsega: spremljanje stanja rastlinskih in živalskih vrst, njihovih habitatov, habitatnih tipov, ekološko pomembnih območij, posebnih varstvenih območij in ekosistemov in spremljanje stanja na področju naravnih vrednot. Podatke o stanju ohranjanja narave zbiramo z namenom pregleda stanja na področju naravnih vrednot in z namenom izvajanja programa ohranjanja biotske raznovrstnosti.

Monitoring biodiverzitete nima enotne metodologije, saj zaradi heterogenosti živih bitij univerzalni in standardizirani monitoring populacij ni mogoč. Večina vrst živih bitij ima specifično metodologijo ugotavljanja njihove prisotnosti in ocenjevanja številčnosti. Posamezne stroke na področju biologije živali, rastlin, mikroorganizmov in sorodnih disciplin so razvile specifične raziskovalne metode.

Časovno obdobje izvajanja monitoringa je odvisno na narave pojava, npr. klimatske spremembe spremljajo celotno obdobje merjenja temperatur; npr. populacijsko dinamiko izbrane ptičje vrste spremljamo v desetletjih; npr. spremljanje populacijske dinamike izbranih mikroorganizmov lahko poteče v nekaj dnevih.

Dobrobit in preživetje organizmov, vključno z ljudmi, je odvisno od delovanja ekosistemov. Kakovostne ekološke in okoljske informacije, zbrane za dolgo obdobje, zagotavljajo poglobljen vpogled v spremembe ekosistemov, v potek ekoloških procesov in v zagotavljanje ekosistemskih storitev. Brez monitoringa tudi ne moremo vedeti, kako učinkovito je upravljanje na primer: z izpusti plinov, ki razgrajujejo ozonsko plast ali prispevajo k globalnemu segrevanju; z invazivnimi vrstami; s programi obnove habitatov; z aktivnostmi za pomoč ogroženim vrstam ali z vnosom hranil/polutantov v okolje.

Dolgoročni podatki so temeljnega pomena za: – dokumentiranje in zagotavljanje temeljnih informacij o stanju okolja; – vrednotenje ekološkega odziva na motnje; – zaznavo in vrednotenje sprememb v strukturi in delovanju ekosistema; – oblikovanje novih strokovnih/znanstvenih vprašanj o dinamiki populacij, združb in ekosistemov; – zagotavljanje empiričnih podatkov za testiranje ekoloških teorij in modelov; – zagotavljanje platforme za različne vidike raziskav; – podatkovno rudarjenje in iskanje vzorcev v meritvah; – vključevanje širše družbe v raziskovalno prakso in sodelovanje pri zbiranju podatkov; – prispevek k oblikovanju politik in zakonodaje v upravljanju z okoljem in naravo.

Monitoringi omogočajo spremljanje treh oblik procesov: počasne procese, redke dogodke in dolgoročne trende in so ključnega pomena za razumevanje in napovedovanje dogajanj v ekologiji in biodiverziteti.

Monitoring (ang. Monitoring, Long-term Studies) je načrtno in sistematično spremljanje oz. merjenje izbranih parametrov, pojavov in procesov daljše časovno obdobje, po opredeljeni metodologiji in z definiranim namenom.

Monitoring biodiverzitete je pri nas urejen v *Zakonu o varstvu okolja* ter v *Zakonu o ohranjanju narave*. V *Zakonu o ohranjanju narave* (2020) – ZON, v 1. odstavku 108. člena piše, da monitoring ohranjenosti narave obsega: spremljanje stanja rastlinskih in živalskih vrst, njihovih habitatov, habitatnih tipov, ekološko pomembnih območij, posebnih varstvenih območij in ekosistemov, spremljanje stanja na področju naravnih vrednot. V naslednjem, 109. členu je zapis, da se podatki o stanju ohranjanja narave zbirajo z namenom:

- pregleda stanja na področju naravnih vrednot in
- izvajanja programa ohranjanja biotske raznovrstnosti.

Uredba o posebnih varstvenih območjih (območjih Natura 2000) (2004) je podzakonski akt ZON, ki v 10. členu določa izvajanje monitoringa kazalcev, ki omogoča: 1. spremljanje stanja rastlinskih in živalskih vrst ter habitatnih tipov; 2. ugotavljanje učinkovitosti ukrepov varstva glede doseganja ugodnega stanja rastlinskih in živalskih vrst ter habitatnih tipov; 3. monitoring se prednostno izvaja za rastlinske in živalske vrste ter habitatne tipe, zaradi katerih je bilo območje Natura opredeljeno, pri čemer je posebna pozornost namenjena prednostnim vrstam in prednostnim habitatnim tipom ter vrstam, s stanjem katerih se najbolj očitno odražajo spremembe v habitatih drugih vrst oz. v habitatnih tipih. Države podpisnice Natura mreže naravnih območij poročajo Evropski komisiji, skupna poročila pa so objavljena na: <https://www.eea.europa.eu/publications/state-of-nature-in-the-eu-2020>.

V *Zakonu o varstvu okolja* (2020) je v 3. odstavku 96. člena zapisano, da monitoring stanja okolja obsega spremljanje in nadzorovanje kakovosti tal, voda in zraka ter biotske raznovrstnosti.

Monitoring stanja okolja z metodološkega, tehnološkega in izvedbenega vidika vključuje zelo različne pristope, npr. mednarodno standardizirane meritve parametrov vode in zraka. Medtem monitoring biodiverzitete nima in tudi ne more imeti enotne metodologije, saj zaradi heterogenosti živih bitij, univerzalni in standardizirani monitoring populacij ni mogoč. Še več, večina vrst živih bitij ali vsaj ekološko sorodnih skupin organizmov ima specifično metodologijo ugotavljanja njihove prisotnosti in ocenjevanja številčnosti. Posamezne stroke na področju biologije živali, rastlin, mikroorganizmov in sorodnih disciplin so razvile specifične raziskovalne metode (za dodatne vsebine glej Poglavje 16 Vzorčenje), obseg predstavljanja posameznih metod presega namen tega dela. Sistematično je pristop k vzorčenju predstavljen npr. v delih o praksi v terenski ekologiji (Wheater in sod. 2020) in v priročniku o biodiverzitetnih metodah (Hill in sod. 2012).

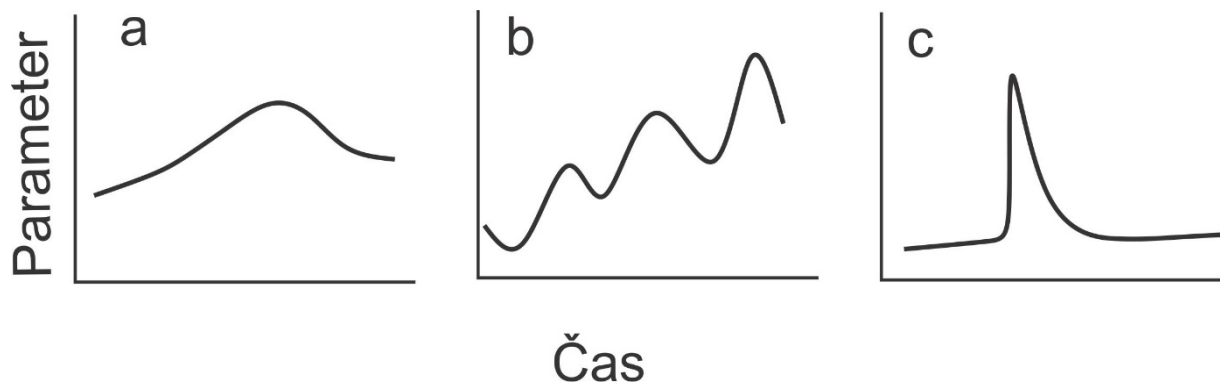
Relevantnost časovnega obdobja izvajanja monitoringa se ravna glede na naravo pojava, npr. klimatske spremembe spremljamo tekom celotnega obdobja merjenja temperatur – v stoletjih; npr. spremljanje populacijske dinamike izbranih mikroorganizmov lahko poteče v nekaj dneh; npr. populacijsko dinamiko izbrane ptičje vrste pa spremljamo v desetletjih itd.

Dobrobit in preživetje organizmov, vključno z ljudmi, je odvisno od delovanja ekosistemov. Kakovost ekološke in okoljske informacije, zbrane za dolgo obdobje, zagotavljajo poglobljen vpogled v spremembe ekosistemov, v potek ekoloških procesov in v zagotavljanje ekosistemskih storitev. Brez teh informacij nimamo vpogleda v stanje ekoloških procesov in sistemov, ki zagotavljajo življenje na planetu. Brez monitoringa tudi ne moremo vedeti, kako učinkovito je upravljanje na primer: z izpusti plinov, ki razgrajujejo ozonsko plast ali prispevajo h globalnemu segrevanju; z invazivnimi vrstami; s programi obnove habitatov; z aktivnostmi za pomoč ogroženim vrstam ali z vnosom hranil/polutantov v okolje.

Dolgoročni podatki (ang. Long-term Data) so potrebni zaradi več vzrokov (prirejeno po Lindenmayer in Likens 2018), temeljnega pomena so za:

- dokumentiranje in zagotavljanje temeljnih informacij o stanju okolja, kar omogoča spremljanje in vrednotenje sprememb (npr. klimatskih sprememb);
- vrednotenje ekološkega odziva na motnje;
- zaznavo in vrednotenje sprememb v strukturi in delovanju ekosistema (npr. spremembe v akumulirani biomasii gozda);
- oblikovanje novih strokovnih/znanstvenih vprašanj o dinamiki populacij, združb in ekosistemov;
- zagotavljanje empiričnih podatkov za testiranje ekoloških teorij in modelov;
- zagotavljanje platforme za različne vidike raziskav;
- podatkovno rudarjenje in iskanje vzorcev v meritvah;
- vključevanje širše družbe v raziskovalno prakso in sodelovanje pri zbiranju podatkov (ang. Citizen Science);
- prispevek k oblikovanju politik in zakonodaje v upravljanju z okoljem in naravo.

Monitoringi oz. dolgoročne študije omogočajo vpogled v tri oblike procesov: 1. **počasne procese**, 2. **dolgoročne trende** in 3. **redke dogodke** (Slika 15.1) in so ključnega pomena za razumevanje in napovedovanje dogajanj v ekologiji in biodiverziteti.



Slika 15.1 Primeri različnih oblik ekoloških procesov v času, a. počasni procesi, b. dolgoročni trendi c. redki dogodki (prirejeno po Strayer in sod., 1986).

Monitoring biodiverzitete in izbranih okoljskih parametrov praviloma izvajamo z namenom iskanja povezav med stanjem biodiverzitete in njenim odzivom na ekološke dejavnike. V nadaljevanju navajamo nekaj primerov z različnimi metodološkimi pristopi. Sprememba v gnezditveni biologiji muharja, kot odziv na globalne klimatske spremembe v Primeru 15.1. Zastrupitev orlov s svincem zaradi vnosa šiber v prehranjevalno verigo v Primeru 15.2. Zlom in okrevanje populacije sokola selca zaradi vnosa insekticida DDT v prehranjevalno verigo Primer 15.3. Napovedljivost invazije ptic je predstavljena v Primeru 15.4.



PRIMER 15.1

Globalno segrevanje in odziv gnezdenja črnoglavega muharja

Winkel in Hudde (1997) sta v obdobju 1970–1995 izvedla 25-letni monitoring gnezdenja črnoglavega muharja *Ficedula hypoleuca* (Slika 15.2). Monitoring je obsegal podatek o povprečnem dnevu izleganja iz jajc in o povprečnem številu mladičev v prvem gnezdu v posamezni gnezditveni sezoni oz. za posamezno leto (Preglednica 15.1).

Za isto časovno obdobje smo pridobili klimatske podatke o povprečni letni temperaturi zraka (Preglednica 15.1), (Opazovani in merjeni meteorološki podatki po Sloveniji 2020). Povprečna temperatura izkazuje stalen trend naraščanja s stopnjo 0,05 stopinje na leto, v 25-letnem obdobju se je temperatura dvignila za 1,4 °C (Sliki 15.3 in 15.4). Mladiči se izlegajo vsako let bolj zgodaj, ta trend je konstanten in ima stopnjo 0,3, kar pomeni, da se v povprečju mladiči izležejo vsako leto za 0,3 dneva prej, oz. v celotnem obdobju je ta razlika 5,6 dni (Slika 15.4). Povprečno število mladičev narašča s stopnjo 0,02, v raziskovanem obdobju se je povprečno število mladičev povečalo za 0,5 mladiča na gnezdo (Slika 15.3).

Vzrok sprememb v gnezditveni biologiji muharja lahko povežemo z višanjem temperature in posledičnim zgodnejšim razvojem vegetacije in zgodnejšo aktivnostjo žuželk, glavno hrano muharja. Najbrž je tudi žuželk več, kar ugodno vpliva na število mladičev.

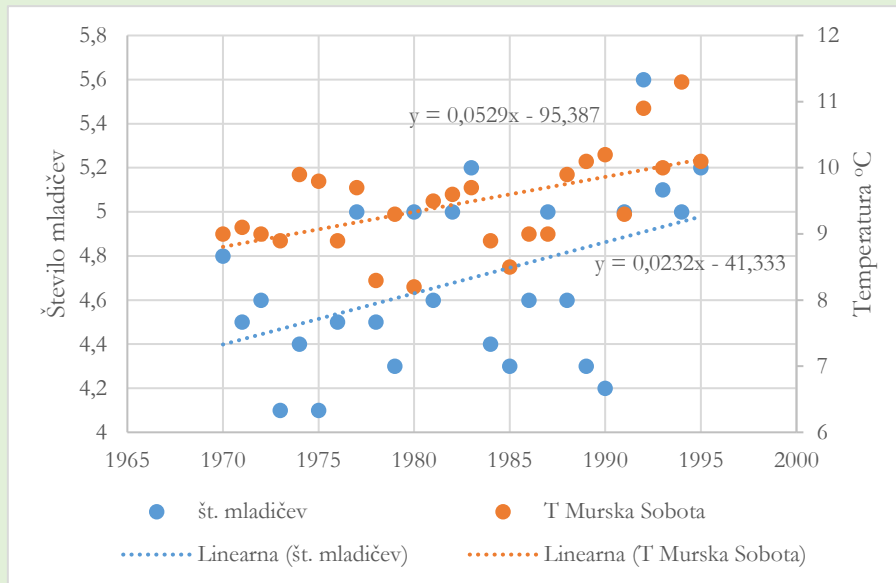


Slika 15.2 Črnoglavi muhar *Ficedula hypoleuca*, na globalno segrevanje se odziva z zgodnejšim gnezdenjem.

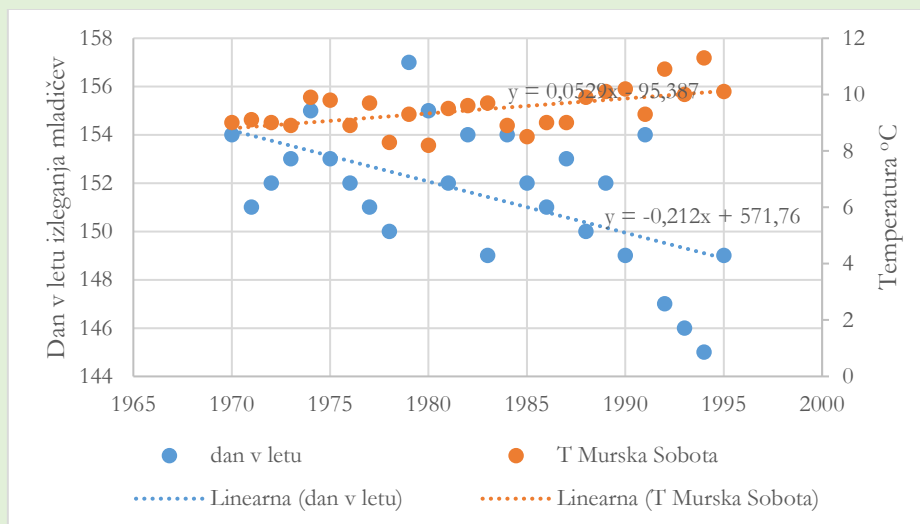
Foto. Gregor Bernard.

Preglednica 15.1 Povprečni dan izleganja iz jajc in povprečno število mladičev črnoglavega muharja ter povprečna letna temperatura, v obdobju od 1970 do 1995; dan v letu odčitamo iz priloge »Dan v letu«

| Izleganje | Dan v letu | Št. mladičev | T °C |
|------------|------------|--------------|------|
| 3.06.1970 | 154 | 4,8 | 9 |
| 31.05.1971 | 151 | 4,5 | 9,1 |
| 1.06.1972 | 152 | 4,6 | 9 |
| 2.06.1973 | 153 | 4,1 | 8,9 |
| 4.06.1974 | 155 | 4,4 | 9,9 |
| 2.06.1975 | 153 | 4,1 | 9,8 |
| 1.06.1976 | 152 | 4,5 | 8,9 |
| 31.05.1977 | 151 | 5,0 | 9,7 |
| 30.05.1978 | 150 | 4,5 | 8,3 |
| 6.06.1979 | 157 | 4,3 | 9,3 |
| 4.06.1980 | 155 | 5,0 | 8,2 |
| 1.06.1981 | 152 | 4,6 | 9,5 |
| 3.06.1982 | 154 | 5,0 | 9,6 |
| 29.05.1983 | 149 | 5,2 | 9,7 |
| 3.06.1984 | 154 | 4,4 | 8,9 |
| 1.06.1985 | 152 | 4,3 | 8,5 |
| 31.05.1986 | 151 | 4,6 | 9 |
| 2.06.1987 | 153 | 5,0 | 9 |
| 30.05.1988 | 150 | 4,6 | 9,9 |
| 1.06.1989 | 152 | 4,3 | 10,1 |
| 29.05.1990 | 149 | 4,2 | 10,2 |
| 3.06.1991 | 154 | 5,0 | 9,3 |
| 27.05.1992 | 147 | 5,6 | 10,9 |
| 26.05.1993 | 146 | 5,1 | 10 |
| 25.05.1994 | 145 | 5,0 | 11,3 |
| 29.05.1995 | 149 | 5,2 | 10,1 |



Slika 15.3 Trend naraščanja povprečne letne temperature zraka od leta 1970 do 1995 in trend naraščanja povprečnega števila mladičev v gnezdu.



Slika 15.4 Trend naraščanja povprečne letne temperature zraka od leta 1970 do 1995 in trend vse zgodnejšega izleganja mladičev.



PRIMER 15.2

Bioakumulacija svineca v jetrih planinskega orla

Planinski orl *Aquila chrysaetos* je plenilec in mrhovinar in člen v delovanju ekosistema, s prehranjevanjem s poškodovanimi živalmi ali mrhovino prispeva h kroženju biomase in energije. Zaužitje svinčenih šiber pri prehranjevanju z ustreljeno mrhovino ali obstreljenimi živalmi je vzrok zastrupitve s svincem. Zastrupijo se tudi ptice ujede.

Zastrupitev ptic s svincem se kaže v izgubi ravnotežja, pojavu encefalitisa, slabšanju vida, težavami z dihanjem, zmanjšanje mišične moči, dehidracije, poškodb ledvic in jeter. Stopnjevanje simptomov vodi v slabšo kondicijo živali in stradanje. V končni stopnji zastrupitev s svincem pripelje do epileptičnih napadov in smrti.

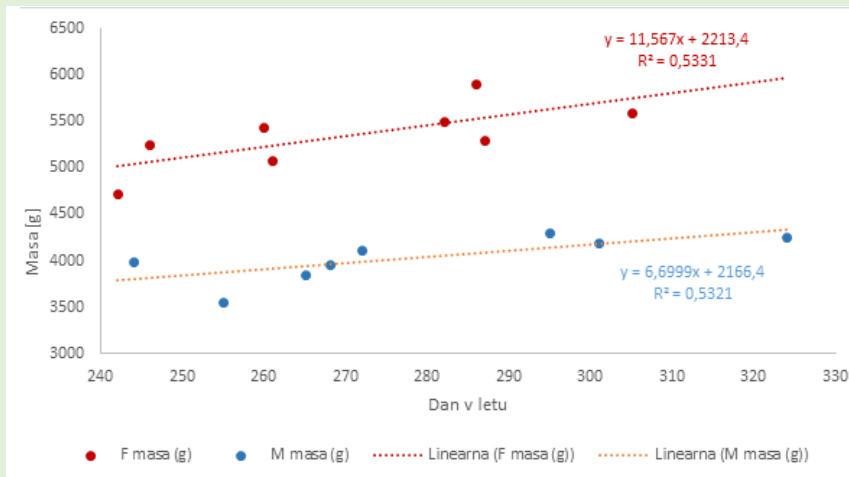
Proces je dokumentiran s primerjavami koncentracij svinca v tkivih planinskega orla pred lovno sezono in z obdobjem v lovni sezoni. V obdobju lovne sezone se vrednost svinca v tkivih poviša. Na Norveškem se lovna doba prične s 1. septembrom, Ecke in sod. (2017) so zbrali podatke o koncentracijah svinca v jetrih planinskega orla (Preglednica 15.2).

Spremljanje procesa vzdolž časovne vrste, o telesni masi orlov in koncentraciji svinca v jetrnih celicah so strnili v tri ključne ugotovitve. Samice so težje od samcev, obema spoloma se v jesenskem obdobju zvišuje telesna masa (Slika 15.5). Pri obeh spolih narašča koncentracija svinca v jetrnih celicah (Slika 15.6). Z večanjem telesne mase se večja koncentracija svinca, pri samcih s stopnjo 0,44 in pri samicah s stopnjo 0,38 (Slika 15.7).

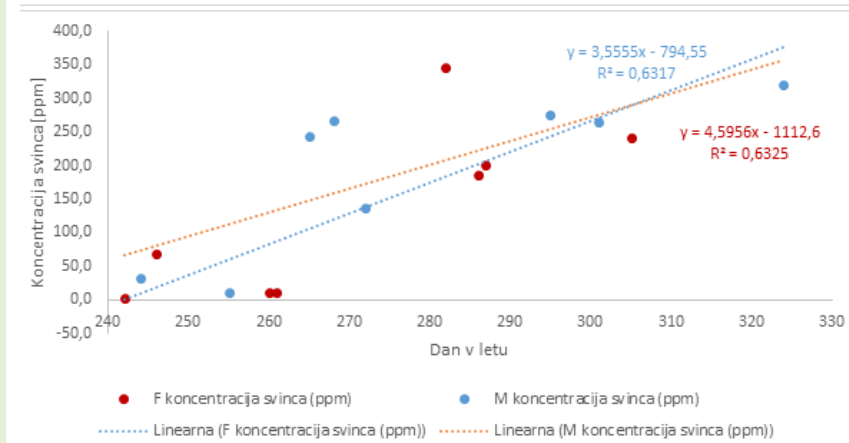
Raziskava, ki je del monitoringa poginulih orlov, prikazuje primer bioakumulacije svinca. Bioakumulacija je pojav, ko pride do kopičenja onesnažil v organizmih, ki te snovi zaužijejo neposredno s hrano ali posredno po prehranjevalni verigi, organizem pa jih ni sposoben izločiti. Najpogosteje prihaja do kopičenja kovin in klorovih organskih spojin.

Preglednica 15.2 Podatki o koncentraciji svinca v jetrih planinskega orla, masi in spolu živali ter datum vzorčenja

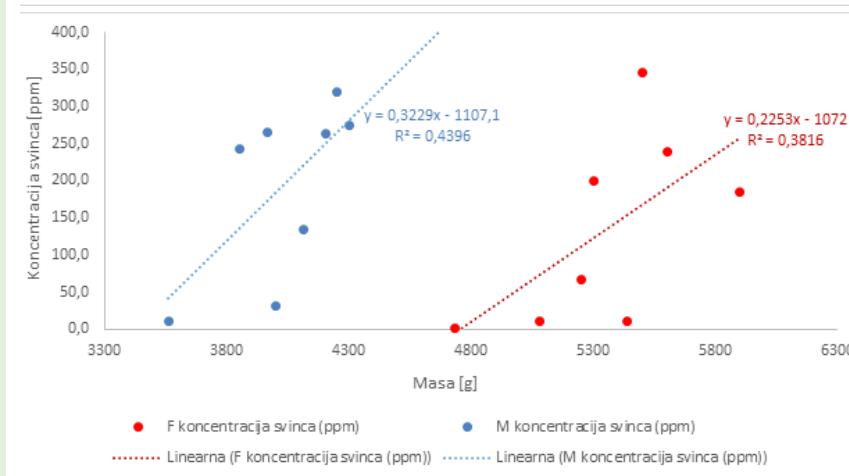
| Masa (g) | Spol | Datum (YYYYMMDD) | Koncentracija svinca (ppm) |
|----------|------|------------------|----------------------------|
| 4730 | F | 20140830 | 2,2 |
| 5250 | F | 20150903 | 68,7 |
| 5435 | F | 20140917 | 11,8 |
| 5080 | F | 20140918 | 11,5 |
| 5500 | F | 20151009 | 346,0 |
| 5900 | F | 20151013 | 186,0 |
| 5300 | F | 20151014 | 200,0 |
| 5600 | F | 20151101 | 241,0 |
| 4000 | M | 20150901 | 32,0 |
| 3560 | M | 20140912 | 11,1 |
| 3850 | M | 20150922 | 244,0 |
| | M | 20150925 | 267,0 |
| 4110 | M | 20150929 | 136,0 |
| 4300 | M | 20151022 | 276,0 |
| 4200 | M | 20151028 | 265,0 |
| 4250 | M | 20151120 | 320,0 |



Slika 15.5 Rezultat regresijske analize odnosa med maso planinskih orlov in dnevom v letu. V jesenskem obdobju od 1. septembra se masa povečuje. Samice (rdeča barva) so težje od samcev (modra barva).



Slika 15.6 Rezultat regresijske analize odnosa med količino svınca v jetrih planinskih orlov in dnevom v letu. V jesenskem obdobju od 1. septembra se koncentracija povečuje. Samice (rdeča barva) so težje od samcev (modra barva).



Slika 15.7 Koncentracija svınca v jetrih planinskih orlov glede na maso živali. Samice (rdeča barva) so težje od samcev (modra barva).



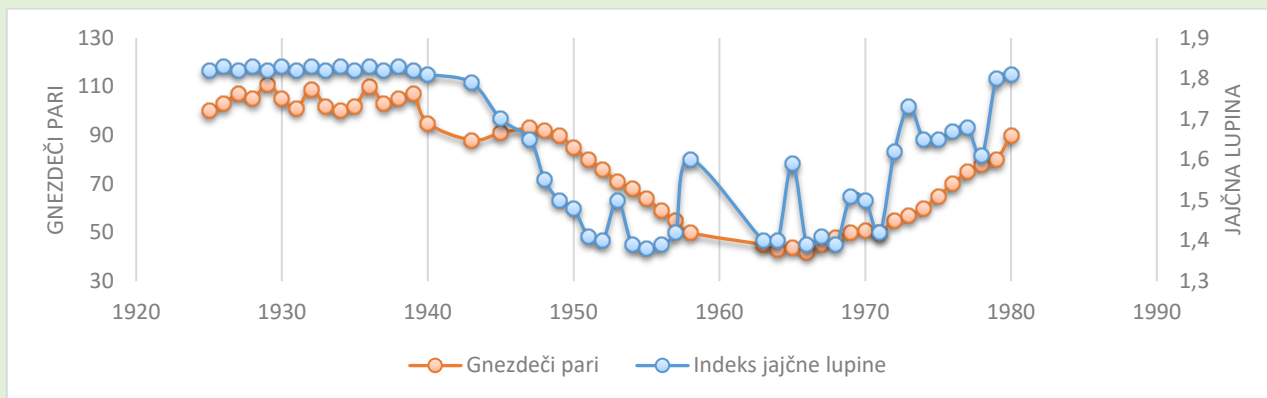
PRIMER 15.3

Uporaba insekticida DDT in biomagnifikacija v sokolu selcu

Dikloro-Difenil-Trikloroetan DDT je vsestranski insekticid, ki so ga v letih od 1940 do 1970 množično uporabljali na globalni ravni. Po letu 1970 so nekatere države DDT začele postopno umikati. Z množično uporabo je DDT prišel v prehranjevalno verigo. Zaradi

biomagnifikacije vzdolž prehranjevalne verige se je njegov vpliv najbolj odrazil v zlomu populacije sokola selca *Falco peregrinus* (Slika 15.9). Zmanjševanje števila gnezdečih parov je bila posledica tanjšanja jajčne lupine do te mere, da le-to samice med valjenjem zdrobijo pod lastno težo. DDT negativno vpliva na metabolizem kalcija. Podatke o debelini jajčne lupine za obdobje od leta 1926 do 1978, izraženi kot indeks jajčne lupine, je za Anglijo zbral Spellerberg (1991), k temu je dodal število gnezdečih parov, ki jih je dopolnil s podatki iz Ratcliffe (1980) in Cramp (1980). Dogajanje v časovni vrsti je predstavljeno na Sliki 15.8, do WW2 je variabilnost jajčne lupine in števila gnezdečih sokolov relativno stabilna. Po drugi svetovni vojni pride do strmega zmanjšanja vrednosti indeksa, temu sledi upadanje števila gnezdečih parov. Po letu 1970 se obe vrednosti počasi vračata v izhodišče. Zlom populacije sokola selca sovpada z uporabo DDT in njegovim vstopom v prehranjevalno verigo.

Biomagnifikacija je skokovito večanje koncentracije polutantov v tkivih organizmov vzdolž prehranjevalne verige.



Slika 15.8 Variabilnost indeksa jajčne lupine in števila gnezdečih sokolov selcev v obdobju od 1926 do 1978.



Foto 15.9. Sokol selec *Falco peregrinus*.

Foto. Gregor Bernard.





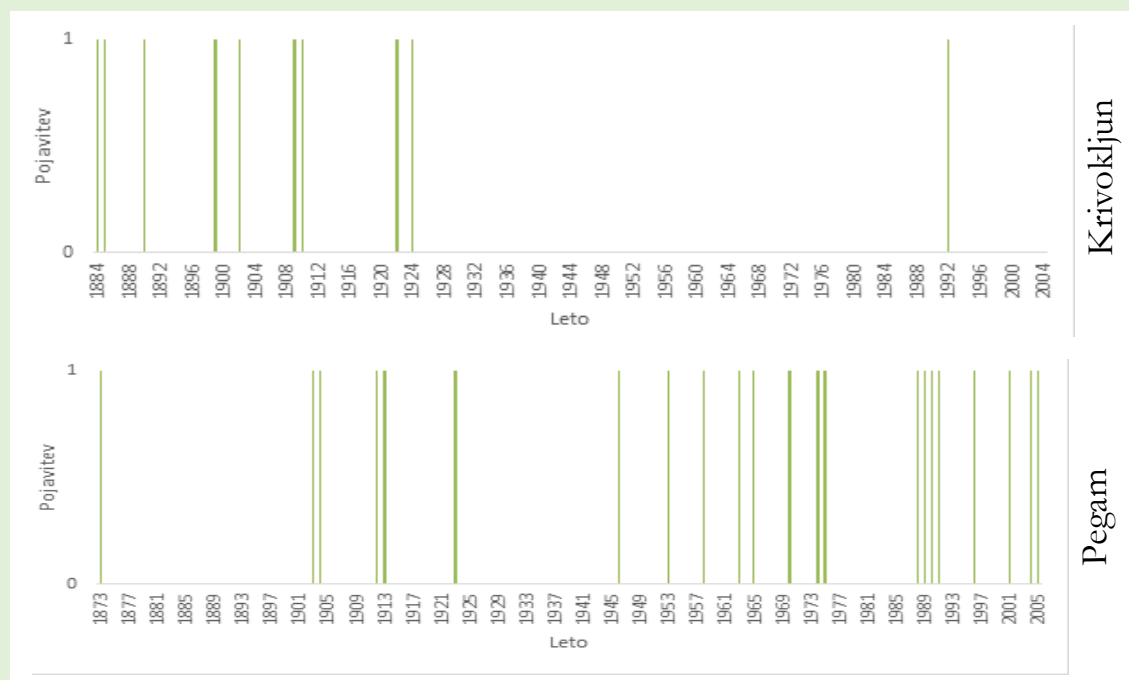
PRIMER 15.4

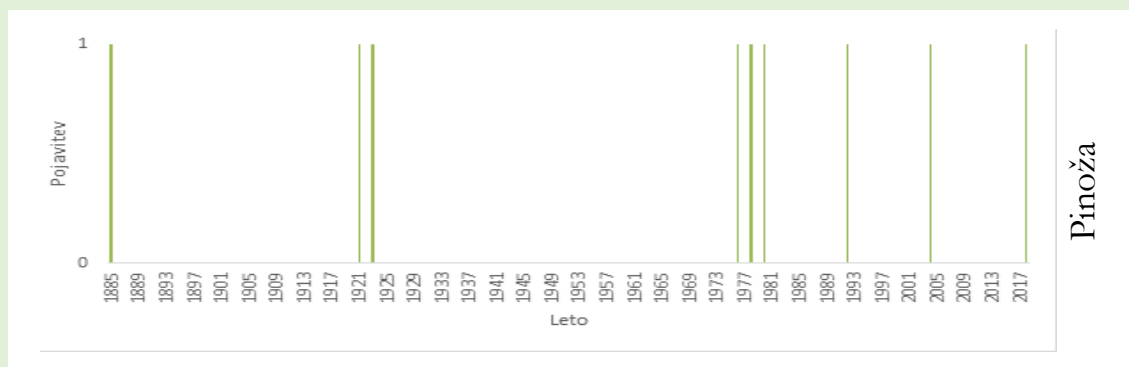
Invazije ptic

Dolgoročni monitoring omogoča evidentiranje izjemnih dogodkov. Praviloma gre za dogodke, ki jih ni moč napovedati; primer so zimske invazije ptic. Tovrstne invazije zbudijo precej pozornosti ljudi in bi jih zabeležili tudi brez načrtovanega dolgotrajnega monitoringa. Ptice so izjemno mobilni organizmi, pri tem je najbolj znan fenomen ptičjih selitev. Posebna oblika selitve so invazije, pri tem gre za množični in usklajeni premik ptic, ki ni periodičen in se ga ne da napovedati vnaprej niti krajevno niti časovno niti po številčnosti osebkov. V Sloveniji se srečujemo z invazijami treh vrst ptic: pegam *Bombycilla garrulus*, pinoža *Fringilla montifringilla* (Slika 15.12) in krivokljun *Loxia curvirostra* (Slika 15.11) podatki za zadnjih 150 let so zbrani v Preglednici 15.3 in prikazani na grafikonih (Sliki 15.10).

Preglednica 15.3 Invazijske vrste in invazijska leta, povzeto po Vrezec in sod. (2006)

| Vrsta | Invazijska leta v Sloveniji |
|--|--|
| Pegam, <i>Bombycilla garrulus</i> | 1873, 1903, 1904, 1912, 1913, 1923, 1946, 1953, 1958, 1963, 1965, 1970, 1974, 1975, 1988, 1989, 1990, 1991, 1996, 2001, 2004, 2005 |
| Pinoža, <i>Fringilla montifringilla</i> | 1885, 1921, 1923, 1976, 1978, 1980, 1992, 2004, 2018 |
| Krivokljun, <i>Loxia curvirostra</i> | 1884, 1885, 1890, 1899, 1902, 1909, 1910, 1922, 1924, 1992 |





Slika 15.10. Časovna porazdelitev invazij treh vrst ptic v Sloveniji, v obdobju 132 let.



Slika 15.11 Krivokljun *Loxia curvirostra*.
Foto. Gregor Bernard.



Slika 15.12 Množični приход pinož *Fringilla montifringilla*.
Foto. Gregor Bernard.



Zaključki

- Monitoring je načrtno in sistematično spremljanje in merjenje izbranih parametrov, pojavov ali procesov v daljšem časovnem obdobju, z opredeljeno metodologijo in z definiranim namenom.
- Monitoring biodiverzitete je pri nas urejen v *Zakonu o varstvo okolja* ter v *Zakonu o ohranjanju narave*, monitoring stanja okolja obsega spremljanje in nadzorovanje kakovosti tal, voda in zraka ter biotske raznovrstnosti.
- Monitoring ohranjenosti narave obsega: spremljanje stanja rastlinskih in živalskih vrst, njihovih habitatov, habitatnih tipov, ekološko pomembnih območij, posebnih varstvenih območij in ekosistemov, spremljanje stanja na področju naravnih vrednot in da podatke o stanju ohranjanja narave zbirajo z namenom pregleda stanja na področju naravnih vrednot in z namenom izvajanja programa ohranjanja biotske raznovrstnosti.
- Monitoring biodiverzitete nima enotne metodologije, zaradi heterogenosti živih bitij standardizirani monitoring populacij ni mogoč.
- Večina vrst živih bitij ima specifično metodologijo ugotavljanja njihove prisotnosti in ocenjevanja številčnosti. Posamezne stroke na področju biologije živali, rastlin, mikroorganizmov in sorodnih disciplin so razvile specifične raziskovalne metode.
- Časovno obdobje izvajanja monitoringa je odvisno na narave pojava.
- Kakovost ekološke in okoljske informacije, zbrane skozi dolgo obdobje, zagotavljajo poglobljen vpogled v spremembe ekosistemov, v potek ekoloških procesov in v zagotavljanje ekosistemskih storitev. Brez monitoringa ne moremo vedeti, kako učinkovito je upravljanje na primer: z izpusti plinov, ki razgrajujejo ozonsko plast ali prispevajo k globalnemu segrevanju; z invazivnimi vrstami; s programi obnove habitatov; z aktivnostmi za pomoč ogroženim vrstam ali z vnosom hranil/polutantov v okolje.
- Dolgoročni podatki so temeljnega pomena za: – dokumentiranje in zagotavljanje temeljnih informacij o stanju okolja: – vrednotenje ekološkega odziva na motnje; – zaznavo in vrednotenje sprememb v strukturi in delovanju ekosistema; – oblikovanje novih strokovnih/znanstvenih vprašanj o dinamiki populacij, združb in ekosistemov; – zagotavljanje empiričnih podatkov za testiranje ekoloških teorij in modelov; – zagotavljanje platforme za različne vidike raziskav; – podatkovno rudarjenje in iskanje vzorcev v meritvah; – vključevanje širše družbe v raziskovalno prakso in sodelovanje pri zbiranju podatkov; – prispevek k oblikovanju politik in zakonodaje v upravljanju z okoljem in naravo.
- Monitoringi omogočajo spremljanje treh oblik procesov: počasne procese, redke dogodke in dolgoročne trende in so ključnega pomena za razumevanje in napovedovanje dogajanj v ekologiji in biodiverziteti.

Priloga. Dan v letu je preglednica za pretvorbo datuma v zaporedni dan v letu. Pretvornik uporabljamo za linearizacijo časovnega zaporedja, npr. v primerih analize biodiverzitete v časovni vrsti.

| Jan | Feb | Mar | Apr | Maj | Jun | Jul | AVg | Sep | OKt | Nov | Dec |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1 | 32 | 60 | 91 | 121 | 152 | 182 | 213 | 244 | 274 | 305 | 335 |
| 2 | 33 | 61 | 92 | 122 | 153 | 183 | 214 | 245 | 275 | 306 | 336 |
| 3 | 34 | 62 | 93 | 123 | 154 | 184 | 215 | 246 | 276 | 307 | 337 |
| 4 | 35 | 63 | 94 | 124 | 155 | 185 | 216 | 247 | 277 | 308 | 338 |
| 5 | 36 | 64 | 95 | 125 | 156 | 186 | 217 | 248 | 278 | 309 | 339 |
| 6 | 37 | 65 | 96 | 126 | 157 | 187 | 218 | 249 | 279 | 310 | 340 |
| 7 | 38 | 66 | 97 | 127 | 158 | 188 | 219 | 250 | 280 | 311 | 341 |
| 8 | 39 | 67 | 98 | 128 | 159 | 189 | 220 | 251 | 281 | 312 | 342 |
| 9 | 40 | 68 | 99 | 129 | 160 | 190 | 221 | 252 | 282 | 313 | 343 |
| 10 | 41 | 69 | 100 | 130 | 161 | 191 | 222 | 253 | 283 | 314 | 344 |
| 11 | 42 | 70 | 101 | 131 | 162 | 192 | 223 | 254 | 284 | 315 | 345 |
| 12 | 43 | 71 | 102 | 132 | 163 | 193 | 224 | 255 | 285 | 316 | 346 |
| 13 | 44 | 72 | 103 | 133 | 164 | 194 | 225 | 256 | 286 | 317 | 347 |
| 14 | 45 | 73 | 104 | 134 | 165 | 195 | 226 | 257 | 287 | 318 | 348 |
| 15 | 46 | 74 | 105 | 135 | 166 | 196 | 227 | 258 | 288 | 319 | 349 |
| 16 | 47 | 75 | 106 | 136 | 167 | 197 | 228 | 259 | 289 | 320 | 350 |
| 17 | 48 | 76 | 107 | 137 | 168 | 198 | 229 | 260 | 290 | 321 | 351 |
| 18 | 49 | 77 | 108 | 138 | 169 | 199 | 230 | 261 | 291 | 322 | 352 |
| 19 | 50 | 78 | 109 | 139 | 170 | 200 | 231 | 262 | 292 | 323 | 353 |
| 20 | 51 | 79 | 110 | 140 | 171 | 201 | 232 | 263 | 293 | 324 | 354 |
| 21 | 52 | 80 | 111 | 141 | 172 | 202 | 233 | 264 | 294 | 325 | 355 |
| 22 | 53 | 81 | 112 | 142 | 173 | 203 | 234 | 265 | 295 | 326 | 356 |
| 23 | 54 | 82 | 113 | 143 | 174 | 204 | 235 | 266 | 296 | 327 | 357 |
| 24 | 55 | 83 | 114 | 144 | 175 | 205 | 236 | 267 | 297 | 328 | 358 |
| 25 | 56 | 84 | 115 | 145 | 176 | 206 | 237 | 268 | 298 | 329 | 359 |
| 26 | 57 | 85 | 116 | 146 | 177 | 207 | 238 | 269 | 299 | 330 | 360 |
| 27 | 58 | 86 | 117 | 147 | 178 | 208 | 239 | 270 | 300 | 331 | 361 |
| 28 | 59 | 87 | 118 | 148 | 179 | 209 | 240 | 271 | 301 | 332 | 362 |
| 29 | | 88 | 119 | 149 | 180 | 210 | 241 | 272 | 302 | 333 | 363 |
| 30 | | 89 | 120 | 150 | 181 | 211 | 242 | 273 | 303 | 334 | 364 |
| 31 | | 90 | | 151 | | 212 | 243 | | 304 | | 365 |
| Jan | Feb | Mar | Apr | Maj | Jun | Jul | Avg | Sep | Okt | Nov | Dec |

Literatura

- Cramp S. (ur.). 1980. Handbook of the Birds of Europe the Middle East and North Africa. The Birds of the Western Palearctic. Volume II Hawks to Bustards. Oxford University Press. Oxford.
- Ecke F., Singh N.J., Arnemo J.M., Bignert A., Helander B., Berglund Å.M.M., Borg H., Brojer C., Holm K., Lanzone M., Miller T., Nordstrom Å., Raikonen J., Rodushkin I., Ågren E., Hornfeldt B. 2017. Sublethal Lead Exposure Alters Movement Behavior in Free-Ranging Golden Eagles. *Environmental Science & Technology*, 51: 5729–5736.
- Hill D., Fashman M., Tucker G., Shewry M., Shaw P. 2012. Handbook of Biodiversity Methods. Survey, Evaluation and Monitoring. Cambridge University Press. Cambridge.
- Lindenmayer B.D., Likens E.G. 2018. Effective Ecological Monitoring. CSIRO Publishing. Clayton.
- Opazovani in merjeni meteorološki podatki po Sloveniji. 2020. Državna meteorološka služba. ARSO. MKO. Pridobljeno dne 1. 12. 2020 iz: <http://www.arso.gov.si/>.
- Ratcliffe D. 1980. The Peregrine Falcon. T & AD Poyser. London.
- Spellerberg I.F. 1991. Monitoring ecological change. Cambridge University Press. Cambridge.
- Strayer D., Glitzenstein J.S., Jones C.G., Kolasa J., Likens G.E., McDonnell M.J., Parker G.G., Pickett S.T.A. 1986. Long-Term Ecological Studies: an illustrated account of their design, operation, an importance to ecology. Intitute of Ecosystem Studies. New York.
- Uredba o posebnih varstvenih območjih (območjih Natura 2000). 2004. Uradni list RS, št. 49/04, 110/04, 59/07, 43/08, 8/12, 33/13, 35/13 – popr., 39/13 – odl. US, 3/14, 21/16 in 47/18.
- Vrezec A., Tome D., Denac D. 2006. Selitev in izjemni selitveni pojavi pri pticah. *Ujma*, 20: 125–136.
- Wheater C.P., Cook A.P., Bell R.J. 2020. Practical Field Ecology. Second Edition. Wiley Blackwell. Hoboken, NJ.
- Winkel W., Hudde H. 1997. Long-term trends in reproductive traits of tits (*Parus major*, *P. caeruleus*) and pied flycatchers *Ficedula hypoleuca*. *Journal of Avian Biology*, 28: 187–190.
- Zakon o ohranjanju narave – ZON-E. 2020. Uradni list RS, št. 82/20.
- Zakon o varstvu okolja – ZVO-1J. 2020. Uradni list RS, št. 158/20.

16. poglavje



VZORČENJE

Povzetek

16. poglavje: Vzorčenje

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

Raziskave vzorcev in procesov biodiverzitete temeljijo na vzorčenju biote in meritvah abiotskih dejavnikov okolja. Vzorčenje živih organizmov je specifično za vsako skupino posebej, zato ni standardiziranih protokolov; oblikovane so različne prakse vzorčenja. Ključnega pomena je strokovna usposobljenost raziskovalca z ustrežno izobrazbo in s praktičnimi izkušnjami. Strokovna usposobljenost vključuje poznavanje osnovne biologije organizmov in njihovo prepoznavanje, poznavanje značilnosti habitatov in mikrolokacijskih zahtev organizmov, načine njihovega vzorčenja, izbiro ustreznih metod glede na konkretne okoliščine v naravi, praktični odvzem v naravi, morebitno pripravo vzorcev na terenu, etiketiranje vzorcev in vodenje zapisnika o vzorčenju, transport in hranjenje v laboratoriju, determinacijo vrst, trajno skladiščenje vzorcev ter oblikovanje zbirke podatkov.

Za izvedbo vzorčenja biote moramo pridobiti dovoljenje Agencije RS za okolje in soglasje Zavoda RS za varstvo narave, upoštevati moramo relevantno naravovarstveno zakonodajo. Pri vzorčenju postopamo etično do organizmov, da jim ne povzročamo nepotrebne bolečina in da jih iz narave odvzame samo toliko, kolikor je nujno za izvedbo raziskave. Izmed abiotskih dejavnikov merimo tiste fizikalne, geografske in klimatske, ki so relevantni za razlago biodiverzitete.

Relevantnost zaključkov zagotovimo z: nepristranostjo – brez vnaprejšnjih odločitev o rezultatih; ponovljivostjo – z uporabo istih metod dobimo identične rezultate; preverljivostjo – ponovljene raziskave drugih raziskovalcev vrnejo identične rezultate; veljavnostjo – načrt raziskave in metode vzorčenja so ustrezne za odgovor na zastavljeno vprašanje.

Vzorec v biodiverzitetni raziskavi je enkratni popis izbranih organizmov, lahko tudi drugih biotskih in abiotskih dejavnikov vzorčenega mesta, v definiranim časovnem oknu. Pristopi k vzorčenju so: enostavno naključno vzorčenje, vsaka enota populacije ima enakovredno možnost in verjetnost, da bo zajeta v vzorec; stratificirano naključno vzorčenje, populacijo ali območje vzorčenja, razdelimo v več neskladnih, a homogenih podskupin; sistematično vzorčenje; vzorčenje s presojo.

V biodiverzitetni raziskavi zbiramo štiri osnovne kategorije podatkov: 1. biološki podatki, taksonomija vrst z latinskimi imeni, nanjo se navezujejo druge biološke lastnosti vrste; 2. podatki o številčnosti posamezne vrste; 3. časovni podatki; 4. podatki o vzorčenem mestu, geolokacija ter fizikalne, geološke, klimatske, kemijske in druge lastnosti. Podatke vnašamo npr. v Excelovo preglednico, podatke več preglednic lahko med seboj povežemo v relacijsko matriko in zgradimo podatkovno bazo v Accessu ali v drugih orodjih.

Rezultate raziskav in raziskovalne podatke delimo v odprti znanosti po načelu FAIR.

Lokalitete vzorčenj predstavimo na zemljevidu.

Zbirke organizmov trajno hranimo v naravoslovnih muzejih.

Raziskave vzorcev in procesov biodiverzitete temeljijo na **vzorčenju živih bitij** in **meritvah abiotskih dejavnikov** habitata, ki ga proučevana živa bitja naseljujejo. Pojem vzorec v tem besedilu nastopa v dveh pomenih, najprej kot ponavljajoča se podenota urejenosti raziskovane enote (entitete) (ang. Pattern) in drugič kot majhen del raziskovane enote, primerne za neposredne raziskave (meritve, ocene) (ang. Sample). Vzorčenje obravnavamo kot pridobivanje informacij o živih bitjih in njihovem okolju. V mnogih primerih moramo iz narave odvzeti del organizmov ali ugotoviti njihovo prisotnost in številčnost, tj. vzorce organizmov. V tem pomenu je vzorec reprezentativen delež raziskovane populacije ali združbe, to je tak, da dobro odraža značilnosti populacije ali združbe.

Zaradi velike heterogenosti organizmov glede velikosti, oblike, zgradbe, presnove, razmnoževanja ipd. in njihovih prilagoditev na abiotske in biotske razmere v okolju je vzorčenje za vsako skupino organizmov specifično. Standardiziranih protokolov vzorčenja je malo; vpeljeni so predvsem za izbrane skupine mikroorganizmov, medtem ko uporabljajo raziskovalci za raziskave biodiverzitete drugih organizmov različne, ne poenotene načine vzorčenja. Akreditirani preizkuševalni in kalibracijski laboratoriji po sistemu standarda ISO/IEC 17025:2017, za vzorčenje biodiverzitete, še ne delujejo.

Za raziskovanje posameznih skupin organizmov so bolj ali manj uveljavljene specifične metode vzorčenje v naravi, ocenjevanja velikosti populacij ali monitoringa, npr. za vzorčenje členonožcev (Santos in Fernandes 2020), ptičev (Bibby in sod. 2000), sesalcev (Sikes 2016) itd. Pregled pristopov in metod za sistematično vzorčenje organizmov v naravi je predstavljen npr. v priročnikih: o praksi v terenski ekologiji (Wheater in sod. 2020, Ecological sampling methods 2020, McCabe 2011), o biodiverzitetnih metodah (Hill in sod. 2012) in analizi ekoloških podatkov (Krebs 2014, Legendre in Legendre 2012).

Za uspešno vzorčenje je ključnega pomena ustrezna strokovna usposobljenost raziskovalca z ustrezno izobrazbo in s praktičnimi izkušnjami z organizmi. Strokovna usposobljenost se nanaša na poznavanje organizmov, načine njihovega vzorčenja oz. popisovanja. V primeru odvzema organizmov iz narave vzorcev na terenu fiksiramo, vzorce etiketiramo (označimo) in vodimo zapisnik o vzorčenju. Sledi transport vzorcev v laboratorij in njihova obdelava, skladno s cilji raziskave, determinacija vrst, trajno hranjenje odvzetih vzorcev in oblikovanja zbirke. Delo se nadaljuje s statistično obdelavo in vrednotenjem podatkov in se konča z izdelavo poročila o opravljeni raziskavi.

Za odvzem organizmov iz narave moramo pridobiti dovoljenje naravovarstvenega organa, to je Agencija Republike Slovenije za okolje, ki deluje znotraj Ministrstva za okolje in prostor, soglasje pa mora podati tudi Zavod Republike Slovenije za varstvo narave. Raziskovalci moramo pri svojem delu spoštovati naravovarstveno zakonodajo, in sicer *Uredbo o zavarovanih prostoživečih živalskih vrstah* (2019), *Zakon o obranjanju narave* (2020) in

Uredbo o posebnih varstvenih območjih (območjih Natura 2000) (2018) ter druge podzakonske predpise. Pri vzorčenju moramo postopati etično do organizmov, ki jih vzorčimo, da jim ne povzroča nepotrebne bolečina in da jih preštejemo v naravnem okolju, iz narave jih odvzame samo toliko, kolikor je nujno za izvedbo raziskave.

Zbrane podatke organiziramo v skladu z namenom raziskave, jih analiziramo, z njimi opisujemo stanje in razlagamo dogajanja v naravi ter vključimo v preverjanje hipotez.

Vzorčenje abiotskih dejavnikov

Abiotske dejavnike, ki jih merimo ob vzorčenju biote, poenostavljeno razdelimo v fizikalne, geografske in klimatske (Preglednice 16.1, 16.2, 16.3). Enota za temperaturo je Kelvin, vendar je v vsakdanji ekološki praksi ustaljena uporaba izpeljane enote za temperaturo v stopinjah Celzija. Abiotski dejavniki v ekoloških raziskavah so vse fizikalno-kemične lastnosti okolja ali medija, v katerem organizmi živijo, kot so npr. lastnosti tal, lastnosti zraka ali lastnosti vode. Navajamo samo nekaj osnovnih okoljskih dejavnikov.

Izvajanje meritev abiotskih dejavnikov okolja je nabor standardiziranih opravil, s katerimi določamo velikost njihovih količin, za meritev potrebujemo ustrezne merilne naprave. Meritev praviloma primerjamo s standardom, zato morajo biti merilne naprave umerjene. Rezultat meritve so fizikalne količine, ki jih sestavljata merska vrednost in enota. Merske enote so določene na podlagi mednarodnih dogovorov in standardov. Mednarodni standardi so podani v okviru *Mednarodnega sistema enot SI* (fr. *Systeme International*) ter implementirani v slovenski zakonodaji v *Zakonu o meroslovju* (2005) in v *Uredbi o merskih enotah* (2001).

Lego vzorčenih lokalitet v prostoru navedemo z GPS koordinatami: geografsko širino, geografsko dolžino in nadmorsko višino (GPS – Globalni sistem pozicioniranja, ang. *Global Positioning System*) (Preglednica 16.2). Ob geografskih koordinatah navedemo tudi ime bližnjega kraja ali drugo relevantno geografsko ime npr. naselja, regije, hriba, reke ipd.

Klimatski dejavniki so standardizirani z 19 globalnimi bioklimatskimi indikatorji (Preglednica 16.3) (O'Donnell in Ignizio 2012), podatki so prosto dostopni na spletu (glej primere v Poglavju 10 in 11).

Preglednica 16.1 Osnovnih sedem fizikalnih količin; * v vsakdanji praksi se v ekoloških raziskavah uporablja izpeljana enota za temperaturo v stopinjah Celzija

| Količina | Enota | Oznaka |
|----------------|----------------------------|--------|
| Čas | Sekunda | s |
| Dolžina | Meter | m |
| Masa | Kilogram | kg |
| Temperatura | Kelvin (stopinja Celzija)* | K (°C) |
| Električni tok | Amper | A |
| Množina snovi | Mol | mol |
| Svetilnost | Kandela | cd |

Preglednica 16.2 Tri količine geografskega koordinatnega sistema

| Količina | Enota | Oznaka |
|--------------------|------------------|--------|
| Geografska širina | Stopinja | ° |
| Geografska dolžina | Stopinja | ° |
| Nadmorska višina | Metri nad morjem | m.n.m. |

Preglednica 16.3 19 bioklimatskih količin

| Koda | Opis |
|-------|---|
| BIO1 | Povprečna letna temperatura (*100) |
| BIO2 | Povprečni dnevni razpon temperatur (mesečno povprečje (Maks. tem. – Min. tem.)) |
| BIO3 | Izotermalnost (BIO2/BIO7) (*100) |
| BIO4 | Sezonskost temperature (standardna deviacija * 100) |
| BIO5 | Maksimalna temperatura vročega meseca |
| BIO6 | Minimalna temperatura hladnega meseca |
| BIO7 | Letni razpon temperature (BIO5-BIO6) |
| BIO8 | Povprečna temperatura mokrega četrletja |
| BIO9 | Povprečna temperatura suhega četrletja |
| BIO10 | Povprečna temperatura vročega četrletja |
| BIO11 | Povprečna temperatura hladnega četrletja |
| BIO12 | Letne padavine |
| BIO13 | Padavine mokrega meseca |
| BIO14 | Padavine suhega meseca |
| BIO15 | Sezonskost padavin (Koefficient variabilnosti) |
| BIO16 | Padavine mokrega četrletja |
| BIO17 | Padavine suhega četrletja |
| BIO18 | Padavine toplega četrletja |
| BIO19 | Padavine hladnega četrletja |

Kakovost

Tako kot vsi procesni postopki, ki jih ljudje vodimo in izvajamo, tudi raziskave biote načrtujemo v skladu s cikličnim zaporedjem **Demingovega kroga**: načrt, izvedba, nadzor/kontrola, ukrepi. Raziskovalec oz. vodja projekta mora delo na področju biodiverzitete znati načrtovati v skladu z namenom in cilji projekta ter časovnimi, kadrovskimi, materialnimi in finančnimi zmožnostmi ter etičnimi normami; ga pravilno izvesti v skladu s protokolom vzorčenja; znati meriti in interpretirati učinke ter se ustrezno odzvati oz. ukrepati glede na ugotovitve.

Načrtovanje biodiverzitetne raziskave se prične z **raziskovalnim vprašanjem** in definiranjem **cilja**, ki ga želimo doseči z raziskavo. Oblikovanju stališča o **namenu raziskave** – to je stališče o tem, kaj želimo izvedeti; sledi postavitve **hipoteze** – to je opredelitev predpostavke, ki jo v raziskavi preverjamo. Na osnovi dobro definirane namena raziskave in dobro zastavljene raziskovalne hipoteze predvidimo, katere podatke potrebujemo. Sledi priprava načrta za zbiranje podatkov, določitev njihove natančnosti in potrebnega števila podatkov. Zelo napačna je zamisel, da bo »statistična analiza« vrnila vpogled v strukturo podatkov. Razumevanje ekoloških odnosov med organizmi in njihovim okoljem oz. odnosov med organizmi na eni strani, na drugi strani pa razumevanje ekoloških in statističnih metod, omogoča uspešno razreševanje biodiverzitetnih vprašanj.

Biodiverzitetne raziskave, ki temeljijo na **primarnih podatkih**, se praviloma pričnejo s terenskim in/ali laboratorijskim zbiranjem podatkov oz. izvajanjem meritev. Drug pristop je analiza že zbranih, praviloma objavljenih podatkov, iz več raziskav, kar imenujemo **meta-analiza** (Harrer in sod. 2019). Pogosta je **kombinacija** obeh pristopov, v katerem primarne podatke dodamo k že objavljenim podatkom ter združene podatke na novo analiziramo in interpretiramo.

Relevantnost sklepov raziskave zagotovimo na več načinov:

- Z **nepriustranostjo** (ang. Impartiality) oz. z nepristranskim (ang. Unbiased) izvajanjem meritev in sprejemanjem sklepov, t.j. brez vnaprejšnjih odločitev o rezultatih.
- S **ponovljivostjo** meritev (ang. Repeatability), t.j. zmožnost pridobitve enakih rezultatov z uporabo istih metod.
- S **preverljivostjo** rezultatov (ang. Reproducability), kar pomeni, da dobijo drugi raziskovalci, ki raziskavo ponovijo, identične rezultate.
- Z **verodostojnostjo** raziskave (ang. Validity), kar pomeni, da so načrt raziskave in uporabljene metode vzorčenja ustrezne za odgovor na zastavljeno vprašanje.

V poglavju o vzorčenju so predstavljene naslednje vsebine:

- Vzorec.
- Način izbire vzorčenih lokalitet v biodiverzitetni raziskavi.
- Vzorčenje v biodiverzitetni raziskavi.
- Organizacija podatkov.
- Odprta znanost.
- Zemljevidi.
- Naravoslovne zbirke.

16.1 Vzorec

Vzorec (ang. Sample) v biodiverzitetni raziskavi je enkraten popis izbranih organizmov, lahko tudi drugih biotskih in abiotskih dejavnikov, na definiranem območju – vzorčeno mesto (npr. v točki, transektu, kvadratu, poligonu, volumnu itd.) in v definiranem časovnem obdobju (npr. v uri, dnevu itd.).

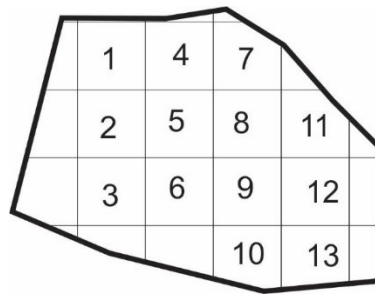
Popis vrst in njihove številčnosti na izbranem območju izvajamo z njihovim prepoznavanjem (identifikacijo) in s preštevanjem osebkov. Po navadi ni mogoče pregledati celotnega območja in prešteti vseh osebkov. Ta izziv rešimo z odvzemom več vzorcev v različnih delih raziskovanega območja ob predpostavki, da imajo vzorci značilne lastnosti celotnega območja, t.j. da so reprezentativni. Za zagotovitev čim boljše reprezentativnosti vzorcev moramo skrbno načrtovati vzorčenja. Vzorčenje praviloma izvajamo s standardno vzorčno enoto, ki jo določimo glede na lastnosti vzorčenih organizmov. Najpogostejši enota vzorčenja sta kvadratna ali pravokotna ploskev z znano površino, kar imenujemo metoda kvadrata. Namen uporabe te metode je zagotavljanje primerljivosti med vzorci. Izbira velikosti kvadrata je odvisna predvsem od velikosti organizmov, ki so predmet raziskave. Pri vzorčenju metuljev na travniku zadostuje transekt približno 5 x 100 m, pri vzorčenju gozdnih ptic transekt 20 x 1.000 m. Pred izvedbo popisa ocenimo, kolikšna natančnost popisa je potrebna za določen namen ter koliko raziskovalnega dela in časa je potrebno vložiti v izvedbo načrtovanega popisa.

16.2 Način izbire lokalitet v biodiverzitetni raziskavi

Pri izbiri lokalitet za vzorčenje živih bitij razlikujemo štiri osnovne pristope: enostavno naključno vzorčenje, stratificirano naključno vzorčenje, sistematično vzorčenje in vzorčenje s presojo.

Enostavno naključno vzorčenje (ang. Simple Random Sampling) je vzorčenje, pri katerem imajo vsi osebki populacije enako možnost in verjetnost, da bodo zajeti v vzorec. Izbor lokalitet vzorčenja je v celoti naključen in zanj velja načelo verjetnosti. Za tak pristop k vzorčenju mora populacija izpolnjevati dva pogoja, biti mora homogena in brez trendov ali kopičenja osebkov (grupiranja) na posameznih mestih.

Priporočen način za izbiro vzorčenih mest je razdelitev celotne raziskovane ploskve (npr. habitata) v mrežo oštevilčenih celic (Slika 16.1), ki jo lahko projiciramo na zemljevid območja. Izbiro kvadratov za izvedbo popisa opravimo s pomočjo generatorju naključnih števil.



Slika 16.1 Raziskovano območje je razdeljeno v mrežo s 13 kvadrati. Izbira kvadratov je naključna.

Stratificirano naključno vzorčenje (ang. Stratified Random Sampling) je način vzorčenja, pri katerem populacijo ali območje vzorčenja razdelimo na več podskupin (stratumov) glede na njihove skupne značilnosti, ki jih druge skupine nimajo. V vsaki podskupini vzorčimo naključno, torej tako, da ima vsak osebek populacije enako verjetnost, da bo zajet v vzorec (tako kot pri enostavnem naključnem vzorčenju). Primer stratificiranega naključnega vzorčenja je predstavljen v Poglavju 7 Analiza združbe, kjer je opisano vzorčenje malih sesalcev v naključno izbranih ploskvah v dveh stratumih glede na preteklo rabo gozda (trajna raba, pašništvo).

Sistematično vzorčenje (ang. Systematic Sampling) je vzorčenje, pri katerem vzorčena mesta izberemo vnaprej na način, da vsaka enota vzorca predstavlja enak delež celotne populacije. Sistematično vzorčenje dosežemo z izbranim prostorskim in/ali časovnim zaporedjem odvzema vzorcev. Izvedba sistematičnega vzorčenja biodiverzitete zahteva veliko vložnega raziskovalnega dela. Primer sistematičnega vzorčenja je predstavljen v nadaljevanju. V Projektu 16.1 Jamska diverziteteta je opisano vzorčenje živih bitij in abiotskih dejavnikov v podzemeljskih jamskih habitatih.

V raziskavah biodiverzitete pogosto **vzorčimo s presojo** (avtoritativno vzorčenje). Izvajamo ga v primerih, ko se osredotočimo samo na del populacije ali na specifični življenjski prostor. Praviloma imamo precej znanja in izkušenj z organizmom, ki ga vzorčimo. Opisano selektivno vzorčenje je časovno manj zahtevno in potrebnih je manj vzorcev. Zelo je izpostavljeno napakam in napačnim presojam raziskovalca.

16.3 Načrt vzorčenja

Glede na specifične značilnosti skupine organizmov, ki jo nameravamo vzorčiti, pripravimo načrt vzorčenja, ki obsega naslednje.

Način vzorčenja – opredelitev kraja in časa vzorčenja; izbira metod vzorčenja, določitev natančnosti meritev in cenitev (npr. ugotavljamo samo prisotnost, ali prisotnost in številčnost; opredelitev metod zbiranja podatkov: z opazovanjem, z ulovom, z ulovom označevanjem in ponovnim ulovom itd.).

Usposobljenost vzorčevalca – vzorčenje organizmov lahko izvaja samo ustrezno strokovno usposobljen strokovnjak z izkušnjami, zagotoviti moramo ustrezno delovno ali zaščitno opremo, aparature za izvajanje meritev, opremo za opazovanje ali lov organizmov ter imeti urejeno dokumentacijo za izvajanje vzorčenja (kjer je ta potrebna).

Odvzem vzorcev in izvajanje meritev na terenu – vzorčenje reprezentativnih vzorcev živih bitij ob sočasnem upoštevanju znanstvenih ciljev in etike (npr. odvzem minimalnega števila osebkov iz narave, ki še omogoča izpolnitev cilja raziskave).

Označevanje vzorcev in izdelava zapisnika o vzorčenju – sprotno označevanje vzorcev na terenu in vodenje evidence o drugih podatkih posameznih lokalitet (GPS koordinate, datum vzorčenja in drugi specifični podatki (npr. opis lokalitete, meritve itd.) v skladu z namenom raziskave.

Shranjevanje in transport vzorcev – popisovanje živih organizmov, ali ustrezno konzerviranje organizmov in označitev vzorcev ter transport v laboratorij za nadaljnje preiskave.

Vnos merskih podatkov v podatkovno zbirko – zasnova zbirke podatkov za vnos merskih podatkov v stopnji načrtovanja raziskave.

Analiza in vrednotenje zbranih podatkov – statistična analiza podatkov v skladu s cilji raziskave.

Priprava poročila ali znanstvene publikacije – zaključek raziskave s pripravo ustrezne objave.



PROJEKT 16.1

Projekt »Jamska biodiverziteta«

Novak (2005) poroča, da je sistematično raziskal kopensko favno v jamah severne in osrednje Slovenije. V 37 letnem obdobju je vzorčil v 64 jamah, na 785 stalnih vzorčenih mestih, praviloma v vsakem letnem času enkrat in je skupaj zbral 3.416 vzorcev, 23.091 krat je zabeležil prisotnost 515 različnih organizmov, s skupno abundanco vseh taksonov 219.661 osebkov.

Metode vzorčenja je prilagodil glede na biologijo skupine. Talne nevretenčarje (npr. hrošči in druge žuželke, stonoge ipd., polže ipd.) je vzorčil s talnimi pastmi (lovne jame) in prešteval po podlagi, leteče nevretenčarje (npr. metulje, kožekrilce ipd.) je ročno prešteval

ali pobiral s sten, netopirje je prešteval, male sesalce (npr. miši, voluharice) je lovil s pastmi, večje vretenčarje je prešteval in beležil njihove sledi (npr. iztrebki, stopinje ipd.). Posamezno vzorčenje je definiral kot enkratno registracija vseh organizmov (ne glede na metodo), na enem stalnem vzorčenem mestu, v enem časovnem oknu.

Ob vsakem vzorčenju je registracijo (opazovanje) organizma zabeležil in opremil s/z:

- taksonomsko oznako (vrstna pripadnost in taksonomska klasifikacija živali),
- abundanco (število osebkov),
- metodo, na kakšen način je potrdil in preštel organizme,
- geolokacijo (z geografskima koordinatama in nadmorsko višino jame in oznako vzorčenega mesta) in
- časovno definicijo (z navedbo datuma vzorčenja).

Okoljske dejavnike na posameznem vzorčenem mestu je razdelil na fiksne in variabilne. Variabilne dejavnike okolja je izmeril ob posameznem vzorčenju, to so:

- temperatura zraka (°C),
- temperatura tal (°C),
- zračna vlaga (%),
- vlaga tal (%) in
- prepih (hitrost premikanja zraka v m/s).

Fiksni dejavniki se nanašajo na lastnosti vzorčenega mesta in obsegajo:

- oddaljenost vzorčenega mesta od (najbližjega) vhoda v jamo (m),
- oddaljenost od površja (m),
- presek rova (m²),
- pH,
- žaroizguba,
- konsistenco podlage (kategorije) in
- osvetljenost (kategorije).

Vsako posamezno jamo je klasificiral glede na:

- nastanek (naravna, umetna),
- tip zunanje vegetacije in
- kamninsko podlago.

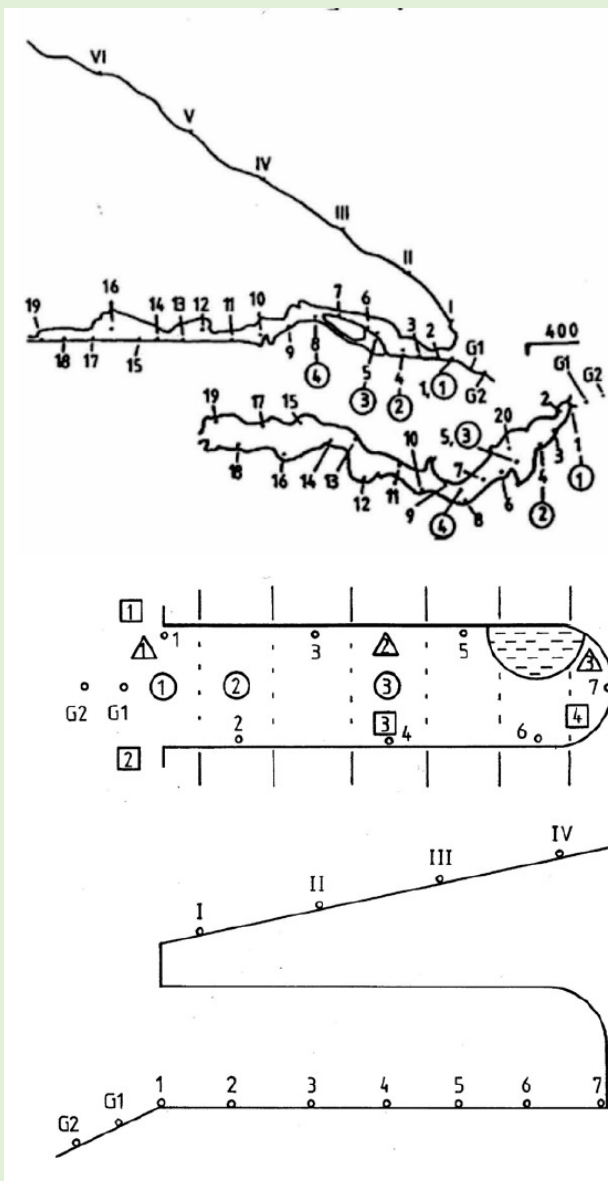




PRIMER 16.1

Vzorčenje jamske biodiverziteta – vzorčna mesta

Novak (2005) je v okviru ekoloških raziskav jamske biodiverzitete zastavil strukturirano raziskovalno vprašanje: »katere kopenske živali naseljujejo jamske habitate, kolikšno je vrstno bogastvo in vrstna pestrost živali, kako bogastvo in pestrost variirata v prostoru in času ter kolikšen del te variabilnosti lahko pojasnimo z ekološkimi dejavniki?«. Glede na prisotnost živali v jamah jih je klasificiral v tri kategorije: troglokseni (občasno zaidejo v jamsko okolje), troglofili (del življenja ali posamezni življenjski cikel preživijo v jami) in troglobionti (cel življenjski cikel so v jami).



Slika 16.2 Načrt jame Babja luknja, zgoraj vzdolžni prerez, spodaj prečni prerez jame (vzeto iz Novak 2005). Z arabskimi številkami je označenih 19 vzorčenih mest v jami; G1 in G2 sta oznaki za vzorčna mesta pred jamo; z rimskimi številkami so označena vzorčna mesta nad jamo; s krogi s številkami so označene pasti za male sesalce.

Slika 16.3 Shematski prikaz razdelitve jame v odseke in razporeditev vzorčenih mest (vzeto iz Novak 2005).

Zgoraj. Shema prečnega prereza jame z regionalizacijo na pet metrske odseke (prečne črte); z označenimi vzorčenimi mesti. G1 in G2 pred jamo; 1, 2, 3, 4, 5, 6 in 7 vzorčna mesta v jami; s pravokotniki in krogi s številkami so označena mesta s pastmi.

Spodaj. Shema vzdolžnega prereza jame in območja nad jamo. Z arabskimi številkami so označena vzorčna mesta v jami; z G1 in G2 vzorčna mesta pred jamo; z rimskimi številkami so označena vzorčna mesta nad jamo.

Glede na kompleksnost raziskave, ki je potekala v 64 jamah, je vzorčenje potekalo na dveh nivojih. Znotraj posamezne jame je potekalo sistematično vzorčenje, jama je bila razdeljena v bolj ali manj enakomerne odseke (glede na specifične terena), v vsakem odseku

je bilo definirano vzorčeno mesto, na katerem je bil odvzet vzorec oz. več vzorcev (Sliki 16.2 in 16.3). Vzorčenje je potekalo v več jamah, ki so bile izbrane med množico jam po (bolj ali manj) naključnem vzorcu, vsaka posamezna jama predstavlja prostorsko in ekološko drugačno okolje, zato je potekalo tudi na nivoju stratificiranega naključnega vzorčenja.

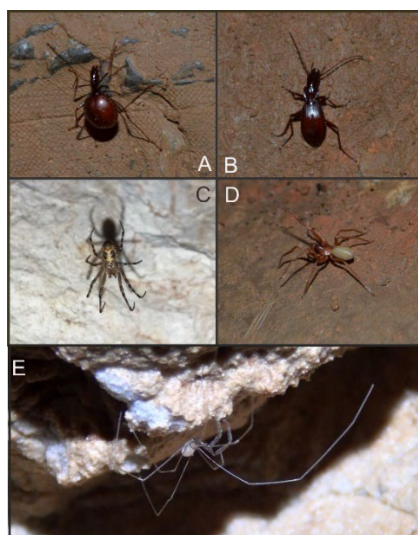


16.4 Organizacija podatkov

V biodiverzitetni raziskavi zbiramo štiri osnovne kategorije podatkov: 1. **biološki podatki**, taksonomija vrst z latinskimi imeni, na informacijo o vrstni pripadnosti se navezujejo biološke lastnosti vrste t.i. funkcionalni znaki (ang. Functional Traits) npr. masa, telesna dolžina, velikost legla, življenjska doba, mikrohabitat ... (Slika 16.4). 2. **podatki o številčnosti** (abundanca) oz. prisotnosti posamezne vrste (prezenca). 3. **časovni podatki**, navajamo jih z datumom, lahko tudi uro opazovanja oz. izvajanja vzorčenja. 4. **podatki o vzorčenem mestu**, geolokacija vzorčenega mesta, navajamo z geografsko širino in geografsko dolžino ter imenom bližnjega kraja. Glede na namen raziskave lahko izvedemo meritve fizikalnih, geoloških, klimatskih, kemijskih in drugih lastnosti za vzorčeno mesto. Namen zbiranja navedenih meritev je, da z njimi razlagamo variabilnost v biodiverziteti.

Glede na namen in cilj raziskave definiramo nabor spremenljivk. V stopnji načrtovanja raziskave moramo dobro pretehtati namen izvajanja posamezne meritve, protokol zapisovanja meritev in statistično analizo za vrednotenje meritev. Podatke praviloma zapisujemo sproti v vnaprej pripravljen formular, v kabinetu jih nato prenesemo v Excelovo preglednico. Za izvajanje prostorskih, fizikalnih in kemičnih meritev lahko uporabimo beležnike podatkov (data-loggerje), ki omogočajo prenos zbranih meritev v ustrezno aplikacijo.

Temeljni metodološki izziv za raziskovalca je organizacija podatkov različnih kategorij za vrednotenje biodiverzitete v prostoru, času in odziva biodiverzitete na biotske, fizikalne, kemijske, klimatske in druge gradiente. Ena od možnih rešitev je predstavljena v primeru, v kateri so raziskovalci oblikovali podatkovno bazo v Accessu, podatke zbrane v več Excelovih preglednicah so povezali med seboj v relacijsko matriko, povezali so vzorčena mesta in njihove lastnosti, vrste organizmov z njihovo abundanco in njihove lastnosti, ter zaporedje vzorčenj v prostoru in času.



Slika 16.4 Nekaj primerov jamskih vrst kopenskih členonožcev. **A.** drobnovratnik *Leptodirus hochenwartii*, troglobiontna vrsta hrošča, prva opisana jamska žival. Foto. Peter Kozel. **B.** veliki jamski brzec *Laemostenus schreibersii*, troglofilna vrsta hrošča. Foto. Peter Kozel. **C.** evropski jamski pajek *Meta menardi* troglofilna vrsta pajka. Foto. Franc Janžekovič. **D.** podzemeljski šesterookec *Stalita taenaria*; prvi opisan troglobiontni in brezoki pajek. Foto. Peter Kozel. **E.** Ferranova hadžinija *Hadzinia ferrani*, brezoka troglobiontna vrsta suhe južine. Foto. Peter Kozel.



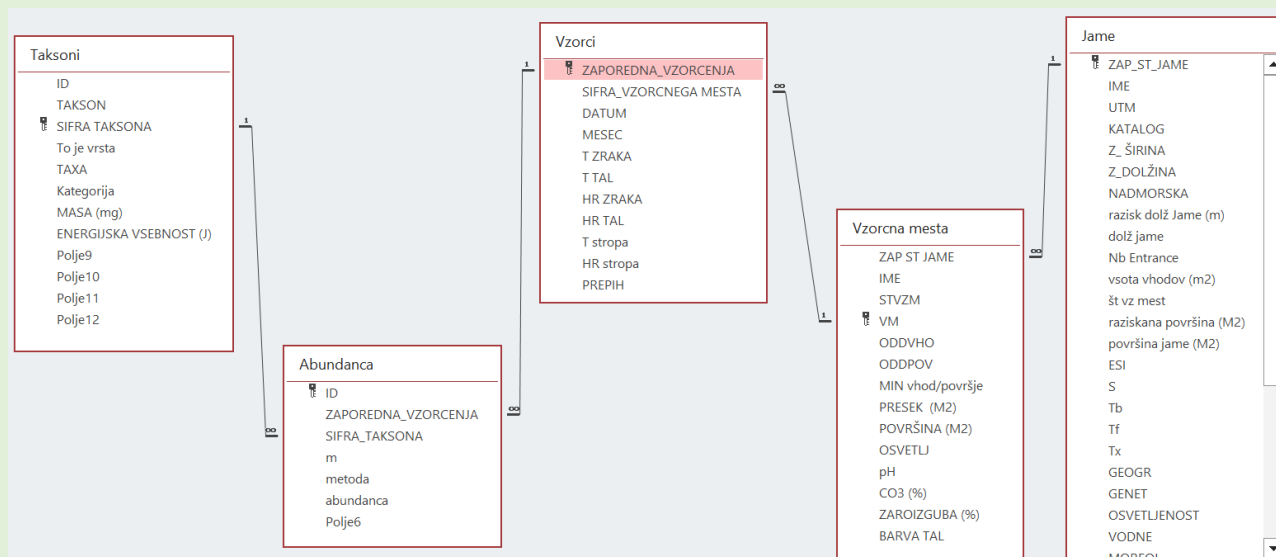
PRIMER 16.2

Vzorčenje jamske biodiverzitete – podatkovna baza

Podatke o vzorčenju biodiverzitete jamskih kopenskih živali ter ekoloških fiksnih in variabilnih spremenljivkah v jamskem okolju ter časovne in prostorske informacije o vzorcih smo razporedili v pet ločenih Excelovih preglednic. Sestavili smo Accessovo bazo, ki sestoji iz petih med seboj povezanih preglednic različnih dimenzij (vrstic x stolpcev):

- taksoni: 515 x 5 (515 taksonov, s petimi spremenljivkami),
- jame: 64 x 20 (64 jam, vsako jamo spremlja 20 spremenljivk),
- vzorčena mesta: 784 x 14 (784 vzorčenih mest s 14 spremenljivkami),
- vzorci: 3.416 x 11 (3.416 vzorcev z 11 spremenljivkami) in
- abundanca: 23.091 x 5 (23.091 abundanc posameznih taksonov s 5 spremenljivkami).

Vseh pet preglednic smo uvozili v Access in jih med seboj povezali v relaciji 1:vse (Slika 16.5). Navedeno pomeni, da je vzpostavljena referenčna integriteta, vsak podatek ima enoznačno povezavo na vse preostale podatke. Vzpostavljena je podatkovna baza, ki omogoča logično navzkrižno povpraševanje v vse smeri.



Slika 16.5 Logična struktura podatkovne zbirke z referenčno integriteto med petimi preglednicami.



16.5 Politike odprte znanosti in ravnaje z raziskovalnimi podatki

V Sloveniji, v drugih državah Evropske Unije in marsikje drugod po svetu uveljavljamo načela odprte znanosti. **Odprta znanost** vključuje prosto deljenje vseh rezultatov raziskovalnega dela, vključno z **raziskovalnimi podatki**. V domači zakonodaji načela odprte znanosti določajo 40., 41. in 42. člen *Zakona o znanstvenoraziskovalni in inovacijski dejavnosti (ZZrID)* (2021). 40. člen ZZrID definira: »*Odprta znanost obsega predvsem odprt dostop do raziskovalnih rezultatov, vrednotenje kakovosti in vpliva znanstvenoraziskovalnega dela z uporabo odgovornih metrik ter povezovanje in vključevanje zainteresirane javnosti v raziskovalni proces...*«. 41. člen ZZrID govori o tem, da: 1. morata financer in raziskovalec v okviru raziskav, sofinanciranih z javnimi viri zagotoviti **odprti dostop** do vseh recenziranih znanstvenih objav in raziskovalnih podatkov ter drugih rezultatov raziskav; 2. da morajo odprti raziskovalni podatki biti objavljeni ali drugače dostopni na način, ki omogoča njihovo najdljivost, dostopnost, interoperabilnost ter vnovično uporabo po **načelih FAIR** (ang. Findability, Accessibility, Interoperability, Reusability).

Na ravni EU je odprta znanost opredeljena v *Paktu za raziskovanje in inovacije v Evropi*, <https://www.consilium.europa.eu/sl/policies/european-research-area/>. Javna objava raziskovalnih podatkov je postala obvezna leta 2021 z začetkom finančnega mehanizma Obzorje Evropa (ang. Horizon Europe).

Definicija OECD o raziskovalnih podatkih: »*Raziskovalni podatki so definirani kot stvarni zapisi (numerični rezultati, besedilni zapisi, slikovno in zvočno gradivo), ki se uporabljajo kot primarni viri za namene znanstvenih raziskav in so v znanstveni skupnosti splošno sprejeti kot nujni za potrditev*

raziskovalnih izsledkov. Nabor raziskovalnih podatkov predstavlja sistematično, delno predstavitev raziskovalne tematike. Ta pojem ne vključuje: laboratorijskih dnevnikov, preliminarnih analiz, osnutkov znanstvenih člankov, načrtov bodočih raziskav, strokovnih recenzij, komunikacije s kolegi in predmetov (npr. laboratorijskih vzorcev, bakterijskih sevov in testnih živali, kot so miši)« (OECD Principles and Guidelines for Access to Research Data from Public Funding).

Centralna tehniška knjižnica Univerze v Ljubljani je razvila spletno mesto za ravnanje z raziskovalnimi podatki DiRROS Data: <https://dirrosdata.ctlk.uni-lj.si/>. DiRROS Data raziskovalcem nudi individualno svetovanje o pripravi načrta za ravnanje z raziskovalnimi podatki, izbiro in formatiranje podatkov za javno deljenje ter svetovanje o izbiri primernega repozitorija. Po gradivu s spletne strani DiRROS Data (2023) so povzeta priporočila v nadaljevanju.

DiRROS priporoča raziskovalcem, da javnosti omogočimo dostop najmanj do podatkov, ki so nujno potrebni, da lahko drugi raziskovalci ponovijo eksperimente. To pomeni najmanj podatke, ki so bili podlaga za objavo raziskave. V nabor podatkov lahko vključuje tudi surove podatke, vendar to ni nujno. Presoja je vedno na strani avtorja. Kjer je smiselno lahko v nabor odprtih podatkov vključite surove podatke iz območja okrog povprečja in obeh ekstremov porazdelitve ali pa nek drug smiseln presek podatkov. Nabor raziskovalnih podatkov naj vključuje spremljevalne dokumente (npr. raziskovalne protokole), izvorno računalniško kodo ipd., če brez tega podatki niso berljivi ali razumljivi. V primeru, da gre za edinstvene podatke, ki jih časovno in prostorsko ni mogoče ponovno pridobiti, priporočamo, da delite vse podatke, tudi surove, razen če gre za upravičene izjeme iz razloga varstva intelektualne lastnine ali drugih vidikov.

16.5.1 Tipi in formati podatkov

Glede na način zbiranja podatkov in njihove dostopnosti, jih razvrščamo v več kategorij. Navajamo samo formate podatkov relevantne za področje biodiverzitetnih raziskav.

Podatki pridobljeni z opazovanjem. To so podatki pridobljeni na mestu vzorčenja (*in situ*) in jih ni mogoče ponovno pridobiti, poustvariti ali nadomestiti. V to skupino sodi monitoring okolja in terenska opazovanja, vključno s popisom biodiverzitete.

Podatki pridobljeni z eksperimentiranjem. To so podatki pridobljeni pod nadzorovanimi pogoji na mestu vzorčenja (*in situ*) ali v laboratoriju, načeloma so ponovljivi. V to skupino sodi npr. mikroskopiranje biološkega materiala (morfologija organizmov), sekvenciranje DNK, kemijske analize zgradbe biološkega materiala (kemijska sestava organizmov) ipd.

Izpeljani ali združeni podatki. To so podatki pridobljeni z združevanjem podatkovnih nizov in podatkovnih baz (meta-analize) (Meta-analysis 2023), izpeljane spremenljivke, podatki pridobljeni s podatkovnim rudarjenjem ipd.

Podatki pridobljeni s simulacijami. To so podatki, ki so rezultat modelov in modeliranja o delovanju dejanskega ali teoretičnega sistema. Modeli, ki napovedujejo razvoj dogodkov in vhodni podatki so lahko pomembnejši od samega izhoda modela. V to skupino sodijo napovedni modeli npr. razvoja podnebja, spreminjanja gladine morske vode, odziva biodiverzitete na spreminjanje temperature ipd.

Referenčni ali uradni podatki. To so urejene statistične ali dinamične zbirke podatkov, praviloma o javnih zadevah ali na izbrano tematiko. V to skupino podatkov sodijo npr. državni statistični popisi, javne prostorske informacije (geografske, geološke, geodetske, klimatske ipd.), genske knjižnice z nukleotidnimi in aminokislinskimi zaporedji, knjižnice funkcionalnih znakov organizmov ipd.

16.5.2 Načelo FAIR

Odprti, oz. javno dostopni podatki morajo izpolnjevati načela FAIR, to pomeni:

1. Da so **odloženi** v zaupanja vrednem repozitoriju, kjer so **opremljeni** s trajnim identifikatorjem in bogatimi metapodatki (ang. Findability – najdljivost in Accesibility – dostopnost).
2. Da so **opisani** v formalnem, splošno dostopnem in široko uporabljanem jeziku za razširjanje znanja (ang. Interoperability – interoperabilnost).
3. Da so **licencirani** z odprto licenco in opremljeni z vsemi informacijami (npr. metodami, protokoli, programsko opremo ipd.), ki drugim raziskovalcem omogočajo razumevanje in **ponovno uporabo** (ang. Reusability – ponovna uporabljivost).

Raziskovalni podatki arhivirani v repozitorij se citirajo kot spletni vir. Prav tako arhivirane raziskovalne podatke vpišemo v COBISS pod rubriko *2.20 Zaključena znanstvena zbirka raziskovalnih podatkov*.

16.6 Zemljevid

Zemljevid (sin. karta) je grafični prikaz porazdelitve podatkov, dogodka, vizualizacija vzorcev ali pojava v prostoru, npr. porazdelitev vzorčenih mest ali lokalitete najdb izbrane vrste organizma. Razlikujemo različne tipe zemljevidov: statistična karta, kartodiagram, simbolna karta, tematska karta.

Statistična ali koropletna karta je kartografski prikaz ploskovne vizualizacije pojava. Pojav ali več pojavov je prikazanih s senčenjem ali barvnim likom. V biodiverzitetnih prikazih je pogosta predstavitev areala oz. razširjenosti izbrane vrste, npr. geografska razširjenost debelorepega skakača (Slika 12.2).

Kartodiagram so na geografsko podlago umeščeni grafikoni (stolpci, krogi, črtni diagrami ...). Primerni so za prikaz enega, redkeje več pojavov, z njim izrazimo medsebojna razmerja. Praviloma so narejeni v kombinaciji s koropletno karto, primer kombiniranja koropletno in kartodiagrama je Slika 12.5.

Simbolne karte predstavljajo v prostor umeščene simbole, ki so vezani na lokacijo pojava v prostoru. Simboli so lahko različnih oblik in barv, z njimi podajamo informacijo o porazdelitvi pojava. Simboli in barve predstavljajo diskretne vrednosti, npr. lokalitete več vrst organizmov. Posebna oblika simbolnih kart je **zemljevid s točkami** (ang. Dot Distribution Map, Dot Density Map), s točkami predstavimo prisotnost oz. prostorski vzorec obravnavanega fenomena. **Ena na ena** (ang. One-to-one) je oblika zemljevida s točkami, ki predstavlja prostorski vzorec enega samega dogodka oz. pojava. Primer je predstavitev prostorske razporeditve vzorčenih mest, vsaka točka predstavlja geografski položaj posamezne lokalitete. **Ena na mnogo** (ang. One-to-many) je oblika zemljevida s točkami, v katerem vsaka točka predstavlja več kot samo en dogodek oz. pojav. Število informacij, ki jih predstavlja posamezna točka, določi avtor glede na podatke, s katerimi razpolaga oz. jih želi predstaviti. Predstavitev prostorske razporeditve izbranih vrst oz. razširjenost posamezne vrste, predstavimo z različno barvo točk za vsako izbrano vrsto. Posamezna točka vsebuje dve informaciji: 1. informacijo o prostorski razporeditvi vrste ali pojava in 2. informacijo o imenu vrste oz. pojava.

Tematska karta, izdelana na podlagi geometrične metode je prikaz podatkov na pravilni kvadratni mreži, ki jo tvorijo celice enakih oblik in velikosti. To je najpogostejši način predstavitve prisotnosti izbrane vrste organizmov. V biodiverzitetnih kartiranjih živali so najpogostejše mreže velikosti 10 x 10 km in 50 x 50 km, na ta način je narejena večina atlasov razširjenosti, npr. ptic in sesalcev.



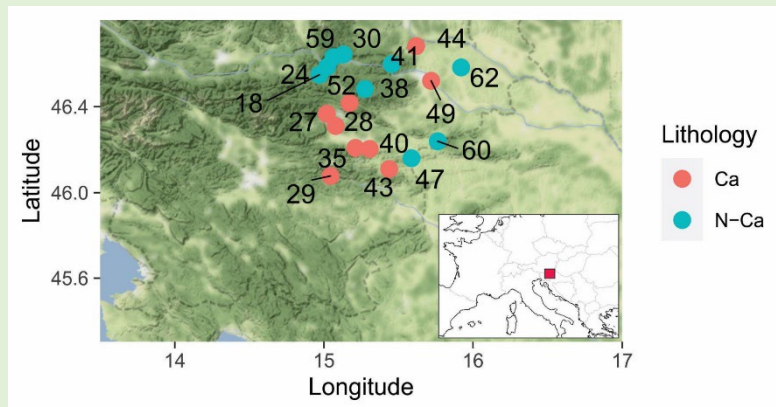
PRIMER 16.3

Vzorčenje jamske biodiverziteta – zemljevid z lokalitetami jam

Osnova za izris zemljevida so geografske koordinate: geografska širina in geografska dolžina (Preglednica 16.4). Zemljevid (Slika 16.6) izdelamo v RStudio.

Preglednica 16.4 Oznake in geografske koordinate jam

| ID | LAT | LON | Litologija |
|----|-------|-------|------------|
| 18 | 46,55 | 14,97 | N-Ca |
| 24 | 46,58 | 15,02 | N-Ca |
| 27 | 46,37 | 15,02 | Ca |
| 28 | 46,31 | 15,08 | Ca |
| 29 | 46,08 | 15,04 | Ca |
| 30 | 46,64 | 15,13 | N-Ca |
| 35 | 46,21 | 15,21 | Ca |
| 38 | 46,48 | 15,28 | N-Ca |
| 40 | 46,20 | 15,31 | Ca |
| 41 | 46,60 | 15,46 | N-Ca |
| 43 | 46,11 | 15,44 | Ca |
| 44 | 46,68 | 15,62 | Ca |
| 47 | 46,16 | 15,59 | N-Ca |
| 49 | 46,52 | 15,72 | Ca |
| 52 | 46,42 | 15,17 | Ca |
| 59 | 46,63 | 15,06 | N-Ca |
| 60 | 46,24 | 15,76 | N-Ca |
| 62 | 46,58 | 15,92 | N-Ca |



Slika 16.6 Razporeditev 18 jam v S in SV Sloveniji. Rdeča barva Ca – jame v karbonatni kamnini. Modra barva N-Ca – jame v nekarbonatni kamnini.

```
R > package = »ggplot2«, »ggmap«, »ggrepel«, »rworldmap«, »rgdal«
range(geo$LON)
range(geo$LAT)
glgmap <- get_stamenmap(bbox = c(left = 13.50, bottom = 45.30, right = 17.00, top = 46.80),
  maptype=c("terrain-background"), zoom = 8, crop=TRUE)
pgeo <- ggmap(glgmap)
p1 <- pgeo + theme(text = element_text(size=10))
p2 <- p1 + geom_point(aes(LON, LAT, color=Lithology), data = geo, pch=19, size=3) +
  geom_text_repel(data = geo, aes(x = LON, y = LAT, label = ID))
p3 <- p2 + xlab("Longitude") + ylab("Latitude")
p3
```



16.7 Naravoslovne zbirke

Preparirane oz. fiksirane vzorce organizmov hranimo na ustrezen način. Za posamezne skupine organizmov je vzpostavljena praksa in metodologija prepariranja in trajnega hranjenja. Preparate sesalcev in ptic hranimo kot mehove in skelete, nekatere žuželke kot suhe preparate (npr. metulji in večji hrošči) in večino kot mokre preparate v alkoholu. O prepariranju lahko bralec najde v drugi strokovni literaturi.

Raziskovalci in inštitucije, ki sistematično raziskujejo in zbirajo material v naravi, oblikujejo **naravoslovne zbirke**. Posebno vlogo v družbi imajo **naravoslovni muzeji**. To so vse ustanove, ki zbirajo, razstavljajo in raziskujejo gradivo, zbrano ali pridobljeno iz narave. Standardi za delo naravoslovnih muzejev so zapisani v ICOMovem *Kodeksu etike za naravoslovne muzeje* (2015) (ang. International Council of Museums ICOM, Mednarodni muzejski svet). Poslanstvo naravoslovnih muzejev je: 1. oblikovati in hraniti naravoslovne zbirke; 2. raziskovati in razlagati nova dognanja; 3. podpirati znanstvena dognanja in razvijati biološke postopke konzerviranja; 4. krepiti splošno razumevanje in spoštovanje narave; 5. sodelovati z javnostjo in oblikovati njeno lastno razumevanje (pomen) naravne dediščine, s katero se srečujemo v muzeju in v naravi.

Muzejske naravoslovne zbirke so tridimenzionalni arhiv naravnega sveta in odnosa družbe do okolja.

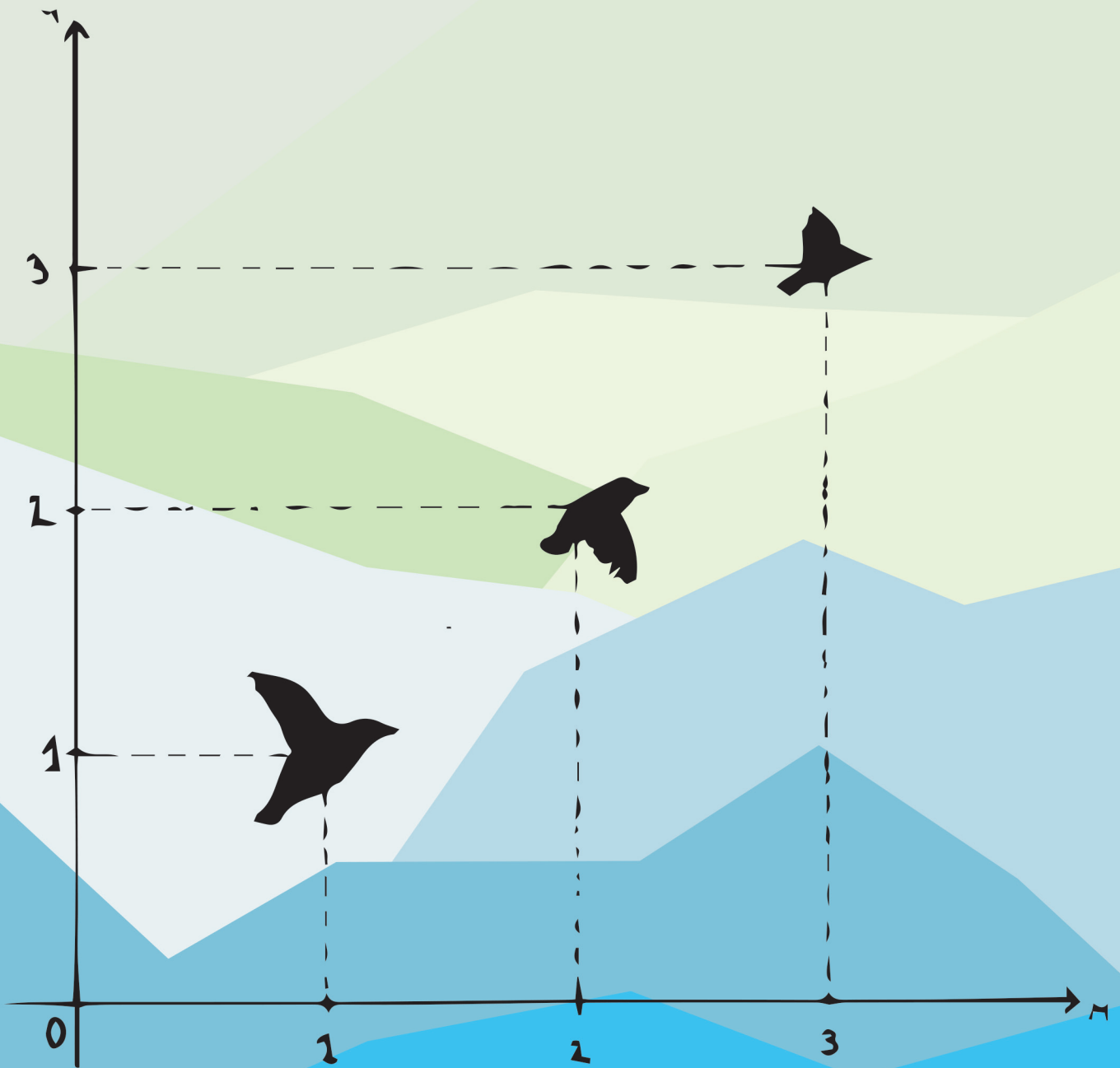
Hramba organizmov je dokaz o njihovem obstoju in o njihovih telesnih lastnostih. Znanstvenikom je zagotovljena možnost ponovnega analiziranja in vrednotenja z novimi metodami (npr. molekularne metode).

Zaključki

- Raziskave vzorcev in procesov biodiverzitete temeljijo na vzorčenju biote in meritvah abiotičnih dejavnikov okolja.
- Vzorčenje živih organizmov je specifično za vsako skupino posebej, zato ni standardiziranih protokolov, oblikovane so različne prakse vzorčenja.
- Raziskovalec/vzorčevalec je strokovno usposobljen z ustrezno izobrazbo in praktičnimi izkušnjami.
- Vzorčevalec mora imeti kompetence: poznavanje osnovne biologije organizmov in njihovo prepoznavanje, poznavanje značilnosti habitatov in mikrolokacijskih zahtev organizmov, načine njihovega vzorčenja, izbiro ustreznih metod glede na konkretne okoliščine v naravi, praktični odvzem v naravi, morebitno pripravo vzorcev na terenu, etiketiranje vzorcev in vodenje zapisa o vzorčenju, transport in hranjenje v laboratoriju, determinacijo vrst, trajno skladiščenje vzorcev in oblikovanje zbirke podatkov.
- Za izvedbo vzorčenja biote moramo pridobiti dovoljenje Agencije RS za okolje in soglasje Zavoda RS za varstvo narave, postopati etično do organizmov, da ne povzročata nepotrebne bolečina in da iz narave odvzame samo toliko, kolikor je nujno za izvedbo raziskave.
- Relevantnost zaključkov raziskave zagotovimo z: nepristranostjo – brez vnaprejšnjih odločitev o rezultatih; ponovljivostjo – z uporabo istih metod dobimo identične rezultate; preverljivostjo – ponovljene raziskave drugih raziskovalcev vrnejo identične rezultate; veljavnostjo – načrt raziskave in metode vzorčenja so ustrezne za odgovor na zastavljeno vprašanje.
- Vzorec v biodiverzitetni raziskavi je enkratni popis izbranih organizmov, lahko tudi drugih biotičnih in abiotičnih dejavnikov, na definiranem vzorčenem mestu, v definiranem časovnem oknu.
- Glede na prostor je vzorčenje lahko: enostavno naključno vzorčenje, stratificirano naključno vzorčenje, sistematično vzorčenje, vzorčenje s presojo.
- V biodiverzitetni raziskavi zbiramo štiri kategorije podatkov: biološki podatki, taksonomija vrst z latinskimi imeni, biološke lastnosti vrste; podatki o številčnosti posamezne vrste; časovni podatki; podatki o vzorčenem mestu, geolokacija ter fizikalne, geološke, klimatske, kemijske in druge lastnosti.
- Podatke vnašamo v Excelove preglednice, podatke več preglednic lahko med seboj povežemo v relacijsko matriko, podatkovna baza v Accessu.
- Rezultate raziskav in raziskovalne podatke delimo v odprti znanosti po načelu FAIR.
- Lokalitete vzorčenj predstavimo na zemljevidu.
- Zbirke organizmov trajno hranimo v naravoslovnih muzejih.

Literatura

- Bibby C., Burgess N., Mustoe S. 2000. Bird Census Techniques. Academic Press. London.
- DiRROS Data. 2022. DiRROS Data. Centralna tehniška knjižnica Univerze v Ljubljani. Pridobljeno 31.1.2023, iz: <https://dirrosdata.ctk.uni-lj.si/>.
- Ecological sampling methods. 2020. The Offwell Woodland & Wildlife Trust. Pridobljeno dne 1. 7. 2020, iz: http://www.countrysideinfo.co.uk/biol_sampl_cont.htm.
- Harrer M., Cuijpers P., Furukawa T.A., Ebert D.D. 2019. *Doing Meta-Analysis in R: A Hands-on Guide*. DOI: 10.5281/zenodo.2551803. https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/.
- Hill D., Fashman M., Tucker G., Shewry M., Shaw P. 2012. Handbook of Biodiversity Methods. Survey, Evaluation and Monitoring. Cambridge University Press. Cambridge.
- ICOMov kodeks etike za naravoslovne muzeje. 2015. ICOM – Mednarodni muzejski svet. Ljubljana.
- ISO/IEC 17025:2017. 2017. General requirements for the competence of testing and calibration laboratories. Pridobljeno 15. 11. 2020, iz: <https://www.iso.org/standard/66912.html>.
- Krebs C.J. 2014. Ecological Methodology. Harper Collins Publisher Inc. New York.
- Legendre P., Legendre L. 2012. Numerical Ecology. Elsevier. Amsterdam.
- McCabe D.J. 2011. Sampling Biological Communities. Nature Education Knowledge, 3(10): 63. <https://www.nature.com/scitable/knowledge/library/sampling-biological-communities-23676556/>.
- Meta-analysis. 2023. Wikipedia. Pridobljeno 31.1.2023, iz: <https://en.wikipedia.org/wiki/Meta-analysis>.
- Novak T. 2005. Kopenska favna votlin severne in osrednje Slovenije ter pregled sistematično ekološko raziskanih votlin. Acta carsologica, 34/1: 169–210.
- O'Donnell M.S., Ignizio D.A. 2012. Bioclimatic predictors for supporting ecological applications in the conterminous United States: U.S. Geological Survey Data Series 691.
- Odredba o merskih enotah*. 2001. Uradni list RS, št. 26/2001. Pridobljeno 31. 1. 2023 iz: <https://www.uradni-list.si/glasilo-uradni-list-rs/vsebina/2001-01-1594/>.
- Santos J.C., Fernandes G.W. 2020. Measuring Arthropod Biodiversity. A Handbook of Sampling Methods. Springer. Cham. DOI <https://doi.org/10.1007/978-3-030-53226-0>.
- Sikes R.S. 2016. Guidelines of the American Society of Mammalogists for the use of wild mammals in research and education. Journal of Mammalogy, 97(3): 663–688.
- Uredba o posebnih varstvenih območjih (območjih Natura 2000)*. 2018. Uradni list RS, št. 47/18. Pridobljeno 31. 1. 2023 iz: <http://www.pisrs.si/Pis.web/pregledPredpisa?id=URED283>.
- Uredba o zavarovanih prosto živečih živalskih vrstah*. 2019. Uradni list RS, št. 62/19. Pridobljeno 31. 1. 2023 iz: <http://www.pisrs.si/Pis.web/pregledPredpisa?id=URED2386>.
- Wheater C.P., Cook A.P., Bell R.J. 2020. Practical Field Ecology. Second Edition. Wiley Blackwell. Hoboken, NJ.
- Zakon o meroslovju – Zmer*. 2005. Uradni list RS, št. 26/05. Pridobljeno 31. 1. 2023 iz: <http://pisrs.si/Pis.web/pregledPredpisa?id=ZAKO1615>.
- Zakona o obranjanju narave – ZON-E*. 2020. Uradni list RS, št. 82/20. Pridobljeno 31. 1. 2023 iz: <http://pisrs.si/Pis.web/pregledPredpisa?id=ZAKO1600>.
- Zakona o znanstvenoraziskovalni in inovacijski dejavnosti (ZZrID)*. 2021. Uradni list RS, št. 186/21. Pridobljeno 31.1.2023, iz: <http://www.pisrs.si/Pis.web/pregledPredpisa?id=ZAKO7733>.



RAZLAGANJE STRUKTURE BIODIVERZITETNIH PODATKOV

Povzetek

17. poglavje: Razlaganje strukture biodiverzitetnih podatkov

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

V analizi biodiverzitetnih in ekoloških podatkov z regresijsko analizo opišemo povezanost med spremenljivkami. Lastnost je spremenljivka, ki opisuje variabilnost objekta. Kvantitativne lastnosti vključujejo intervalne (zvezne in diskretne) merske vrednosti, ki jih razvrstimo v naraščajočo vrsto realnih števil. S kvalitativnimi lastnostmi opišemo objekt, podatke pa uredimo le po frekvencah. Objekt raziskave je definiran vnaprej.

Z regresijsko analizo opišemo odnos med odvisno spremenljivko Y in eno ali več neodvisnimi spremenljivkami $X_1, X_2 \dots X_m$, da bi lahko napovedali ali predvidevali vrednosti y ob dani vrednosti x . Razmerja med spremenljivkama zapišemo z regresijsko enačbo:

$$Y = a + bX,$$

a je konstantni člen, b je smerni koeficient regresijske premice. Regresijska premica je po metodi najmanjših kvadratov položena skozi oblak točk v takšni smeri, da so ostanki ε najmanjši. Z regresijskim modeliranjem opišemo odnos med lastnostmi objekta, sklepamo o hipotezah ali razlagamo, napovedujemo in predvidevamo.

Z enostavno linearno regresijo podamo linearni odnos med spremenljivkama Y in X . Če neodvisna spremenljivka X ne izkazuje ostankov, je kontrolirana. Vsi ostanki se pripišejo odvisni spremenljivki. Ostanke računamo samo po vertikalni osi, kar velja za regresijski model I.

Kadar obe spremenljivki vsebujeta ostanke, uporabimo regresijski model II. Ostanke po metodi glavne osi računamo kot razdaljo od točke pravokotno na regresijsko premico. Spremenljivki

morata zadoščati določenim kriterijem. Izraženi morata biti v istih fizikalnih enotah, varianci napak morata biti podobni. V primeru različnih fizikalnih enot je naklon regresijske premice odvisen od odločitve, katera spremenljivka je na abscisi (X osi). Vpliv absolutnih vrednosti na rezultat odpravimo z standardizacijo, npr. metodo standardne glavne osi ali z normiranjem z rangiranjem (metoda rangirane glavne osi). Pri metodi navadnih najmanjših kvadratov predpostavimo, da spremenljivka X nima ostankov.

V biodiverzitetnih raziskavah imamo pogoste opravka z eno odvisno in več neodvisnimi spremenljivkami. Odnose med njimi analiziramo z multiplo linearno regresijo. Zgradimo model v obliki regresijske enačbe:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_mX_m + \varepsilon.$$

Pri oblikovanju optimalnega multivariatnega modela se v primeru testiranja ničelne hipoteze ravnamo po načelu varčnosti, z upoštevanjem značilnosti regresijskih koeficientov in prispevka posamezne spremenljivke k višanju R^2 ter k nižanju VIF. V primeru izbire modela z informacijsko – teoretičnim pristopom iz nabora kandidatnih modelov, izberemo modele z najnižjo vrednostjo AIC.

Če je Y nelinearna funkcija X , lahko za prilaganje uporabimo polinom vrednosti X . Model polinomske regresijske enačbe oblikujemo po načelu varčnosti in se glasi:

$$Y = b_0 + b_1X + b_2X^2 + \dots + b_kX^k + \varepsilon.$$

Povezanost simetričnih matrik preverjamo z Mantelovim testom.

Struktura podatkov (ang. Structure) je opis odnosa med vrednostmi spremenljivk glede na lastnosti in objekte. Odnos med spremenljivkami preverjamo vzdolž ekoloških gradientov ali med skupinami.

Lastnost (ang. Descriptor) je v biodiverzitetni oz. ekološki raziskavi spremenljivka (ang. Variable) ali več spremenljivk, s čimer podamo lastnosti objekta ali procesa. Npr. v biodiverzitetni raziskavi je prisotnost vrste organizma lastnost lokalitete ali pa je abundanca vrste lastnost lokalitete, v morfometrični raziskavi je vrsta organizma objekt raziskave in je lokaliteta s svojimi ekološkimi spremenljivkami lastnost okolja s katerimi razlagamo morfometrično variabilnost.

Ekološke spremenljivke imajo lahko kvalitativno ali kvantitativno naravo in jih po vsebini razvrstimo v različne kategorije: npr. biološke (npr. prisotnost, abundanca, biomasa vrste), fizikalne (npr. temperatura, elektoprevodnost), kemijske (npr. pH), časovne (npr. dan, mesec), klimatske (npr. padavine) itd. Kvantitativni tip spremenljivk, ki je v ekoloških raziskavah zelo pogost, vključuje vse intervalne, diskretne in ordinalne vrednosti, ki jih lahko razvrstimo v naraščajočo vrsto realnih števil. Kvalitativne spremenljivke vključujejo kvalitativni opis objekta (npr. habitat, ime vrste, barva, spol ipd.), ki jih uredimo samo po frekvencah. Binarne lastnosti so v biodiverzitetni analizi poseben kvalitativni tip spremenljivk s katerimi vrednotimo prisotnost-odsotnost, ki jih po navadi označimo z 1 ali 0, lahko pa tudi kot + ali –. V biodiverzitetnih raziskavah na ta način pogosto beležimo prisotnost vrst in tudi drugih ekoloških lastnosti.

Spremenljivka ima v regresijski analizi dogovorjeno vlogo, ko je predmet raziskave in njeno lastnost opazujemo oz. vrednotimo, ima vlogo **odvisne spremenljivke** (sin. odzivna spremenljivka). Na variabilnost odvisne spremenljivke vpliva **neodvisna spremenljivka**, ki ima vlogo razlagalne spremenljivke. Vloga posamezne spremenljivke je povezana z namenom raziskave.

Objekt (ang. Object) je v biodiverzitetni raziskavi predmet raziskave (npr. združba organizmov, lokaliteta, individualni organizem, proces ipd.), ki ga definiramo vnaprej (a priori) glede na namen in cilj raziskave. Objekt ima v podatkovnem nizu kategorično vrednost.

Razlaga strukture biodiverzitetnih podatkov ima tri glavne cilje: 1. **razlaganje** (ang. Explanation) ene lastnosti z drugo lastnostjo ali z več lastnostmi; 2. **napovedovanje** (ang. Forecasting) vrednosti ene lastnosti z drugo lastnostjo ali z več lastnostmi in 3. **predvidevanje** (ang. Prediction) vrednosti ene lastnosti z drugo lastnostjo ali več lastnostmi. Napovedovanje in predvidevanje imata različen pomen, čeprav je njuna interpretacija v ekoloških raziskavah pogosto pomešana. Predvidevanje se nanaša na izračun vrednosti odvisne spremenljivke v okviru nabora vrednosti razlagalnih

spremenljivk. Napovedovanje se nanaša na izračun vrednosti odvisne spremenljivke izven nabora vrednosti razlagalnih spremenljivk, vključenih v raziskavo.

Glavni viri za pripravo poglavja o regresijski analizi so Burnham in Anderson (2002), Fox in Weisberg (2019), James in sod. (2021), Jesenko in Jesenko (2007), Legendre in Legendre (2012), Venables in Ripley (2002). Posamezne analize smo izvedli v R programskem okolju (R Core Team 2022) in s statistični paketi: »geosphere« (Hijmans 2019), »lmodel2« (Legendre 2018), »vegan« (Oksanen in sod. 2019), »ggplot2« (Wickham 2016), »dplyr« Wickham in sod. (2020), MASS (Venables in Ripley 2002), "MuMIn" (Bartoń 2022).

V poglavju so predstavljene vsebine:

- Regresija.
 - Enostavna linearna regresija: model I.
 - Enostavna linearna regresija: model II.
 - Multipla linearna regresija.
 - Polinomska regresija.
- Primerjava matrik: Mantelov test.

17.1 Enostavna linearna regresija

Z enostavno **linearno regresijsko analizo** opišemo odnos med odvisno spremenljivko Y (sin. odzivna spremenljivka, ang. Dependent or Response Variable) in eno ali več neodvisnih spremenljivk $X_1, X_2 \dots X_m$, ki so razlagalne spremenljivke (ang. Independent or Explanatory Variables). Cilj regresijske analize je razlaga odnosa med spremenljivkama ter napovedovanje ali predvidevanje v povprečju točne vrednosti y ob dani vrednosti x . Odnos med spremenljivkama Y (ordinata) in X (abscisa) predstavimo grafično z **linearno regresijsko premico** in zapišemo z **linearno regresijsko enačbo**:

$$Y = a + bX + \varepsilon ,$$

koeficient a je **konstantni člen** (ang. Intercept) oz. presečišče regresijske premice na Y osi pri vrednosti $X = 0$; koeficient b je **regresijski koeficient** oz. smerni koeficient (tudi naklon regresijske premice; ang. Slope), ε je **slučajna napaka** (sin. ostanek; ang. Error, Residual). Z regresijsko enačbo lahko za vsako vrednost X izračunamo vrednost Y , izračun pa vrne v povprečju točno napoved.

Pri grafičnem prikazu v presečišču vrednosti posameznega dogodka (na Y in X osi) postavimo **točko**. Množica meritev oblikuje **oblak točk**, skozi katerega položimo linearno regresijsko premico (Slika 17.1). Odmik posamezne točke od regresijske premice je

ostanek. Regresijska premica je položena skozi oblak točk v taki smeri, da je vsota kvadratov ostankov najmanjša.

Neodvisna oz. razlagalna spremenljivka je lahko naključna (ang. Random) ali kontrolirana (znana vnaprej, tudi sistematska; ang. Fixed). Odvisna oz. odzivna spremenljivka je vselej naključna. Odločitev glede tega, ali je razlagalna spremenljivka naključna ali kontrolirana, je pomembna zaradi izbire ustrezne računske metode. **Model I** izberemo v primeru kontrolirane in **model II** v primeru naključne porazdelitve neodvisne spremenljivke.

Matematični model je enostavna matematična formulacija odnosa med spremenljivkami, s katerim preverjamo hipotezo. **Regresijsko modeliranje** uporabimo za iskanje **najboljšega prileganja** (ang. Goodness of Fit) matematičnega modela k dejanskim meritvam. Z namenom razlage odnosa, sklepanja, predvidevanja in napovedovanja iščemo najboljšo funkcijsko povezanost med spremenljivkami.

Pogosto je prva možnost uporabe regresijske analize želja po posplošitvi rezultatov na celotno populacijo. V takšnih primerih lahko sklepamo na podlagi intervala zaupanja, znotraj katerega se nahajajo prave vrednosti proučevane lastnosti. Druga možnost uporabe regresijske analize je, da s statističnim sklepanjem preverjamo tri tipe hipotez: 1. o obstoju odnosa; 2. o usmeritvi oz. predznaku odnosa; 3. po stopnji odnosa.

- Raziskovalno vprašanje ali med spremenljivko Y in spremenljivko X obstaja povezava, nadgradimo s hipotezama. V ničelni hipotezi H_0 trdimo, da povezave ni. V alternativni H_a trdimo, da povezava obstaja. Definiramo stopnjo zaupanja, ki je v naravoslovju praviloma 5 % oz. $\alpha = 0,05$. Izvedemo statistični test in na podlagi vrednosti tveganja p sprejmemo odločitev o izboru hipoteze. Če je stopnja tveganja večja od 5 % ($p > 0,05$), zavrnamo H_a . Če je stopnja tveganja manjša od 5 % ($p < 0,05$), H_a sprejmemo.
- V nekaterih primerih raziskovalno vprašanje zastavimo tako, da ugotovimo ali je odnos usmerjen oz. kakšen predznak ima odnos (negativen ali pozitiven). Pristop je identičen kot je opisan v prvem odstavku.
- Tretji primer preverjanja hipoteze je lahko oblikovan tako, da preverimo ali regresijski koeficient med spremenljivkama dosega izbrano vrednost. V tem primeru H_0 oblikujemo tako, da smerni koeficient dosega izbrano vrednost. Alternativno hipotezo H_a oblikujemo tako, da se smerni koeficient značilno razlikuje od izbrane vrednosti.

Tretja možnost uporabe regresijske analize je napovedovanja in predvidevanje vrednosti odvisne spremenljivke z uporabo regresijske enačbe.

V linearni regresijski analizi morajo spremenljivke izpolnjevati določene predpostavke:

- Spremenljivke so normalne porazdeljene. Njihovo porazdelitev vizualno preverimo s histogramom in Q–Q grafom; za približanje normalni porazdelitvi jih lahko transformiramo npr. z logaritmiranjem.
- Ostanke so normalno porazdeljeni, porazdelitev ostankov vizualno preverimo s histogramom in Q–Q grafom.
- Ostanke niso v korelaciji z neodvisnimi spremenljivkami; odnos med spremenljivkami in ostanki preverimo s korelacijo.
- Odnos med odvisno in neodvisno spremenljivko je homoskedastičen, kar vizualno preverimo z razsevnim diagramom.
- Vrednosti neodvisne spremenljivke morajo biti dovolj razpršene, kar preverimo z razponom.

Populacija (v statističnem kontekstu) sestavljajo vsi obstoječi objekti, ki so predmet raziskave. V praksi le izjemoma uspemo v analizo vključiti vse objekte v populaciji, zato analize praviloma izvajamo na **vzorcu** (ang. Sample). Vključitev zadostnega števila objektov v raziskavo, in s tem zadostitev zahtevam posamezne analize, je predstavljeno v Poglavju 18 Statistično sklepanje. V linearni regresijski analizi je uveljavljeno priporočilo, da za vsako spremenljivko, vključeno v model, povečamo število meritev za 20 in nikakor za manj kot 10. Na velikost vzorca smo pozorni že pri načrtovanju raziskave. Priporočljivo število meritev za regresijsko analizo izračunamo s pomočjo formule:

$$N \geq 20(\sum N_{odvisnih} + \sum N_{neodvisnih}),$$

kjer je N število meritev (velikost vzorca), $N_{odvisnih}$ je število odvisnih spremenljivk in $N_{neodvisnih}$ je število neodvisnih spremenljivk. V enostavni linearni analizi kjer sta vključeni samo ena odvisna in ena neodvisna spremenljivka je priporočljivo število meritev 40.

17.1.1 Enostavna linearna regresija – Model I

Enostavno linearno regresijo uporabimo za izračun linearnega odnosa med spremenljivkama Y in X z enačbo prve stopnje (model), ki ima obliko:

$$Y = b_0 + b_1X.$$

Termin »enostavna« se uporablja zato, ker je v odnos vključena samo ena razlagalna spremenljivka. Koeficient b_0 je konstantni člen oz. vrednost na Y osi, kjer je sečišče regresijske premice (pri vrednosti $X = 0$). Koeficient b_1 je naklon regresijske premice oz. regresijski koeficient.

Razdalja posamezne točke od regresijske premice je ostanek ali slučajna napaka in ga označimo z ε (epsilon). Ostanek ε je razlika med izmerjeno Y in izračunano \hat{Y} vrednostjo odvisne spremenljivke ob dani vrednosti neodvisne spremenljivke (Slika 17.1). Formulo zapišemo:

$$\varepsilon = Y - \hat{Y}.$$

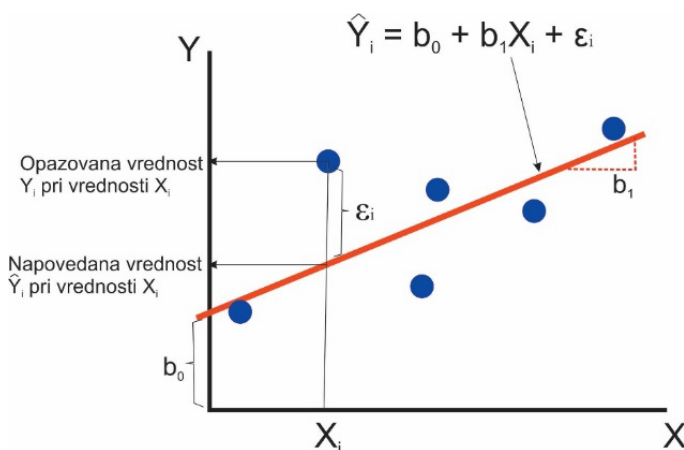
Regresijska premica je položena skozi oblak točk v taki smeri, da je vsota kvadriranih ostankov minimalna:

$$\min \sum (Y - \hat{Y})^2,$$

zato govorimo o **metodi najmanjših kvadratov** (ang. Least Squares). Z vključitvijo ostankov v izračun vrednosti \hat{Y} dobimo **linearni model**, ki ga zapišemo:

$$\hat{Y} = b_0 + b_1X + \varepsilon.$$

Ta oblika linearnega modela temelji na domnevi, da obstaja variabilnost od linearnega modela samo vzdolž Y osi, vzdolž X osi pa takih odmikov ni. Metoda se imenuje tudi **navadni najmanjši kvadrati OLS** (ang. Ordinary Least-squares). Posledično so razlike prikazane kot odmiki od regresijske premice samo v vertikali (navpičnici). V statistični praksi razlikam med opazovano in izmerjeno vrednostjo pravimo »napake« (ang. »Errors«).



Slika 17.1 Linearna regresijska premica in linearna regresijska enačba Modela I. Konstantni člen b_0 , regresijski koeficient b_1 in ostanek ε_i .



Foto 17.2 Turški gekon *Hemidactylus turcicus* je splošno razširjena vrsta v priobalnem območju Mediterana.

Foto Franc Janžekovič.

Kadar predpostavimo linearni odnos med spremenljivkama, v katerem neodvisna spremenljivka X ne izkazuje napake (je kontrolirana), pripišemo vsa odstopanja od regresijske premice (napake) odvisni spremenljivki Y (vertikalni ostanki). Govorimo o

regresijskem modelu I. Model I uporabimo pri analizi odnosa, pri katerem neodvisno spremenljivko kontrolira raziskovalec ali jo definira na terenu (npr. čas, prostor, definiran s koordinatami, volumen, razdalje, temperatura ipd).

V regresijski analizi nas zanima, kolikšen delež variance odvisne spremenljivke razloži neodvisna spremenljivka. Vrednost podamo s **koeficientom determinacije R^2** (sin. determinacijski koeficient), ki je enak količini (odstotku) razložene variance.



PRIMER 17.1

Plazilci v gradientu nadmorske višine – Model I

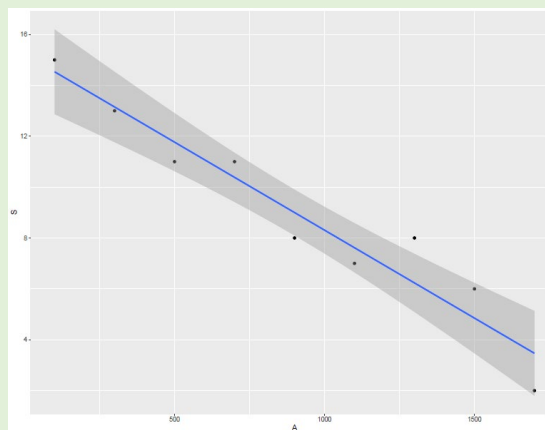
Zbrali smo podatke o številu vrst plazilcev (Slika 17.2) vzdolž gradienta nadmorske višine na Velebitu (glej Poglavlje 10). Podatki so oblikovani v pravokotni matriki velikosti 9×2 , razredi nadmorske višine so v vrsticah in število vrst v stolpcih (nadmorska višina A , število vrst plazilcev S). Nadmorska višina je neodvisna spremenljivka X in ne vsebuje napak, napake vsebuje samo odvisna spremenljivka Y število vrst. Ker vse napake pripišemo odvisni spremenljivki Y , izberemo Model I.

Preverjali smo ali je vrstno bogastvo plazilcev povezano z nadmorsko višino. Postavili smo H_0 – da povezave ni ($b_1 = 0$) in H_a – da povezava obstaja ($b_1 \neq 0$). Podatke smo vključili v enostavno linearno regresijsko analizo ELRA Model I. Model vrne visoko značilno povezavo $F_{1,7} = 87,52$ $p < 0,0001$, delež razložene variance $R^2 = 0,926$, tudi oba regresijska koeficienta (tako b_0 kot b_1) sta značilna (Preglednica 17.1), zato sprejmemo H_a . Regresijski koeficient ima negativni predznak in vrednost $b_1 = -0,007$. Ker je nadmorska višina podana v metrih, to pomeni, da z vsakim višinskim metrom, upade število plazilcev v povprečju za 0,007 vrste, oz. če pretvorimo, na vsakih 100 višinskih metrov se vrstno bogastvo plazilcev zmanjša v povprečju za 0,7 vrste. Odnos med obema spremenljivkama je predstavljen na Sliki 17.3. Regresijsko analizo smo izvedli v R okolju (R Core Team 2022), grafično predstavitev z razsevnim diagramom in regresijsko premico smo izdelali s paketom »ggplot2« (Wickhan 2016).

Statistično značilni model in visok delež razložene variance omogoča dobro predvidevanje števila vrst znotraj višinskega obsega od 0 do 1.800 m, ki je v skladu z modelom: $S = 15,225 - (0,0069 * A)$. Izračun vrst na nadmorski višini 1.000 m vrne vrednost $S = 8,33$. Z drugimi besedami, na nadmorski višini 1000 m predvidevamo 8 vrst plazilcev.

Preglednica 17.1 Rezultati enostavne linearne regresije

| Koeficient | Vrednost koef. | Std. napaka | t | p |
|------------|----------------|-------------|-------|----------|
| b_0 | 15,225 | 0,7672 | 19,85 | < 0,0001 |
| b_1 | -0,0069 | 0,0007 | 9,36 | < 0,0001 |



Slika 17.3 Regresija števila vrst plazilcev (S) v odvisnosti od nadmorske višine (A; v metrih) in regresijska premica (v modri barvi) s 95 % intervalom zaupanja (v sivi barvi).

```
R >
#linearna regresija
lm.nv<-lm(S~A, data=nv)
summary(lm.nv)
```

```
R > package = »ggplot2«
#graf, bi-plot+premica
nv.graf<-ggplot(nv, aes(x=A, y=S)) + geom_point()
nv.graf
na.graf.pr<-nv.graf+geom_smooth (method="lm")
na.graf.pr
```



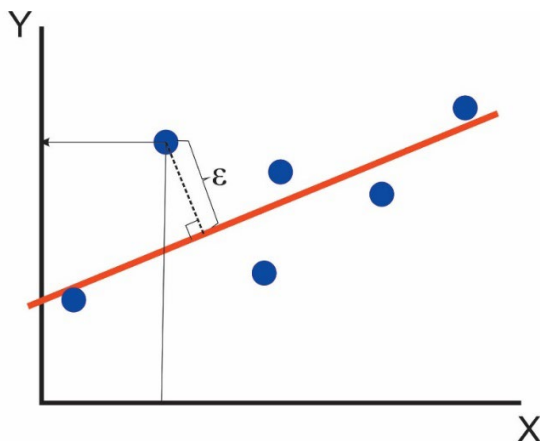
17.1.2 Enostavna linearna regresija – Model II

Kadar imata obe spremenljivki Y in X v modelu naključno porazdelitev, potem obe spremenljivki vsebujeta napake. V takšnih primerih imamo opravka z **enostavno linearno regresijo model II**. Izračun poteka po drugačnem postopku, kot je bil opisan v predhodnem poglavju (razlaga je opisana v Sokal in Rohlf 2012). Za vsako analizo posebej sprejmemo odločitev katera spremenljivka bo na grafičnem prikazu na Y in katera na X osi, oziroma katera bo odvisna in katera bo neodvisna spremenljivka. Odločitev je odvisna od namena in cilja raziskave.

Domneva v izračunu je, da so ostanki na Y spremenljivki in tudi ostanki na X spremenljivki, neodvisni, kar pomeni, da sta obe napaki naključnega izvora. Opisana domneva vpliva na izračun naklona regresijske premice oz. regresijskega koeficienta, ki ga računamo po **metodi največjega verjetja ML** (ang. Maximum Likelihood).

Za izračun enostavne linearne regresijske analize modela II obstaja več metod: glavna os, standardna glavna os, rangirana glavna os in navadni najmanjši kvadrati.

Metoda glavne osi MA (ang. Major Axis Regression) temelji na izračunu ostankov na podlagi Evklidske enačbe med regresijsko premico in posamezno točko (Slika 17.3). Razdalja med regresijsko premico in točko je pravokotna na premico in ni vzporedna niti z ordinato niti z absciso (Slika 17.4). Predpostavka v tej analizi je, da sta varianci obeh napak podobni. Druga predpostavka je, da sta obe spremenljivki v istih fizikalnih enotah, sicer je naklon premice različen (nehomogena dimenzionalnost podatkov), odvisno od tega, katera od spremenljivk je definirana kot neodvisna. Pri uporabi MA pogosto uporabljamo logaritemsko transformacijo podatkov.



Slika 17.4 Ostanek ε po modelu II. Ostanek je razdalja med točko in pravokotnico na regresijsko premico.

Metoda standardne glavne osi SMA (ang. Standard Major Axis); pri analiziranju spremenljivk v različnih fizikalnih enotah absolutne vrednosti vplivajo na izračun odnosa med njima. Zato je naklon regresijske premice odvisen od odločitve, katera spremenljivka bo neodvisna oz. bo na abscisi. Za odpravo vpliva absolutnih vrednosti spremenljivke predhodno **standardiziramo** po enačbi (z standardizacija):

$$z = \frac{x - \bar{x}}{StD}.$$

Standardizirane vrednosti so praviloma v naboru od -3 do 3. Vključitev z standardiziranih vrednosti v Model II enostavne linearne regresije se imenuje metoda standardne glavne osi.

Metoda rangirane glavne osi RMA (ang. Ranged Major Axis) je alternativa metodi SMA. Razlikujeta se v načinu normiranja. V RMA so podatki normirani z **rangiranjem** vzorčnih vrednosti po enačbi:

$$x' = \frac{x_i - x_{min}}{x_{max} - x_{min}}.$$

Normirane vrednosti so v naboru vrednosti od 0 do 1.

Metoda **navadnih najmanjših kvadratov OLS** (ang. Ordinary Least Squares) ima identičen postopek izračuna kot v Modelu I. Predpostavka v izračuna je, da spremenljivka X nima ostankov, z drugimi besedami, napaka neodvisne spremenljivke $X = 0$. V tem pogledu sta metodi OLS v Modelu I in II identični.

Kriteriji za izbiro ustrezne metode

Izbira ustrezne metode regresijskega modela II temelji na osnovi porazdelitve napak ε . Na tej osnovi so bila izdelana naslednja priporočila:

- Napaka spremenljivke Y je bistveno večja kot napaka spremenljivke X : OLS.
- Obe spremenljivki Y in X sta približno normalno porazdeljeni ->
 - Obe spremenljivki sta v istih fizikalnih enotah, njuni napaki sta podobni: MA.
 - Spremenljivki sta v različnih fizikalnih enotah ->
 - Ne vsebuje osamelcev, korelacijski koeficient ni signifikanten: RMA.
 - Lahko vsebuje osamelce: SMA.
- Porazdelitev podatkov zelo odstopa od normalne porazdelitve; povezava med spremenljivkama Y in X je linearna: OLS.
- Cilj napovedovanje vrednosti je \hat{Y} : OLS.
- Cilj je predvidevanje: MA.



PRIMER 17.2

Razmerje med dolžino in maso pri polhu Model II – podatki

Iz podatkovne baze Prirodoslovnega muzeja Slovenije smo pridobili telesne dolžine in mase 299 polhov (*Glis glis*) – objekt raziskave (Slika 17.5). Preglednico smo oblikovali v obliki pravokotne matrike dimenzije 299×2 z objekti (posamezna žival) v vrsticah in deskriptorji (telesna dolžina L in telesna masa M) v stolpcih. Vrednosti opisne statistike za obe spremenljivki imajo dimenzije, L : $171,7 \pm 12,0$ mm, 130 – 200 mm in M : $137,5 \pm 50,2$ g, 22 – 300 g.

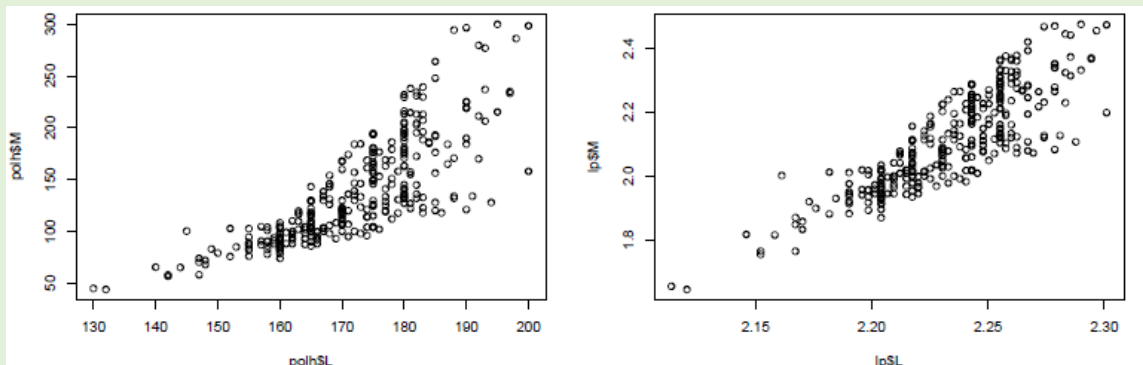
Zastavili smo raziskovalno vprašanje, ali sta dolžina telesa in masa povezani? Hipotezi: H_0 : med spremenljivkama ni povezave ($b_1 = 0$). H_a : spremenljivki sta povezani ($b_1 \neq 0$). Stopnja značilnosti $\alpha = 5\%$. Odnos med dolžino in maso polhov preverimo z enostavno linearno regresijsko analizo.

Prva odločitev je povezana z izbiro regresijskega modela. Ker obe spremenljivki izkazujeta napake in imata porazdelitev vrednosti blizu normalne, je ustrezna metoda enostavna linearna regresijska analiza ELRA Model II. Naslednja odločitev je izbira neodvisne spremenljivke oz. grafično, katera bo na abscisi (na X osi). Za razlagalno spremenljivko izberemo telesno dolžino, masa bo odvisna spremenljivka, nanašamo jo na Y os. Naslednji korak je razsevni diagram med obema spremenljivkama, zaradi preverjanja odnosa med spremenljivkama in iskanja osamelcev (Slika 17.6). Odnos ni linearen, ampak potenčen, zato vrednosti transformiramo z logaritmiranjem z osnovo 10. Postopek se imenuje popolna logaritemska transformacija. Razsevni diagram logaritmiranih vrednosti vrne bolj linearen odnos (Slika 17.4). Na diagramu ni osamelcev. Korelacijski koeficient ima vrednost $r = 0,85$ in vzorčno značilnost $p < 0,0001$ ter determinacijski koeficient $R^2 = 0,719$. To pomeni, da je povezanost med dolžino in maso telesa statistično značilna in močna, z dolžino razložimo 71,9 % variabilnosti v masi polhov.



Slika 17.5 Polh *Glis glis*.

Foto Alenka Kryštufek.



Slika 17.6 Projekcija telesne mase M Y os na telesno dolžino L X os. Levo – surove vrednosti. Desno – logaritmirane vrednosti.

```
R > package = »dplyr«  
lp<-log10(polh) # Logaritmiranje
```

```
R >  
#Bi-plot, na surovih, na LOG  
plot(polh$M~polh$L)  
plot(log10(polh$M)~log10(polh$L))
```





PRIMER 17.3

Razmerje med dolžino in maso pri polhu Model II – analiza

Obe spremenljivki sta blizu normalne porazdelitve, ne vsebujeta osamelcev in sta v korelaciji. Na podlagi navedenih lastnosti porazdelitve vrednosti izvedemo dve analizi, najprej po metodi navadnih najmanjših kvadratov OLS na surovih vrednostih in nato po metodi glavne osi MA na logaritmiranih vrednostih.

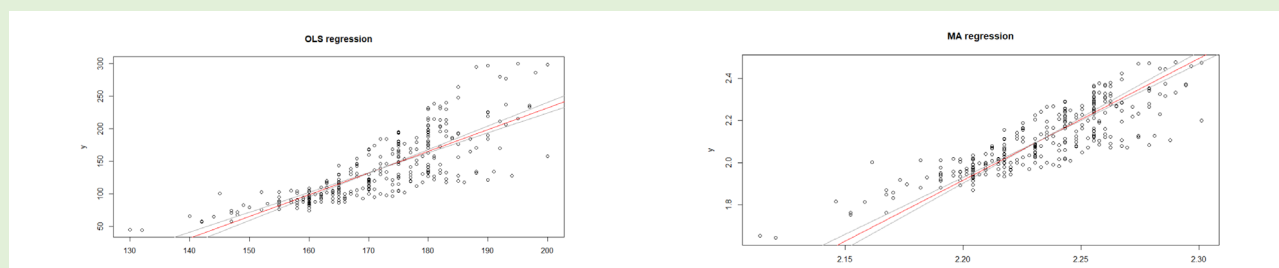
Smerni koeficient po metodi OLS je pozitiven, ima vrednost $b_1 = 3,34$ in je statistično značilen $p < 0,05$, kar vodi v sprejem H_a (Preglednica 17.2). To pomeni, da se telesna masa polha v povprečju poveča za 3,34 g, ko se mu dolžina poveča za 1 mm. Interval 95 % zaupanja smernega koeficienta je v razponu od 3,1 do 3,6 (Preglednica 17.2).

Smerni koeficient po metodi MA je pozitiven, ima vrednost $b_1 = 5,78$ in je statistično značilen $p < 0,05$, kar vodi v sprejem H_a (Preglednica 17.2). To pomeni, da se telesna masa polha v povprečju poveča za 5,8 %, ko se mu dolžina poveča za 1 %. Interval 95 % zaupanja smernega koeficienta je v razponu od 5,4 do 6,2 (Preglednica 17.2).

Spomnimo, da so vrednosti logaritmirane. To upoštevamo pri oblikovanju regresijske enačbe, ki se v skladu z modelom MA glasi: $\log M' = -10,81 + 5,78 * \log L'$. Ker sta obe spremenljivki logaritmirani, se model imenuje log-log oz. log-linearni regresijski model. Razsevna diagrama med vrednostmi mase in dolžine telesa z vrisanima regresijskima premicama in intervalom zaupanja so predstavljeni na Sliki 17.7. Masa in dolžina telesa nista v izometričnem temveč v alometričnem razmerju, masa raste hitreje kot dolžina. Zato bi ob zamenjavi mase in dolžine na abscisni osi (po metodi OLS) opisali regresijski premici kot $3,3^0$.

Preglednica 17.2 Rezultati enostavne linearne regresijske analize Modela II, dobljeni po dveh metodah

| Metoda | Konstantni člen, b_0 | Smerni koeficient, b_1 | P | 2,5-97,5% Konstantni člen | 2,5-97,5% Smerni koeficient |
|--------|------------------------|--------------------------|------|---------------------------|-----------------------------|
| OLS | -435,14 | 3,34 | 0,01 | -484,86 – -385,43 | 3,05 – 3,63 |
| MA | -10,81 | 5,78 | 0,01 | -11,80 – -9,95 | 5,40 – 6,23 |



Slika 17.7 Razsevni diagram vrednosti mase na Y osi in dolžine telesa na X osi z regresijsko premico (rdeča črta) in 95 % intervalom zaupanja (sivi krivulji) po metodah: levo – OLS na surovih vrednostih, desno – MA na logaritmiranih vrednostih.

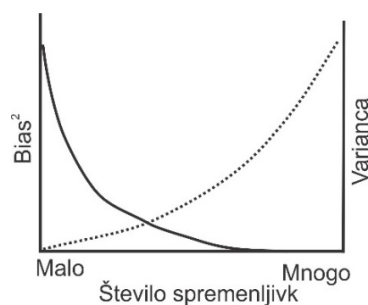
```
R > package = »lmodel2«
res<- lmodel2(M ~ L, data=polh, "relative", "relative", 99) # surove vrednosti
res
plot(res, "OLS")
lp.res<- lmodel2(M ~ L, data=lp, "relative", "relative", 99) # logaritmirane vrednosti
lp.res
plot(lp.res, "MA")
```



17.2 Multipla linearna regresija

V biodiverzitetnih raziskavah želimo variabilnost ene odvisne spremenljivke Y razložiti z več neodvisnimi spremenljivkami X_1, X_2, \dots, X_m . Potencialno vse razlagalne spremenljivke vplivajo na vrednost Y . Raziskovalni izziv je vrednotenje prispevka posamezne neodvisne spremenljivke in vrednotenje skupnega vpliva več razlagalnih spremenljivk na odvisno spremenljivko.

Multipla linearna regresija razumemo kot gradnjo modela iz neodvisnih spremenljivk. Pri čemer iščemo kompromis med maksimalnim prileganjem podatkov k regresijski premici s čim manjšim številom spremenljivk. Z dodajanjem razlagalnih spremenljivk v model se večja celokupna varianca modela, obenem se zmanjšuje pristranost modela (bias) oz. večja se natančnost ocene vrednosti odvisne spremenljivke. Obraten proces se dogaja pri manjšanju števila razlagalnih spremenljivk, manjša se varianca modela in večja se pristranost ocene za odvisno spremenljivko. Z drugimi besedami, iščemo model z maksimalnim deležem razložene variance ob čim bolj preprosti zgradbi. Za enostavnejši model z manj spremenljivkami pravimo, da je bolj varčen. Odnos med pristranostjo in varianco je tak, da se z naraščanjem variance hitro zmanjšuje bias (Slika 17.8). Naloga raziskovalca (analitika) je, da išče kompromisno rešitev med obema ekstremoma.



Slika 17.8 Princip varčnosti: koncept odnosa med pristranostjo ocene odvisne spremenljivke (bias na kvadrat, Bias^2) ter varianco glede na število spremenljivk v modelu. Izbira optimalnega modela je kompromis med pristranostjo ocene in varianco. Za optimalni model ni nujno, da leži v točki presečišča obeh krivulj, navadno je optimalni model nekoliko desno od presečišča. Vzeto iz Burnham in Anderson (2002).

Multipla linearna regresijska analiza je optimiranje modela v obliki regresijske enačbe:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_mX_m + \varepsilon.$$

Za vse neodvisne spremenljivke velja domneva, da so brez napak, zato vse ostanke računamo vzdolž Y osi.

Regresijski koeficienti b imajo naslednji pomen: **konstantni člen** b_0 , predstavlja pričakovano vrednost odvisne spremenljivke v primeru, če bi bile vrednosti vseh neodvisnih spremenljivk enake 0. **Regresijski koeficienti** b_n predstavljajo pričakovano spremembo vrednosti odvisne spremenljivke, ko se X_i spremeni za eno enoto, vse ostale spremenljivke ostanejo nespremenjene. Izračunamo lahko dva regresijska koeficienta: 1. **navadni regresijski koeficient** b (ang. Ordinary Regression Coefficient) izračunamo na surovih podatkih, vrednost koeficienta je v fizikalnih enotah neodvisne spremenljivke. Ta oblika koeficienta je uporabna za izračun ocenjenih vrednosti Y s pomočjo regresijske enačbe. 2. **Standardizirani regresijski koeficient** b' (ang. Standard Regression Coefficient) izračunamo na standardiziranih podatkih, vrednost koeficienta nima dimenzije. Ta oblika koeficienta je uporabna za primerjanje relativnega prispevka posamezne neodvisne spremenljivke. Višji koeficient b'_i pomeni večji prispevek spremenljivke X_i k tvorbi vrednosti Y . Obe obliki regresijskih koeficientov imata značaj **delnega regresijskega koeficienta** (ang. Partial Regression Coefficient). Termin »delni« se nanaša na velikost spremembe odvisne spremenljivke, ko se neodvisna spremeni za eno enoto. Velja pod pogojem, da ostanejo regresijski koeficienti vseh preostalih neodvisnih spremenljivk v modelu nespremenjeni.

Poseben problem v multipli regresiji predstavlja vključevanje povezanih razlagalnih spremenljivk v analizo. Če sta neodvisni spremenljivki povezani torej korelirani, govorimo o **kolinearnosti** (ang. Collinearity). V primeru več povezanih spremenljivk ali funkcijske povezanosti govorimo o **multikolinearnosti** (ang. Multicollinearity). Kolinearne spremenljivke v modelu zelo vplivajo na izračun regresijskih koeficientov. Rezultati analize so lahko zavajajoči in nerealni. Pred izvedbo multiple regresije neodvisne (razlagalne) spremenljivke obvezno preverimo glede kolinearnosti, ki jo nato odpravimo. Pri eksperimentalnem delu se lahko z načrtovanjem eksperimenta in premišljeno izbiro kontroliranih spremenljivk izognemo kolinearnosti. Pri zbiranju meritev terenskih spremenljivk oz. opazovanj v naravi se pogosto srečamo z močno korelacijo med spremenljivkami. Zato moramo biti zelo pazljivi pri vključevanju kolinearnih spremenljivk v model regresijske enačbe. Prvo informacijo o tveganjih na kolinearnost dobimo s preverjanjem neodvisnih spremenljivk in izračunom korelacijskih koeficientov. Močne ali celo funkcionalne povezave so dober kazalnik kolinearnosti. V začetni stopnji iskanja optimalnega modela vključimo v analizo vse korelirane spremenljivke. Odločitev o izločitvi kolinearne spremenljivke sprejmemo na podlagi višine **koeficienta inflacije**

variance VIF (ang. Variance Inflation Faktor). Vrednost $VIF = 1$ ima tista neodvisna spremenljivka, ki ni v korelaciji z nobeno drugo spremenljivko ($r = 0$). Spremenljivke z najvišjimi vrednostmi VIF najbolj zmalčijo model in vplivajo na izračun regresijskih koeficientov, zato jih najprej izločimo iz modela. Ponovimo izračun modela, brez spremenljivke z visokim VIF in ponovno preverimo VIF pri vključenih spremenljivkah. Absolutna višina VIF za izločitev spremenljivke ni definirana. Nekateri raziskovalci zagovarjajo izločitev spremenljivk pri vrednosti $VIF > 5$ ali pri $VIF > 10$, nekateri celo pri $VIF > 20$. Temelj za vključitev spremenljivke v model naj bo načelo varčnosti.

Za oblikovanje modela multiple linearne regresije in izbor razlagalnih spremenljivk obstajajo štiri strategije: 1. izločanje zadnjih, 2. vključevanje prvih, 3. stopenjski postopek in 4. izbor najboljših modelov. Prve tri strategije izgradnje optimalnega modela izhajajo iz testiranja ničelne hipoteze. Pri izgradnji regresijskega modela s pomočjo testiranja ničelne hipoteze preverjamo ali prispevek posamezne spremenljivke statistično značilno vpliva na model. In spremenljivke s statistično značilnim vplivom so vključene v model. Spremenljivke brez značilnega vpliva niso vključene v končni model. Na končni izbor spremenljivk vplivata še cenilki: koeficient inflacije variance VIF in determinacijski koeficient R^2 . Četrta strategija ima popolnoma drugačno izhodišče in se opira na informacijsko – teoretični pristop ITP. V njej oblikujemo vse možne modele, glede na cenilke ITP (AIC, AIC_c, BIC, C_p) naredimo izbor nekaj optimalnih modelov.

V **postopku izločanja zadnjih** (ang. Backward Elimination Procedure) v model vključimo celoten nabor neodvisnih spremenljivk. Izračunamo regresijske koeficiente in spremenljivko, ki ne prispeva značilno k njihovem oblikovanju, izločimo iz naslednjega modela. Postopek ponavljamo, dokler ne oblikujemo modela, v katerem vse neodvisne spremenljivke statistično značilno prispevajo k izračunu modela.

Postopek vključevanja prvih (ang. Forward Selection Procedure) temelji na postopnem vključevanju neodvisnih spremenljivk v model. Najprej vključimo v model spremenljivko, ki vrne najvišji determinacijski koeficient R^2 . V naslednjem koraku dodamo spremenljivko, ki najbolj zviša vrednost obstoječega R^2 . Postopek ponavljamo, dokler se ob dodajanju spremenljivk R^2 značilno viša. Spremenljivke, ki ne izkažejo značilnega prispevka k izračunu regresijskih koeficientov, ne vključimo v model.

Stopenjski postopek (ang. Stepwise Procedure) je kombinacija izločanja in vključevanja spremenljivk. Na vsaki stopnji preverimo: signifikantni prispevek k regresijskim koeficientom in prispevek k višanju R^2 .

Metoda **izbora najboljših modelov** (ang. Best Subset Selection). V informacijsko – teoretičnem pristopu oblikujemo nabor kandidatnih modelov. Izbor modela in s tem ustreznega ravnovesja med biasom in varianco izvedemo s pomočjo prilagojenega deleža

razložene variance R_a^2 in z informacijskimi cenilkami AIC, AIC_C in BIC. Kandidatni modeli so vsi regresijski modeli, ki jih lahko oblikujemo iz nabora neodvisnih spremenljivk p . Število kandidatnih modelov D izračunamo s formulo:

$$D = 2^p - 1.$$

Ob večjem številu razlagalnih spremenljivk (> 6) število kandidatnih modelov preraste obvladljivo število kombinacij. Zato z logičnim sklepanjem oblikujemo nabor relevantnih kandidatnih modelov. V raziskovalnem pristopu (ang. Exploratory Research) vključimo najrazličnejše vidike povezav in iščemo različne odnose med spremenljivkami, zato praviloma zgradimo vse (ali čim več) kandidatnih modelov in z njimi testiramo hipoteze. V razlagalnem pristopu (ang. Explanatory Research) smo bolj fokusirani na vzročno posledične odnose in na podlagi predhodnega znanja gradimo regresijske modele.

Linearni model s štirimi neodvisnimi spremenljivkami $p = 4$, $x_1 \dots x_4$ omogoča oblikovanje 15 kombinacij kandidatnih modelov ($2^4 - 1 = 15$), za razlago odvisne spremenljivke Y :

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \varepsilon \\ y &= \beta_0 + \beta_2 x_2 + \varepsilon \\ y &= \beta_0 + \beta_3 x_3 + \varepsilon \\ y &= \beta_0 + \beta_4 x_4 + \varepsilon \\ y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \\ y &= \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \varepsilon \\ y &= \beta_0 + \beta_1 x_1 + \beta_4 x_4 + \varepsilon \\ y &= \beta_0 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \\ y &= \beta_0 + \beta_2 x_2 + \beta_4 x_4 + \varepsilon \\ y &= \beta_0 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon \\ y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \\ y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_4 x_4 + \varepsilon \\ y &= \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon \\ y &= \beta_0 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon \\ y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon. \end{aligned}$$

Vsem kandidatnim modelom izračunamo prilagojeni determinacijski koeficient R_a^2 in cenilke informacijsko – teoretičnega pristopa: Akaikov informacijski kriterij AIC ali prilagojeni Akaikov informacijski kriterij AIC_C, izbira med cenilkama je odvisna od velikosti vzorca. Izmed kandidatnih modelov izberemo nekaj modelov, ki najbolj razlagajo variabilnost odvisne spremenljivke. O izboru modelov presodimo na podlagi vrednosti cenilk: da imajo najvišjo vrednost R_a^2 , da imajo najnižje vrednosti informacijskih cenilk AIC oz. AIC_C. V informacijsko – teoretičnem pristopu ne preverjamo statistično značilnega prispevka posamezne spremenljivke v modelu. Ker je izbor najboljših modelov bil narejen iz množice modelov, kot rešitev predstavimo nekaj najkonkurenčnejših

modelov. To tudi pomeni, da ni samo ene možne rešitve temveč lahko zadovoljivo rešitev ponudi več regresijskih modelov.

Načelo **varčnosti** (ang. Parsimony). Vsak člen v enačbi multiple regresije, ki ne prispeva značilno k modelu, izločimo. Kriterij za izločitev spremenljivke je statistična nesignifikantnost ($p > 0,05$).

Multipli determinacijski koeficient R^2 (ang. Coefficient of Multiple Determination) je merilo deleža razložene variabilnosti odvisne spremenljivke z modelom neodvisnih spremenljivk. Število neodvisnih spremenljivk vpliva na uspešnost modela. Model je uspešnejši ob manjšem številu razlagalnih spremenljivk. Determinacijski koeficient R^2 ima nabor vrednosti od 0 do 1. Vrednost R^2 po navadi preoblikujemo v delež (%), s katerimi podamo delež razložene variabilnosti odvisne spremenljivke.

Na podlagi R^2 in ocene vpliva števila spremenljivk v modelu ocenimo uspešnost modela multiple linearne regresije s tremi statistikami: prilagojeni R^2_a , VIF in F -statistiko.

Prilagojeni multipli determinacijski koeficient R^2_a (ang. Adjusted Coefficient of Multiple Determination or Adjusted R^2_a) vrne nepristranski izračun pričakovane razložene variance v modelu. V izračun so vključene prostostne stopinje, ki jih računamo iz razmerja med številom spremenljivk in številom meritev. Prilagojeni R^2_a ima nabor vrednosti od 0 do 1 in ga prav tako pretvorimo v %. R^2_a ima nižjo vrednost od R^2 ali maksimalno enako.

Z **F -testom** (ang. F -statistic) preverjamo statistično značilnost determinacijskega koeficienta R^2 na podlagi p vrednosti. Vrednost $R^2 = 0$ pomeni, da ne bomo zavrnil H_0 , pri vrednosti $R^2 > 0$ sprejmemo H_a .

Akaikov informacijski kriterij AIC (ang. Akaike Information Criterion) je cenilka prileganja (ang. Goodnes of Fit) statističnega modela k podatkom in ga računamo iz ostankov ε (ang. Residuals). Ostenek je razdalja med dejansko vrednostjo Y in njeno izračunano vrednostjo \hat{Y} po regresijskem modelu. Izračun temelji na povprečju kvadratov ostankov RMS (ang. Residual Mean Square) oz. varianci ostankov σ^2 , ki jo računamo iz vsote kvadratov ostankov SSE (ang. Sum of Square Error) s formulami:

$$\begin{aligned}\varepsilon &= Y - \hat{Y} \\ SSE &= \sum \varepsilon^2 \\ RMS &= \sigma^2 = \frac{SSE}{n}.\end{aligned}$$

V izračun AIC vključimo število meritev n , število spremenljivk K in varianco ostankov σ^2 in ga izračunamo s formulo (obstajajo tri oblike formule):

$$AIC = n * \ln(\sigma^2) + 2K .$$

Vrednost AIC vrednotimo relativno glede na vrednosti drugih modelov. Z AIC primerjamo različne modele med seboj, model z nižjo vrednostjo AIC je uspešnejši. AIC ne daje informacije ali je model ustrezen ali ne, temveč samo rangira modele, ki smo jih vključili v analizo.

Kadar je število meritev majhno to pomeni, da je razmerje med številom meritev n in številom spremenljivk K manjše od 40:

$$\frac{n}{K} < 40 ,$$

govorimo o majhnem vzorcu. V takih primerih prilaganje modela k podatkom vrednotimo s cenilko **prilagojeni Akaikevi informacijski kriterij AIC_C** (ang. Adjusted Akaike Information Criterion). Dejansko gre za korekcijo AIC, ki jo računamo s formulo:

$$AIC_C = AIC + \frac{2K(K+1)}{(n-K-1)} .$$

Z večanjem števila meritev se vrednost AIC_C približuje AIC.

Pri ocenjevanju najboljših modelov je pomembna razlika med najnižjo vrednostjo AIC_C in naslednjo AIC_C oz. med sosednjimi modeli. Razliko v vrednosti AIC_C med sosednjima modeloma označimo z **delta** Δ (sin. ΔAIC_C) in manjša kot je razlika uspešnejši je model.

Za vsako vrednost AIC_C podamo tudi **težo** w_i (ang. Weight). Teža w_i ima vrednosti od 0 do 1 in podaja vrednost, kako uspešen je model i med obravnavanimi modeli. Težo izračunamo s formulo:

$$w_i = \frac{\exp(-1/2 \Delta_i)}{\sum \exp(-1/2 \Delta)} .$$

Bayesov informacijski kriterij BIC (ang. Bayes Information Criterion) rangira modele po prilaganju k podatkom. Postopek izračuna je zelo podoben AIC s tem, da še dodatno kaznuje kompleksnost modela. BIC računamo po formuli:

$$BIC = n * \ln(\sigma^2) + K * \ln n .$$

Opomba: V literaturi in pri praktičnem računanju se uporabljajo tri različice izračuna AIC_C in BIC. Vse tri različice dajejo zelo podobne vrednosti cenilk, ki se malo razlikujejo med seboj.

AIC

Hiroto Akaike (1927-2009) je bil japonski statistik. Njegov pomemben prispevek k statističnemu vrednotenju je bila formulacija kriterija za statistično sklepanje. V biodiverzitetnih raziskavah njegovo cenilko najpogosteje uporabljamo za izbor regresijskih modelov. Cenilka je poimenovana po njem: Akaikeov informacijski kriterij AIC.



PRIMER 17.4

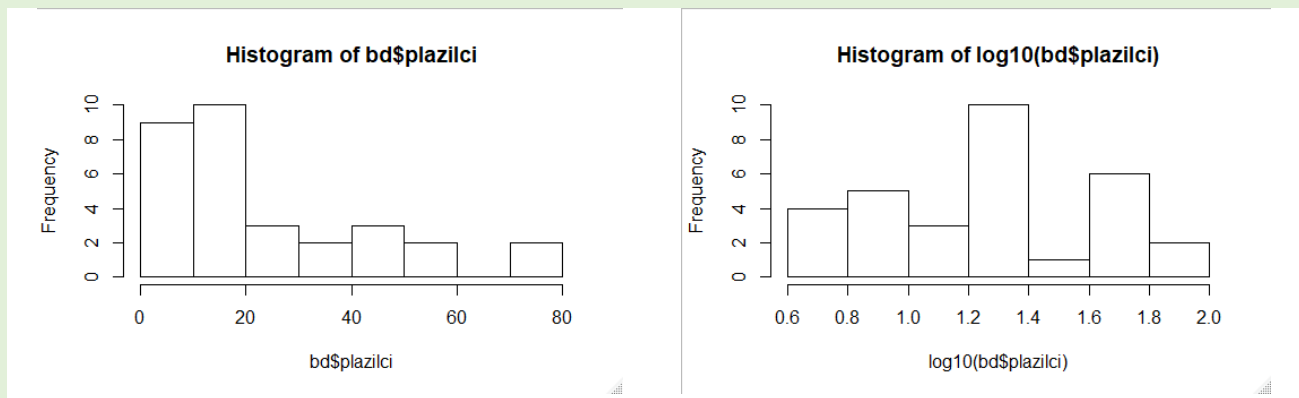
Vrstno bogastvo in multipla linearna regresija – podatki

Za 31 evropskih držav smo zbrali podatke o vrstnem bogastvu plazilcev (Slika 11.8) – odvisna spremenljivka. Bogastvo plazilcev smo želeli razložiti z naborom petih neodvisnih spremenljivk: velikost države – površina v km² (a); geografski položaj države – koordinati geografske širine (lat) in dolžine (lon) glavnega mesta; bioklimatske razmere – povprečna letna temperatura °C (bio1) in povprečne letne padavine l/m² (bio12).

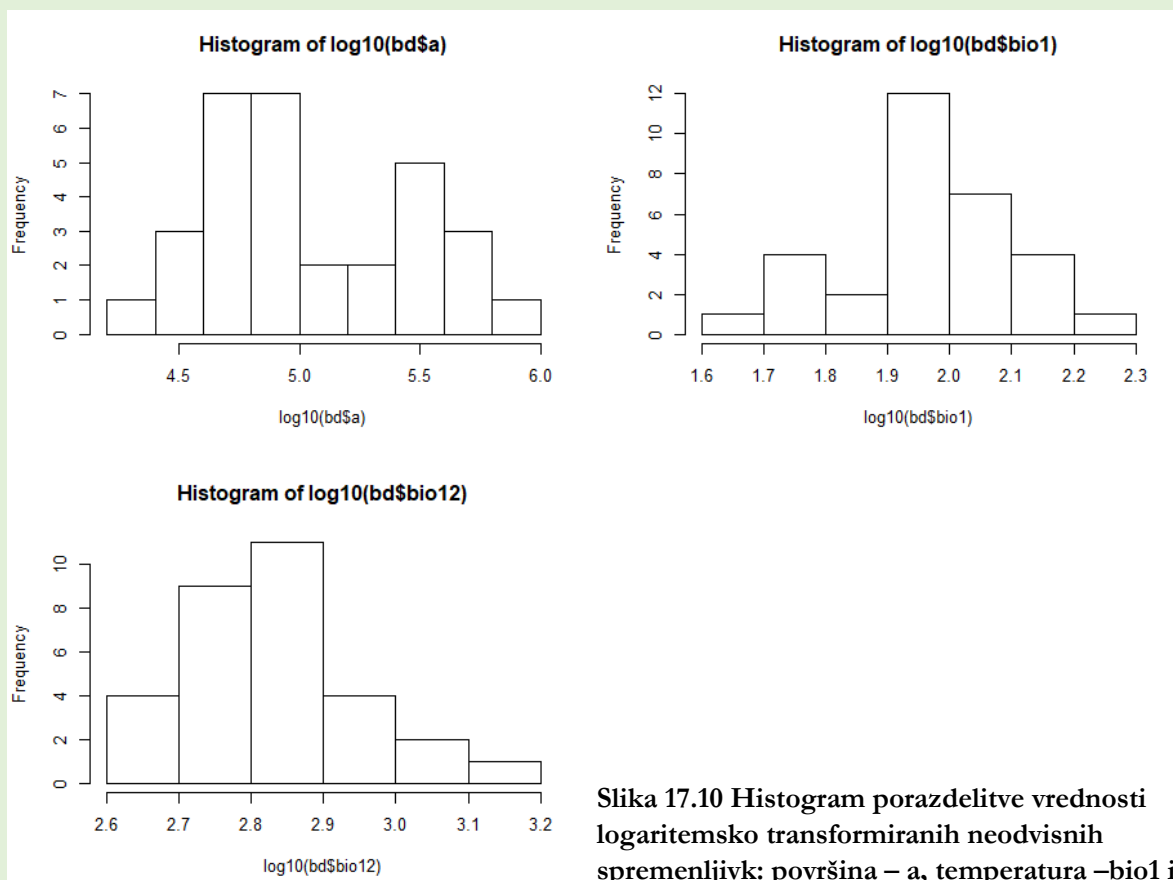
Preverjamo hipotezo, da vrstno bogastvo lahko razložimo s kombinacijo razmerij med površino ozemlja, položajem države v prostoru in bioklimatskimi značilnostmi.

Pri postavitvi multiplega regresijskega modela upoštevamo: 1. spremenljivke morajo izpolnjevati pogoje (normalnost, homoskedastičnost in nekolinearnost); 2. slediti moramo načelu varčnosti. Analizo bomo naredili posebej po informacijsko – teoretičnem pristopu ITP in posebej s testiranjem ničelne hipoteze. Model gradimo bodisi z vključevanjem ali izločanjem spremenljivk, bodisi stopenjsko. Pogoj za signifikantni prispevek preverjamo pred, med in po izvedbi modela. Model iteriramo (ponavljamo), dokler ne dobimo zadovoljujoče rešitve.

Preverjanje normalnosti porazdelitve spremenljivk vrne histograme z desno asimetrijo (Slika 17.9). Za korekcijo asimetrične porazdelitve vrednosti logaritemsko transformiramo (Slika 17.10) in jih približamo normalnejši porazdelitvi (transformirani nista spremenljivki lat in lon).



Slika 17.9 Histogram porazdelitve vrednosti vrstnega bogastva plazilcev. Levo – surove vrednosti, desno – logaritmsko transformirane vrednosti.



Slika 17.10 Histogram porazdelitve vrednosti logaritmsko transformiranih neodvisnih spremenljivk: površina – a, temperatura –bio1 in

```
R >
hist(bd$lat)
hist(bd$lon)
hist(bd$a)
hist(bd$bio1)
hist(bd$bio12)
hist(log10(bd$a))
hist(log10(bd$bio1))
hist(log10(bd$bio12))
```

```
R >
head(bd) #pregled podatkov
summary(bd)
op<-par(mfrow=c(1,2)) #Št. slik na zaslon
#Histogrami, surovi, log10
hist(bd$plazilci)
hist(log10(bd$plazilci))
par(op) #Preklic št. slik na zaslon
```





PRIMER 17.5

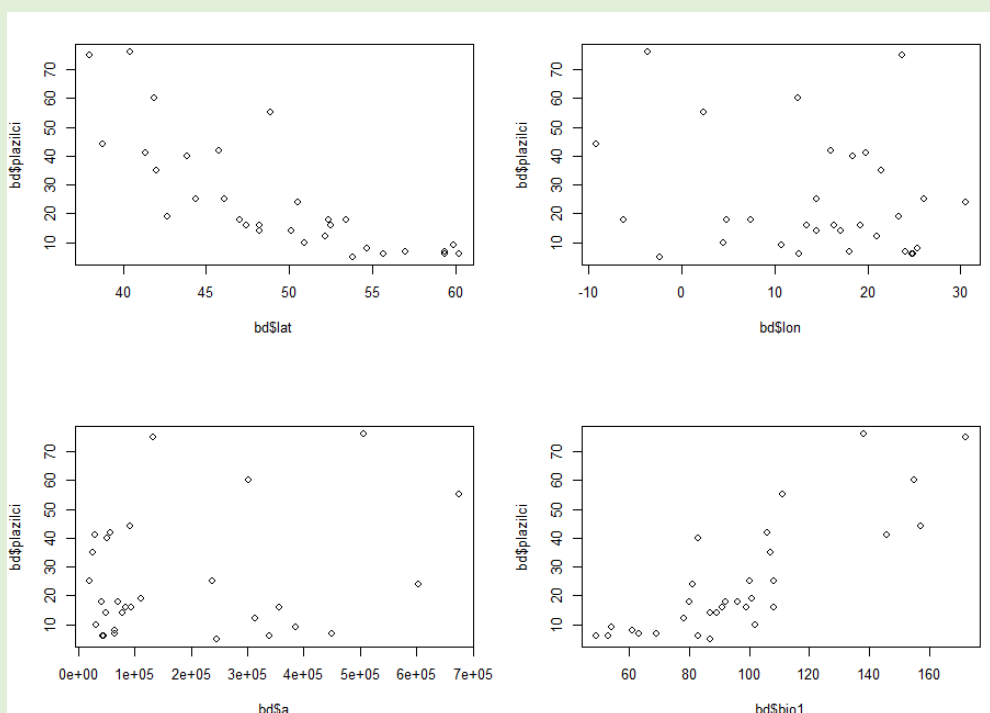
Vrstno bogastvo in multipla linearna regresija – normalnost podatkov

Preverjanje kolinearnosti – med neodvisnimi spremenljivkami izvedemo korelacijsko analizo (Preglednica 17.3). Močno negativno povezanost izkazuje le par geografska širina (lat) in povprečna letna temperatura (bio1). Z drugimi besedami, ko se premikamo iz Mediterana proti Skandinaviji, povprečna letna temperatura upada. Zaradi kolinearnosti med lat in bio1 bo v končni varianti modela ostala samo ena od teh dveh spremenljivk.

Preglednica 17.3 Korelacijski koeficienti med neodvisnimi spremenljivkami, močna korelacija ($|r| > 0,5$) je v krepkem tisku (razloga oznak za spremenljivke je v besedilu)

| | lat | lon | a | bio1 |
|-------|--------------|-------|-------|------|
| lon | 0,13 | | | |
| a | 0,16 | -0,06 | | |
| bio1 | -0,87 | -0,31 | -0,03 | |
| bio12 | -0,07 | -0,28 | -0,34 | 0,01 |

Preverjanje homoskedastičnosti izvedemo z raztresenim diagramom odvisne spremenljivke na niz neodvisnih spremenljivk. Heteroskedastična zveza pomeni, da se v enem segmentu grafa odvisna spremenljivka linearno odziva na spremembo neodvisne spremenljivke. V drugem segmentu grafa se odziv odvisne spremenljivke razcepi, del podatkov se odziva močnejše in del podatkov šibkeje; na grafu je razporeditev točk v obliki ležeče črke Y. Takega primera nimamo (Slika 17.11).



Slika 17.11 Projekcija odvisne spremenljivke (vrstno bogastvo plazilcev) na niz neodvisnih spremenljivk (geo. dolžina in širina, površina in povprečna letna temperatura).

```
R >
aktivni<-bd[,5:9] # Izbor neodvisnih spremenljivk
cor(aktivni) # Korelacija, preverjanje kolinearnost
```

```
R >
#Bi-plot - homoskedastičnost
op<-par(mfrow=c(2,2))
plot(bd$plazilci ~ bd$lat)
plot(bd$plazilci ~ bd$lon)
plot(bd$plazilci ~ bd$a)
plot(bd$plazilci ~ bd$bio1)
plot(bd$plazilci ~ bd$bio12)
par(op)
```



PRIMER 17.6

Vrstno bogastvo in multipla linearna regresija –
pristop s testiranjem ničelne hipoteze s stopenjskim postopkom

Preverjamo možnost razlage variabilnosti vrstnega bogastva plazilcev v prostoru z geografskimi in bioklimatskimi dejavniki. V tem primeru bomo prikazali pristop s testiranjem ničelne hipoteze. To pomeni, da preverjamo signifikantni prispevek posamezne spremenljivke k izgradnji modela. Model bomo gradili s stopenjskim postopkom, z dodajanjem in odvzemanjem spremenljivk in sprotim preverjanjem signifikantnosti. V začetni model multiple linearne regresije vključimo vseh pet neodvisnih spremenljivk; regresijska enačba ima obliko:

$$\log(\text{plazilci}) \sim \text{lat} + \text{lon} + \log(a) + \log(\text{bio1}) + \log(\text{bio12}).$$

Začetni model, s katerim variabilnost v številu vrst plazilcev razlagamo s petimi neodvisnimi spremenljivkami, razloži visoki delež variabilnosti $R^2 = 0,810$ oz. 81 %; prav tako visok je prilagojeni $R^2_a = 0,772$. Determinacijski koeficient je statistično značilen $F_{5, 25} = 21,35$, $p < 0,0001$. Pregled posameznih regresijskih koeficientov vrne samo dve signifikantni spremenljivki (lat in a) (Preglednica 17.4). Preverjanje koeficienta inflacije variance VIF vrne dve spremenljivki (lat in bio1) s previsokim faktorjem (VIF > 5) (Preglednica 17.4).

Model kot celota je signifikanten (značilen) in razloži visok delež variabilnosti vrstnega bogastva plazilcev v evropskih državah. Vendar model krši pravilo varčnosti z vključevanjem spremenljivk, ki ne vplivajo značilno (lon, bio1, bio12), in z vključevanjem spremenljivk, ki kršita kolinearnost.

Preglednica 17.4 Regresijski rezultati začetnega modela, signifikantne p vrednosti in visoki FIV koeficienti so v krepkem tisku

| | Koeficient | Std. napaka | t | p | VIF |
|-------------|------------|-------------|-------|---------------|--------------|
| (Intercept) | 1,86 | 2,09 | 0,89 | 0,3835 | |
| lat | -0,04 | 0,01 | -3,83 | 0,0008 | 6,026 |
| lon | 0,00 | 0,00 | -0,21 | 0,8345 | 1,570 |
| log(a) | 0,17 | 0,08 | 2,14 | 0,0420 | 1,355 |
| log(bio1) | 0,25 | 0,57 | 0,43 | 0,6686 | 6,660 |
| log(bio12) | 0,09 | 0,30 | 0,30 | 0,7640 | 1,466 |

Optimalni multipli model, ki smo ga zgradili s stopenjskim postopkom ohrani samo dve spremenljivki *lat* in *a*, ki signifikantno prispevata k modelu. Postopek izvedemo z ukazom *stepAIC* v paketu MASS.

```
R > package = »MASS«
# Začetni model
mr_bd_plazilci<-lm(log10(plazilci)~lat+lon+log10(a)+log10(bio1)+log10(bio12), data = bd)
summary(mr_bd_plazilci) # Pregled regresijskih koeficientov
car::vif(mr_bd_plazilci) # Pregled VIF
```

```
R > package = »MASS«
#Stopenjski postopek
step_plazilci<-stepAIC(mr_bd_plazilci, direction = "both")
```

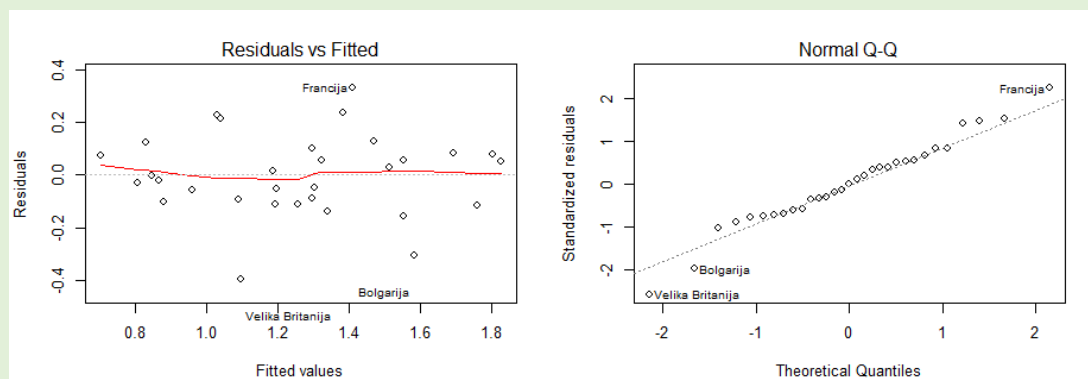
Končni model sestavljata dve neodvisni spremenljivki, geografska širina (*lat*) in površina (*log(a)*): $\log(\text{plazilci}) \sim \text{lat} + \log(a)$, ki razložita 81 % variabilnosti vrstnega bogastva plazilcev v Evropi, $R^2 = 0,806$, prilagojeni $R^2_a = 0,792$; model je visoko statistično značilen $F_{2,28} = 57,97$ $p < 0,0001$. Vsi trije regresijski koeficienti so značilni ($p < 0,05$), vrednosti VIF pa so nizke < 5 (Preglednica 17.5).

Rezultate multiple linearne analize strnemo v ugotovitev, da vrstno bogastvo plazilcev v Evropi dobro razložita geografska širina države in njena površina. Z velikostjo države narašča bogastvo plazilcev s stopnjo $b = 0,16$, od juga proti severu pa bogastvo upada s stopnjo $b = -0,05$. Ne smemo prezreti povprečne temperature, ki deluje v nasprotni smeri od geografske širine in je bila zaradi kolinearnosti izločena iz modela. Njena informacija je deloma ohranjena v geografski širini.

Preglednica 17.5 Regresijski rezultat končnega modela

| | Koeficient | Std. napaka | t | p | VIF |
|-------------|------------|-------------|--------|--------------------|--------------|
| (Intercept) | 2,8462 | 0,3644 | 7,81 | < 0,0001 | |
| lat | -0,0489 | 0,0046 | -10,74 | < 0,0001 | 1,029 |
| log(a) | 0,1632 | 0,0648 | 2,52 | 0,0178 | 1,029 |

Preverimo še porazdelitev ostankov, ki ne smejo izkazovati trenda in morajo biti normalno porazdeljeni, oba kriterija sta izpolnjena (slika 17.12).



Slika 17.12. Projekcija ostankov na funkcijo prileganja – levo, in Q–Q graf – desno.

R > package = »MASS«

#Končni model

```
mr_plazilci<-lm(log10(plazilci)~ lat + log10(a), data=bd)
```

```
summary(mr_plazilci)
```

```
car::vif(mr_plazilci)
```

```
anova(mr_plazilci)
```

R > package = »MASS«

#Kontrola ostankov

```
plot(mr_plazilci)
```



PRIMER 17.7

Vrstno bogastvo in multipla linearna regresija –
Informacijsko – teoretični pristop

Preverjamo možnost razlage variabilnosti vrstnega bogastva plazilcev v prostoru s petimi geografskimi in bioklimatskimi spremenljivkami (predstavljene so v Primeru 17.4). Tukaj bomo prikazali informacijsko – teoretični pristop. To pomeni, da zgradimo vse možne kombinacije modelov, število modelov je 31. Pet modelov je s samo eno razlagalno spremenljivko:

```
m1<-lm(log10(plazilci)~lat, data = bd) #1
m2<-lm(log10(plazilci)~lon, data = bd) #1
m3<-lm(log10(plazilci)~log10(a), data = bd) #1
m4<-lm(log10(plazilci)~log10(bio1), data = bd) #1
m5<-lm(log10(plazilci)~log10(bio12), data = bd) #1
```

Deset modelov z dvema razlagalnima spremenljivkama:

```
m6<-lm(log10(plazilci)~lat+lon, data = bd) #2
```

```

m7<-lm(log10(plazilci)~lat+log10(a), data = bd) #2
m8<-lm(log10(plazilci)~lat+log10(bio1), data = bd) #2
m9<-lm(log10(plazilci)~lat+log10(bio12), data = bd) #2
m10<-lm(log10(plazilci)~lon+log10(a), data = bd) #2
m11<-lm(log10(plazilci)~lon+log10(bio1), data = bd) #2
m12<-lm(log10(plazilci)~lon+log10(bio12), data = bd) #2
m13<-lm(log10(plazilci)~log10(a)+log10(bio1), data = bd) #2
m14<-lm(log10(plazilci)~log10(a)+log10(bio12), data = bd) #2
m15<-lm(log10(plazilci)~log10(bio1)+log10(bio12), data = bd) #2

```

Deset modelov s tremi razlagalnimi spremenljivkami:

```

m16<-lm(log10(plazilci)~lat+lon+log10(a), data = bd) #3
m17<-lm(log10(plazilci)~lat+lon+log10(bio1), data = bd) #3
m18<-lm(log10(plazilci)~lat+lon+log10(bio12), data = bd) #3
m19<-lm(log10(plazilci)~lat+log10(a)+log10(bio12), data = bd) #3
m20<-lm(log10(plazilci)~lat+log10(a)+log10(bio1), data = bd) #3
m21<-lm(log10(plazilci)~lat+log10(bio1)+log10(bio12), data = bd) #3
m22<-lm(log10(plazilci)~lon+log10(a)+log10(bio1), data = bd) #3
m23<-lm(log10(plazilci)~lon+log10(a)+log10(bio12), data = bd) #3
m24<-lm(log10(plazilci)~lon+log10(bio1)+log10(bio12), data = bd) #3
m25<-lm(log10(plazilci)~log10(a)+log10(bio1)+log10(bio12), data = bd) #3

```

Pet modelov s štirimi razlagalnimi spremenljivkami:

```

m26<-lm(log10(plazilci)~lat+lon+log10(a)+log10(bio1), data = bd) #4
m27<-lm(log10(plazilci)~lat+lon+log10(a)+log10(bio12), data = bd) #4
m28<-lm(log10(plazilci)~lat+lon+log10(bio1)+log10(bio12), data = bd) #4
m29<-lm(log10(plazilci)~lat+log10(a)+log10(bio1)+log10(bio12), data = bd) #4
m30<-lm(log10(plazilci)~lon+log10(a)+log10(bio1)+log10(bio12), data = bd) #4

```

In en model s vsemi petimi razlagalnimi spremenljivkami:

```

m31<-lm(log10(plazilci)~lat+lon+log10(a)+log10(bio1)+log10(bio12), data = bd) #5

```

```
summary(m7)
```

```
plot(m7)
```

S pomočjo paketa »MiMIIn« smo izračunali cenilke prilagojeni Akaikov informacijski kriterij AIC_C , delta ΔAIC_C ter težo $wAIC_C$ in modele razvrstili glede na najnižjo vrednost AIC_C (Preglednica 17.6). Najnižjo vrednost AIC_C je imel model 7, ki v razlago vrstnega

bogastva plazilcev vključuje samo dve razlagalni spremenljivki geografsko širino in površino države: $\log_{10}(\text{plazilci}) \sim \text{lat} + \log_{10}(a)$. Model sedem je imel težo 0,397, to pomeni zmerno dobro razlago variabilnosti vrstnega bogastva. Naslednji model po prileganju k podatkom je model 20, ta vključuje tri spremenljivke poleg *lat* in *a* še *bio1* – povprečno letno temperaturo. Vendar model 20 krši pravilo kolinearnosti saj sta geografska širina in letna temperatura v tesni korelaciji. Zato ta model ni primeren.

Preglednica 17.6 Prvi štirje modeli pri katerih je teža večja od 0,1 ($wAIC_c > 0,1$), ki se najbolj prilegajo podatkom

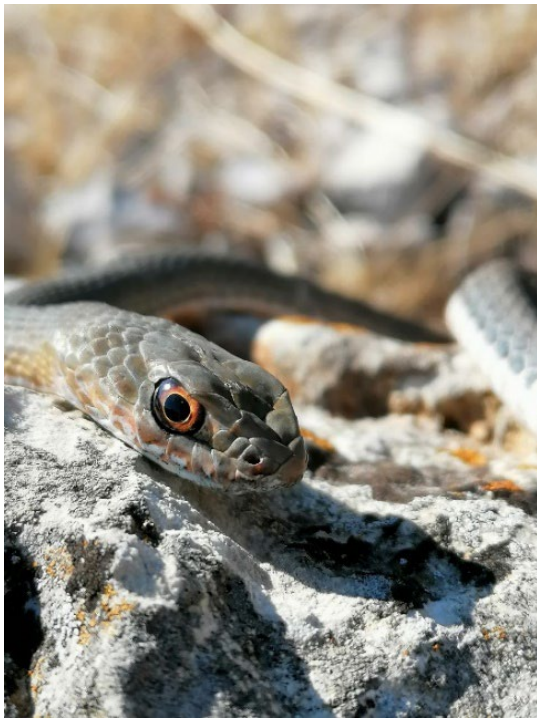
| Model | Stopinje prostosti Df | AICc | $\Delta AICc$ | $wAICc$ |
|-------|-----------------------|--------|---------------|---------|
| m7 | 4 | -19,74 | 0,000 | 0,397 |
| m20 | 5 | -17,38 | 2,358 | 0,122 |
| m16 | 5 | -17,35 | 2,389 | 0,120 |
| m19 | 5 | -17,09 | 2,643 | 0,106 |

Variabilnost vrstnega bogastva smo razložili s pomočjo testiranja ničelne hipoteze in z informacijsko – teoretičnim pristopom. V obeh pristopih smo prišli do enakega modela, ki vključuje samo dve razlagalni spremenljivki površino ozemlja in geografsko širino.

R > package = »MuMIn«

```
AICs <- AICc(m1, m2, m3, m4, m5, m6, m7, m8, m9, m10, m11, m12, m13, m14,
             m15, m16, m17, m18, m19, m20, m21, m22, m23, m24, m25, m26,
             m27, m28, m29, m30, m31) # Ppriprava AICc
MyDf <- AICs[,1]
AICsNum <- AICs[,2]
minAW <- min(AICsNum)
Delta <- AICsNum - minAW
RL <- exp(-0.5 * Delta)
wi <- round(RL / sum(RL), 3)
model <- c("m1", "m2", "m3", "m4", "m5", "m6", "m7", "m8", "m9", "m10", "m11", "m12", "m13",
           "m14", "m15", "m16", "m17", "m18", "m19", "m20", "m21", "m22", "m23", "m24", "m25", "m26",
           "m27", "m28", "m29", "m30", "m31")
Z <- data.frame(model, MyDf, AICsNum, Delta, wi)
Z <- round(Z, digits = 3)
colnames(Z) <- c("Model", "Df", "AICc", "AICc differences", "Akaike weights")
rownames(Z) <- c("m1", "m2", "m3", "m4", "m5", "m6", "m7", "m8", "m9", "m10", "m11", "m12",
               "m13", "m14", "m15", "m16", "m17", "m18", "m19", "m20", "m21", "m22", "m23", "m24", "m25",
               "m26", "m27", "m28", "m29", "m30", "m31")
Z <- Z[order(Z$AICc),]
Z
```





Slika 17.13 Južnoevropska zrva Malpolon monessulanus.

Foto. Franc Janžekovič.

17.3 Polinomska regresija

Odvisna spremenljivka Y je lahko nelinearna funkcija razlagalne spremenljivke X . V takem primeru lahko prileganje razložimo z uporabo polinoma vrednosti x . Model gradimo, kot je predstavljeno pri multipli enačbi, s to razliko, da je Y polinomska funkcija X , model pa se imenuje polinomska regresija k stopnje:

$$Y = b_0 + b_1X + b_2X^2 + \dots + b_kX^k + \varepsilon.$$

Z višanjem stopnje enačbe se povečuje število prevojev, ki oblikujejo krivuljo. Polinomska krivulja prve stopnje je premica, druge stopnje je parabola (z vrhom obrnjenim navzgor ali navzdol), krivulja tretje stopnje ima obliko črke S. Višanje števila stopenj pomeni boljše prileganje in višanje determinacijskega koeficienta R^2 . Cena za višanje stopenj je zmanjševanje prostostnih stopenj in posledično nižanje statistične značilnosti posameznih regresijskih koeficientov in celotnega modela.

Glavni izziv je iskanje modela polinomske enačbe, ki ne krši načela varčnosti. Postopek izgradnje in sprejema končnega modela je identičen tistemu, ki je predstavljen pri multipli linearni regresiji. Model začnemo pri izbranem številu stopenj, npr. k , h kateremu dodajamo ali odvezujemo posamezne člene modele po stopenjskem postopku. Modele ocenjujemo na podlagi testiranja ničelne hipoteze in signifikantnega prispevka posameznega člana k celotnemu modelu. V informacijsko – teoretičnega pristopu pa izdelamo kombinacijo vseh možnih modelov in jih ocenjujemo z Akaikovim informacijskim kriterijem AIC.



PRIMER 17.8

Sesalci in velikost otoka, polinomska regresija – stopenjski model

Pri raziskovanju odnosa med plenom in plenilcem na primeru prehrane pegaste sove na mediteranskih otokih, smo preverjali povezavo med deležem malih sesalcev v prehrani pegaste sove (Slika 17.13) in velikostjo otoka (Janžekovič in Klenovšek 2021). Razmerje ni linearno (Slika 17.14). Začeli smo s polinomskim modelom četrte stopnje:

$$\text{sesalci} \sim a + I(a^2) + I(a^3) + I(a^4),$$

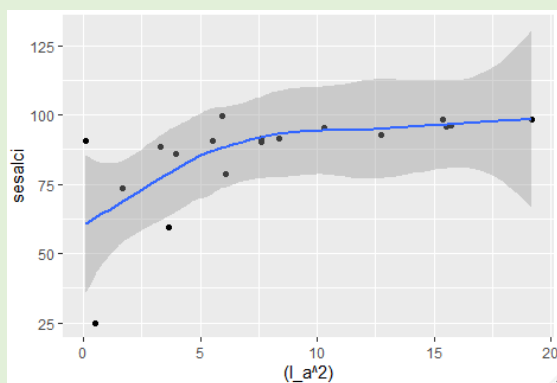
po načelu varčnosti ohranimo kvadratni model (Preglednica 17.7):

$$\text{sesalci} \sim a + I(a^2),$$

ki razloži 64 % variabilnosti $R^2 = 0,638$ in $R^2_a = 0,589$, model je značilen, $F_{2,15} = 13,2$ $p = 0,0005$.

Preglednica 17.7 Regresijski koeficienti, t-test in VIF

| | Koeficient | Std. napaka | t | p | VIF |
|-------------|------------|-------------|------|----------|-------|
| (Intercept) | 54,674 | 6,657 | 8,21 | < 0,0001 | |
| I_a | 19,973 | 5,601 | 3,57 | 0,0028 | 6,803 |
| $I(I_a^2)$ | -2,416 | 1,288 | 1,88 | 0,0804 | 6,803 |



Slika 17.14 Razsevni diagram deleža sesalcev v prehrani pegaste sove in velikost otoka.

```
R >
#Grajenje modela
poli<-lm(sesalci~I_a+I(I_a^2)+I(I_a^3)+I(I_a^4), data =
otok)
summary(poli)
```

```
R > package = »ggplot2«
#Preverjanje razmerja, plot
ggplot(otok, aes((I_a^2), sesalci) ) + geom_point() +
```





Slika 17.15 Pegasta sova *Tyto alba*, pogosto nastopa kot modelni organizem v analizah odnosa med plenilcem in plenom.

Foto. Gregor Bernard.

17.4 Primerjava matrik Mantelov test

Mantelov test (ang. Mantel test) je metoda za primerjanje simetričnih matrik podobnosti S ali razdalj D , ki se nanašata na izračune na istem objektu. Z Mantelovim testom preverjamo hipotezo o povezanosti med matrikama.

Sklepna statistika temelji na raziskovalnem vprašanju o povezanosti med dvema simetričnima matrikama. Ničelno hipotezo oblikujemo H_0 – razdalje med objekti v matriki D_Y niso v povezavi z razdaljami v korespondenčni matriki D_X . Alternativna hipoteza je H_a – razdalje med objekti v matriki D_Y so v povezavi z razdaljami v korespondenčni matriki D_X . Izvedemo izračun Mantelove statistike in na podlagi p vrednosti sklepamo na statistično značilnost povezanost med matrikama, korelacijski koeficient r je merilo moči povezave.

Mantelov test v biodiverzitetnih raziskavah pogosto uporabljajo, npr. pri preverjanju povezave med geografskimi razdaljami med nahajališči, iz katerih izvirajo vzorci, ter vrstnim bogastvom teh lokalitet; v preverjanju podobnosti taksonov v morfometričnem in genetskem kontekstu; v preverjanju vzorcev, nabranih v različnih časovnih obdobjih ipd.



PRIMER 17.9

Morfometrična variabilnost v prostoru – Mantelov test

Za odgovor na raziskovalno vprašanje, ali so vrste, ki geografsko živijo bližje tudi bolj morfološko podobne, potrebujemo dva niza podatkov: morfometrični in geolokacijski. Postavimo hipotezi: H_0 – morfometrične razdalje niso povezane z geografskimi razdaljami; H_a – morfometrične razdalje so povezane z geografskimi razdaljami. Standardizirane povprečne morfometrične vrednosti za 6 vrst so oblikovane v pravokotno matriko 6 x 7 (6 vrst v vrsticah – objekti in 7 morfometričnih znakov v stolpcih – lastnosti).

Med vrstami smo izračunali Evklidsko razdaljo in izrazili odnos med njimi v obliki simetrične matrike 6 x 6. Geolokacija vrst je podana z navedbo geografske širine in dolžine v pravokotni matriki 6 x 2 (6 lokalitet ter Latitude in Longitude). Med lokacijami smo izračunali razdalje in jih predstavili v obliki simetrične matrike 6 x 6.

Izračun povezanosti med matrikama morfometričnih in geografskih razdalj testiramo z Mantelovim testom, ki vrne značilno in močno korelacijo $r = 0,762$, $p = 0,0292$.

Na podlagi p vrednosti ($p < 0,05$) zavrnamo H_0 in trdimo, da so morfometrične razdalje povezane z geografskimi razdaljami. Sprejememo sklep, da so vrste, ki so geografsko bližje skupaj, si tudi bolj podobne med seboj.

```
R > package = »geosphere«  
#Pretvorba geografskih koordinat v diagonalno matriko - haversine distance  
dgeo <- distm(dapo, fun = distHaversine)  
dgeo
```

```
R > package = »vegan«  
# Mantel test  
mantel <- mantel(deud, dgeo)  
mantel
```

```
R > package = »vegan«  
#Izračun Evklidovih razdalj  
deuc <- dist(zapo, method = "euclidean")  
deuc
```



Zaključki

- Podatkovni nizi z ekološkimi in biodiverzitetnimi vrednostmi izkazujejo povezanost med lastnostmi, lahko tudi vzdolž ekološkega gradienta.
- Lastnost je spremenljivka, s katero opišemo objekt. Kvantitativne lastnosti vključujejo intervalne spremenljivke, ki jih lahko razvrstimo v naraščajočo vrsto realnih števil. Kvalitativne lastnosti vključujejo kvalitativni opis objekta, podatke uredimo samo po frekvencah. Binarni deskriptorji vključujejo samo stanje prisotno-odsotno.
- Objekt je predmet raziskave, definiran je vnaprej.
- Z linearno regresijsko analizo opišemo odnos med odvisno spremenljivko Y in eno ali več neodvisnimi spremenljivkami X_1, X_2, \dots, X_m , da bi lahko razložili odnos med njimi ter napovedali ali predvideli vrednost y pri dani vrednostih x .
- Odnos med spremenljivkama zapišemo z linearno regresijsko enačbo: $Y = a + bX + \varepsilon$, a je konstantni člen, b je smerni koeficient premice. Regresijska premica je položena skozi oblak točk v taki smeri, da so ostanki ε najmanjši (metoda najmanjših kvadratov). Z regresijskim modeliranjem opišemo odnos med lastnostmi, sklepamo o hipotezah ali napovedujemo in predvidevamo.
- Enostavna linearna regresija je metoda za opis linearnega odnosa med spremenljivkama Y in X .
- Če neodvisna spremenljivka X ne izkazuje napak, je kontrolirana in vse napake pripišemo odvisni spremenljivki Y ; ostanke računamo po vertikalni osi (regresijski model I).
- Kadar obe spremenljivki izkazujeta napake, gre za regresijski model II. Ostanke računamo kot pravokotne razdalje med točkami in regresijsko premico (metoda glavne osi). Spremenljivki sta v istih fizikalnih enotah. Vpliv absolutnih vrednosti odpravimo s standardizacijo (metoda standardne glavne osi) ali z rangiranjem (metoda rangirane glavne osi). Metoda navadnih najmanjših kvadratov predpostavlja, da spremenljivka X nima ostankov.
- Z multiplo linearno regresijo razlagamo odvisno spremenljivko z več neodvisnimi spremenljivkami: $Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_mX_m + \varepsilon$; model oblikujemo po načelu varčnosti.
- Če je Y nelinearna funkcija X , lahko za prilaganje uporabimo polinom vrednosti x . Enačba modela polinomske regresije: $Y = b_0 + b_1X + b_2X^2 + \dots + b_kX^k + \varepsilon$; model oblikujemo po načelu varčnosti.
- Pri oblikovanju regresijskih modelov iščemo kompromis med enostavnostjo modela - načelo varčnosti in prilaganja modela k podatkom – bias.
- Multipli regresijski model lahko zgradimo s testiranjem ničelne hipoteze, pri tem upoštevamo statistično značilnost regresijskih koeficientov, prispevek razlagalne spremenljivke k višanju R^2 in njen nizek VIF.

- Multipli regresijski model lahko zgradimo z informacijsko – teoretičnim pristopom, kjer izdelamo kandidatne modele in jih razvrstimo glede na najnižjo vrednost AIC.
- Povezanost simetričnih matrik podobnosti ali razdalj preverjamo z Mantelovim testom.

Literatura

- Bartoń K. 2022. MuMIn: Multi-Model Inference. R package version 1.47.1, <https://CRAN.R-project.org/package=MuMIn>.
- Burnham K.P., Anderson D.R. 2002. Model selection and multimodel inference, 2nd Edn. Springer. New York.
- Fox J., Weisberg S. 2019. An R Companion to Applied Regression, Third Edition. Thousand Oaks CA. Sage. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.
- Hijmans R.J. 2019. geosphere: Spherical Trigonometry. R package version 1.5-10. <https://CRAN.R-project.org/package=geosphere>.
- James G., Witten D., Hastie T., Tibshirani R. 2021. An Introduction to Statistical Learning with Applications in R. Second Editions. Springer Science+Business Media, LLC. https://doi.org/10.1007/978-1-0716-1418-1_1.
- Janžekovič F., Klenovšek T. 2020. The biogeography of diet diversity of barn owls on Mediterranean islands. *Journal of Biogeography*. <https://doi.org/10.1111/jbi.13955>.
- Jesenko J., Jesenko M. 2007. Multivariatne statistične metode. Moderna organizacija. Kranj.
- Legendre P. 2018. lmodel2: Model II Regression. R package version 1.7-3. <https://CRAN.R-project.org/package=lmodel2>.
- Legendre P., Legendre L. 2012. Numerical Ecology. Elsevier. Amsterdam.
- Oksanen J., Blanchet F. G., Friendly M., Kindt R., Legendre P., McGlinn D., Minchin P. R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., Szoecs E., Wagner H. 2019. vegan: Community Ecology Package. R package version 2.5-6. <https://CRAN.R-project.org/package=vegan>.
- R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Sokal R.J., Rohlf F.J. 2012. Biometry. W.H. Freeman Company. New York.
- Venables W.N., Ripley B. D. 2002. Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0.
- Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag. New York.
- Wickham H., François R., Henry L., Müller K. 2020. dplyr: A Grammar of Data Manipulation. R package version 0.8.5. <https://CRAN.R-project.org/package=dplyr>.

18. poglavje



STATISTIČNO SKLEPANJE

Povzetek

18. poglavje: Statistično sklepanje

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

V znanosti je splošno sprejeti koncept statističnega sklepanja s testiranjem ničelne hipoteze. V procesu odločanja po navadi zastavimo dve nasprotujoči si trditvi. Ničelna hipoteza H_0 , po kateri ne obstaja razlika med dvema populacijama, in alternativna hipoteza H_a , po kateri obstaja razlika med populacijama. Pri zastavi hipotez ne vemo, katera od hipotez je pravilna in katera predstavlja dejansko stanje v naravi. S pomočjo verjetnosti in statističnega sklepanja poskušamo izluščiti pravilno trditev. Pravilo, po katerem na podlagi vzorca ničelno hipotezo H_0 sprejmemo ali ne, se imenuje statistični test.

Pravilo statističnega testa ne velja vedno. Zaradi slučajnega vzorca se lahko zgodi, da sprejmemo napačno hipotezo. Kadar zavrnilo pravilno ničelno hipotezo H_0 , smo naredili napako prve vrste. Kadar sprejmemo napačno ničelno hipotezo H_0 , smo naredili napako druge vrste. Obe zagrešeni napaki sta v procesu statistične analize nezaželeni, vendar se temu v celoti ne moremo izogniti. Zato si prizadevamo, da sta kar se da malo verjetni. Za napako prve vrste predpišemo zgornjo mejo verjetnosti α praviloma pri $\alpha = 0,05$ in jo imenujemo stopnja značilnosti.

Pri preizkušanju hipoteze, ko alternativna hipoteza ni enostavna hipoteza, po navadi preverjamo statistično značilno razliko samo za en statistični parameter, npr. povprečje. Izračunamo lahko samo verjetnost za napako prve vrste, verjetnost napake druge vrste ne moremo določiti. Ničelno hipotezo ali zavrnilo pri $p < 0,05$ ali o ničelni hipotezi ne odločimo pri $p > 0,05$. Na podlagi testa značilnosti se izognemo napaki druge vrste.

Pri sprejemanju enostavne hipoteze sta znana oba parametra, ki ju testiramo (tako za H_0 , kot za H_a). Izračunamo verjetnost za napako prve vrste in verjetnost za napako druge vrste. Glede na predpisano stopnjo značilnosti za napako prve vrste $\alpha = 0,05$, ničelno hipotezo zavrnilo pri $p < 0,05$ in sprejmemo alternativno hipotezo; lahko pa glede na $p > 0,05$ in $\beta < 0,2$ sprejmemo ničelno hipotezo. Z β označimo verjetnost za sprejetje ničelne hipoteze, ko je ta napačna, β je povezana z močjo testa P , to je verjetnost za zavrnitev ničelne hipoteze, ko je ta napačna $P = 1 - \beta$.

Odnosi med verjetnostjo za napako prve vrste, verjetnostjo za napako druge vrste, velikostjo učinka in velikostjo vzorca so taki, da je vsak od njih funkcija preostalih treh. Ko pridobimo tri kriterije je s tem določen tudi četrti. Velikost učinka je vrednost, s katero podamo velikost razlike med ničelno in alternativno hipotezo. Je merilo odstopanja od ničelne hipoteze in informacija za zanesljivost sprejema alternativne hipoteze. V biodiverzitetni praksi nas zanima predvsem odnos med stopnjo značilnosti α in velikostjo vzorca.

Presoj o velikosti vzorca za oceno vrstnega bogastva naredimo s krivuljo kopičenja vrst, ki je funkcija vložene delo v vzorčenje. Delo v vzorčenje merimo kot: število vzorčenj, število vzorčenih lokalitet, število registriranih osebkov, število ur, dni ali let vzorčenj. V primeru, ko akumulacijska krivulja vrst doseže asimptoto, je to pokazatelj, da smo vrstno bogastvo zadostno raziskali. V vrednotenje statističnih parametrov moramo vključiti zadostno število meritev, da zagotovimo reprezentativnost vzorca. Potrebno število meritev ocenimo s pomočjo velikosti učinka, močjo testa in stopnje značilnosti.

V sodobni naravoslovni znanosti (sploh v ekoloških in biodiverzitetnih raziskavah) je splošno sprejet koncept statističnega sklepanja s **testiranjem ničelne hipoteze NHST** (ang. Null Hypothesis Significance Testing). V procesu odločanja o natančno definirani vsebini, raziskovalec zastavi dve nasprotujoči si trditvi. V prvi trditvi, ki ji pravimo **ničelna hipoteza H_0** , trdi, da ne obstaja razlika med dvema populacijama (npr. pri primerjanju aritmetičnih sredin). V drugi trditvi, **alternativna hipoteza H_a** , trdi, da obstaja razlika med dvema populacijama. Raziskovalec, ki zastavi hipotezi, ne ve, katera od hipotez je pravilna oz. katera od hipotez predstavlja dejansko stanje v naravi. S pomočjo verjetnosti in statističnega sklepanja skuša iz podatkov izluščiti pravilno trditev.

Na začetku vpeljemo nekaj matematičnih pojmov, da lahko pridemo k posamezni biološki oz. biodiverzitetni interpretaciji. **Populacija** v statističnem kontekstu je množica P , ki jo sestavlja znano ali neznan število elementov populacije, ki jih imenujemo **statistične enote**. Na enotah opazujemo ali merimo neko merljivo količino X . Tako dobimo empirične podatke. Podatke urejamo, opisujemo, analiziramo oz. na različne načine statistično vrednotimo. V statistično vrednotenje imamo le izjemoma možnost vključiti celotno populacijo. Običajna praksa je opazovanje samo njene podmnožice H . Podmnožico H populacije P imenujemo **vzorec V** . Vzorec je del populacije, na katerem proučujemo določene lastnosti. Naloga statističnega vrednotenja je tudi definiranje pogojev, s katerimi lahko iz podatkov na vzorcu veljavno sklepamo na lastnost celotne populacije. Vzorec mora biti **represntativen**, to pomeni, da vsebuje zadostno število statističnih enot, ki dobro predstavljajo populacijo. Reprezentativnost vzorca zgotovimo z naključno izbiro zadostnega števila enot, zato mu rečemo **slučajni vzorec**. Z n -kratnim merjenjem količine X na slučajnem vzorcu, dobimo **slučajno spremenljivko X_i** . Število n je **velikost vzorca**.

Domneva o porazdelitvi slučajne spremenljivke X_i se imenuje **statistična hipoteza**. Hipoteza, ki natančno določa bodisi samo statistični parameter (npr. aritmetično sredino) bodisi samo porazdelitev spremenljivke (npr. normalna porazdelitev) se imenuje **enostavna**; vsaka drugačna hipoteza je **sestavljena** (npr. določa parameter in porazdelitev). Hipotezi, ki jo želimo preizkusiti, pravimo **ničelna hipoteza H_0** . Nasproti nje postavimo **alternativno hipotezo H_a** . Pravilo, po katerem na podlagi vzorca ničelno hipotezo H_0 sprejmemo ali ne sprejmemo, se imenuje **statistični test**. Slučajna spremenljivka X_i ima zalogo vrednosti w , ki se imenuje **prostor vzorca**, znotraj prostora vzorca določimo podmnožico vrednosti w_k **kritično območje K** . Pravilo statističnega testa je oblikovano tako, da če vzorec V_i vsebuje elemente podmnožice w_k , potem sprejmemo H_a ; če pa vzorec V_i ne vsebuje elementov podmnožice w_k , potem H_a zavrnamo.

Pravilo statističnega testa ne velja vedno, oz. ni idealno. Zaradi slučajnega vzorca se zgodi, da vzorec V_i ne vsebuje elementov podmnožice w_k iz kritičnega območja in H_a napačno sprejmemo. Prav tako se zgodi, da vzorec V_i vsebuje elemente podmnožice w_k in H_a napačno zavrremo. Kadar zavrremo pravilno ničelno hipotezo H_0 , smo naredili **napako prve vrste**. Kadar sprejmemo napačno ničelno hipotezo H_0 , smo naredili **napako druge vrste**. Obe zagrešeni napaki sta v procesu statistične analize nezaželeni, vendar se temu v celoti ne moremo izogniti. Zato si prizadevamo, da bosta kar se da malo verjetni. Za napako prve vrste po navadi predpišemo zgornjo mejo verjetnosti α , ki jo imenujemo **stopnja značilnosti**. Na tem mestu moramo sprejeti še eno odločitev, povezano z napako druge vrste. Povezana je z alternativno hipotezo H_a . V primeru, da je alternativna hipoteza sestavljena, to pomeni, da parameter za alternativno hipotezo ni predpisan s točno vrednostjo, problem ni rešljiv. Tak postopek ni več statistični test, imenuje se **preskus značilnosti** (sin. test značilnosti).

Preskus značilnosti. V preskusu značilnosti, s katerim preverjamo sestavljene hipoteze, je mogoča samo napaka prve vrste, njena največja dovoljena verjetnost je predpisana s stopnjo značilnosti α . Najpogosteje predpisane stopnje značilnosti so 0,05, 0,01 ali 0,001. Kadar v preskusu značilnosti H_0 zavrremo, pravimo, da je razlika značilna (sin. signifikantna). Kadar pa v preskusu značilnosti H_0 ne moremo zavrniti, oz. ne odločimo, rečemo, da razlika ni značilna. V preizkusu značilnosti **ne sprejemamo ničelne hipoteze**, temveč jo bodisi zavrremo in sprejmemo alternativno hipotezo H_a bodisi o H_0 ne odločimo. Ker o sprejemu ničelne hipoteze ne odločimo, ne moremo zagrešiti napake druge vrste.

Statistični test. V statističnem testu, s katerim preverjamo enostavno hipotezo, sta mogoči obe vrsti napake. Napaka prve vrste, v kateri zavrremo ničelno hipotezo, in napaka druge vrste, v kateri sprejmemo ničelno hipotezo (Preglednica 18.1). Verjetnost za napako prve vrste je vrednost, označena z α in je predpisana praviloma pri $\alpha = 0,05$. Verjetnost za napako druge vrste je vrednost, označena z β in je praviloma predpisana pri $\beta = 0,2$. Tveganje za napačni sprejem ničelne hipoteze se imenuje **moč testa** P in velja $P = 1 - \beta$.

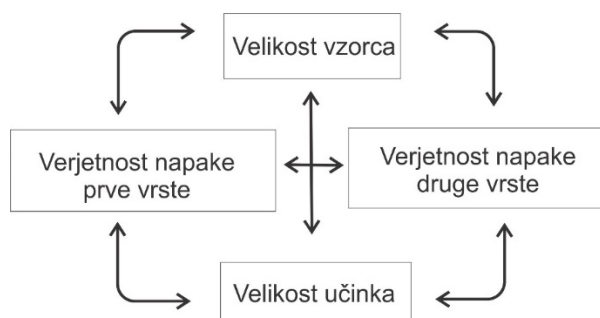
Preglednica 18.1. Odločitvena shema pri sprejemanju hipotez glede na statistični izračun in dejansko stanje

| Statistični test | | Dejansko stanje v naravi | |
|--------------------------------|---------------------------------|--|--|
| | | Resnična je ničelna hipoteza H_0 | Resnična je alternativna hipoteza H_a |
| Naša odločitev, rezultat testa | Potrdimo ničelno hipotezo | Pravilna odločitev Verjetnost = $1 - \alpha$ | Napaka druge vrste Verjetnost = β |
| | Sprejmemo alternativno hipotezo | Napaka prve vrste Verjetnost = α | Pravilna odločitev Verjetnost = $1 - \beta$; moč testa |

V statističnem testu sprejmemo odločitev bodisi o zavrnitvi ničelne hipoteze in sprejetjem alternativne hipoteze bodisi o zavrnitvi alternativne hipoteze in sprejetju ničelne hipoteze. Ker odločamo tako o sprejemu ničelne hipoteze kot o sprejemu alternativne hipoteze, lahko zagrešimo obe vrsti napake, napako prve vrste ali napako druge vrste. Odnos med napako prve vrste in napako druge vrste je praviloma tak, da z zmanjševanjem verjetnosti napake prve vrste višamo verjetnost za napako druge vrste. Pri iskanju ravnovesja med obema napakama lahko na izid vplivamo z velikostjo vzorca ali z velikostjo kritičnega območja.

Odnosi med verjetnostjo za napako prve vrste, ki jo podamo z α , verjetnostjo za napako druge vrste, ki jo podamo z β in predstavlja moč testa $P = 1 - \beta$, velikostjo učinka ES in velikostjo vzorca n so taki, da je vsak od njih funkcija preostalih treh. Z drugimi besedami, če fiksiramo tri kriterije, s tem definiramo tudi četrtega (Slika 18.1). Izhajajoč iz zapsanega, velja:

- P kot funkcija α , ES in n ,
- n kot funkcija α , ES in P ,
- ES kot funkcija α , P in n ter
- α kot funkcija P , ES in n .



Slika 18.1 Odnosi med velikostjo vzorca, napako prve vrste in napako druge vrste ter velikostjo učinka (prirejeno po Krebs 2014).

PRESKUS ZNAČILNOSTI IN STATISTIČNI TEST

Pri testiranju ničelne hipoteze je alternativna hipoteza bodisi enostavna bodisi sestavljena. Enostavno hipotezo preverimo s statističnim testom, sestavljeno hipotezo pa s preskusom značilnosti.

Pri sprejemanju sestavljene hipoteze preverjamo razliko samo za en statistični parameter, npr. povprečje $H_0: \bar{x} = 61; H_a: \bar{x} \neq 61$. Izračunamo lahko samo verjetnost za napako prve vrste, napake druge vrste pa ne moremo določiti. To pomeni, da glede na predpisano stopnjo značilnosti npr. $\alpha = 0,05$, ničelno hipotezo zavrnemo pri $p < 0,05$ ali pa o ničelni hipotezi ne odločimo pri $p > 0,05$. Na podlagi tega pravila se izognemo napaki druge vrste. Pri sprejemanju enostavne hipoteze sta znana oba parametra, ki ju testiramo, tako za H_0 , kot za H_a , npr. $H_0: \bar{x} = 61; H_a: \bar{x} = 65$. V tem primeru lahko izračunamo verjetnost za napako prve vrste in verjetnost za napako druge vrste. Glede na predpisano stopnjo značilnosti za napako prve vrste $\alpha = 0,05$, ničelno hipotezo zavrnemo pri $p < 0,05$ in sprejmemo alternativno hipotezo ali pa glede na $p > 0,05$ in $\beta < 0,2$ sprejmemo ničelno hipotezo. Z β označimo verjetnost za sprejetje ničelne hipoteze, ko je ta napačna, β je povezana z močjo testa P . Moč testa je verjetnost za zavrnitev ničelne hipoteze, ko je ta napačna $P = 1 - \beta$.

V biodiverzitetni praksi nas zanima predvsem odnos med stopnjo značilnosti α in velikostjo vzorca.

V **biodiverzitetni** obravnavi vrstnega bogastva, vrstne pestrosti ali združb organizmov imata statistična termina »populacija« in »vzorec« pomensko vsebinsko drugačen kontekst. Tako v statistični analizi vrstnega bogastva ali vrstne pestrosti »populacija« pomeni vse obravnavane vrste organizmov na raziskovalnem območju, »vzorec« pa reprezentativni del te združbe organizmov. Statistične enote so torej posamezne vrste organizmov, vrednosti slučajne spremenljivke X_i pa bodisi njihova prezenca bodisi njihova abundanca.

Populacijo, v ekološkem kontekstu, tvorijo vsi osebki izbrane vrste na določenem območju in v določenem času. V biodiverzitetni raziskavi npr. morfometrične variabilnosti praviloma opazujemo samo del populacije, ki mu rečemo vzorec. Vzorec je torej le del populacije in smo ga iz populacije izbrali z namenom preučevanja neke morfometrične lastnosti te populacije. Da lahko iz podatkov na vzorcu veljavno sklepamo na razmere v celotni populaciji mora biti vzorec reprezentativen, to pomeni, da dobro predstavlja lastnosti populacije. Reprezentativnost vzorca zagotovimo z naključno izbiro osebkov, ki vzorec sestavljajo, zato pravimo, da so organizmi, vključeni v vzorec,

naključno izbrani. Izziv, s katerim se raziskovalec sreča pri naključnem vzorcu, je heterogenost virov morfološke variabilnosti. Na morfometrične lastnosti osebkov praviloma vplivajo tudi starost, spol in dejavniki okolja. V skladu z namenom raziskave lahko posamezni vir variabilnosti ločeno analiziramo ali pa iz slučajnega vzorca naredimo izbor organizmov, na katerih naredimo statistično analizo. Kriterije za oblikovanje vzorca seveda razložimo.

Vsebina poglavja je narejena po temeljni literaturi Cohen (1988), Faul in sod. (2009), Krebs (2014), Legendre in Legendre (2012), Sokal in Rohlf (2012).

V tem poglavju so predstavljene vsebine:

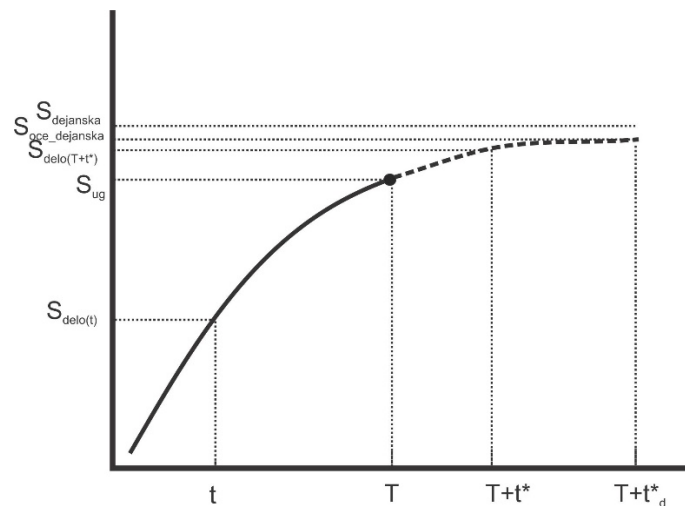
- Velikost vzorca.
- Velikost učinka.
- Analiza moči.

18.1 Velikost vzorca

Vzorčenje in načrt raziskave sta termina iz statističnega žargona, prevedena v jezik biodiverzitetne raziskave pomenita: – kje naj vzorčim? – kako naj zbiram podatke v prostoru in času? in – koliko vzorcev potrebujem? V tem odseku so predstavljene vsebine o velikosti vzorca v raziskavi vrstnega bogastva in velikosti vzorca pri analizi statističnih parametrov.

18.1.1 Velikost vzorca za oceno vrstnega bogastva

Poseben izziv v biodiverzitetnih raziskavah je presoja o velikosti vzorca za oceno vrstnega bogastva. Odločitev o številu vzorčenj najprej izhaja iz namena raziskave. Ko je bistvo raziskave v popisu in oceni vrstnega bogastva ali vrstne pestrosti izbranega območja, **število potrebnih vzorčenj ocenimo s pomočjo krivulje kopičenja vrst SAC** (glej primere v poglavju 3.2), ki je funkcija vložnega dela v vzorčenje. Delo v vzorčenju merimo različno glede na namen raziskave. Po navadi ga izrazimo kot: število vzorčenj, število vzorčenih lokalitet, število registriranih osebkov, število ur, dni ali let vzorčenj in podobno. Koncept dela ima torej prostorsko in/ali časovno dimenzijo. Odnos med vložnim delom in številom vrst, ki je predstavljen z zglašeno krivuljo (Slika 18.2), lahko ekstrapoliramo in podaljšamo. Na tak način predvidimo povečanje števila vrst glede na vloženo delo. V primeru, ko akumulacijska krivulja vrst doseže asimptoto, je to pokazatelj, da smo vrstno bogastvo zadostno raziskali in lahko zaključimo z vzorčenjem.



Slika 18.2 Krivulja kopičenja vrst. Delo, ki je bilo vloženo v vzorčenje, je predstavljeno na X osi in kumulativno število vrst na Y osi. Polna krivulja predstavlja empirični del in črtni del krivulje ekstrapolirani del vrednosti za neko dejansko vrednost vrstnega bogastva $S_{dejanska}$. Delo, ki je vloženo v vzorčenje T , daje število ugotovljenih vrst S_{ug} . Interpolacija omogoča določitev števila vrst $S_{delo(t)}$ ob delu t . Ekstrapolacija omogoča: 1. bodisi določitev števila vrst $S_{delo(T+t^*)}$ ob povečanju raziskovalnega dela $T + t^*$, 2. bodisi določitev količine vložnega dela $T + t^*_d$ za določitev ocene dejanskega števila vrst v združbi $S_{oce_dejanska}$ (Prirejeno po Colwell in sod. 2012).

18.1.2 Velikost vzorca za oceno statističnega parametra

V vrednotenje statističnih parametrov za izbrano spremenljivko X (npr. povprečje, varianca) moramo vključiti zadostno število meritev n (velikost vzorca), da zagotovimo reprezentativnost vzorca oz. zadostimo statističnim zakonitostim, predvsem porazdelitvi vrednosti oz. variabilnosti. Med t.i. statistične legende sodi »pravilo $n=30$ «. Bolje kot nekritično sledenje temu pravilu priporočamo izračun vrednosti n . **Pravilo $n = 30$** (ang. » $n = 30$ Rule of Thumb«) je priporočilo za velikost vzorca za statistično oceno za izračun parametra. V biodiverzitetnih raziskavah pogosto sledimo priporočilu, da je število vzorcev vsaj 30 ali več. Takemu vzorcu rečemo velik vzorec. Za vrednosti s porazdelitvijo blizu normalni je »pravilo $n = 30$ «, uporabno in smiselno. Ko pa imamo opravka z vzorci z zelo izraženo variabilnostjo, v takih primerih 30 meritev ne dosega standardov za izračun ustreznih statistik. Kakorkoli že, v biodiverzitetni praksi se pogosto znajdemo pred izzivom, da je meritev manj kot 30, včasih tudi manj kot 10 in števila meritev ne moremo zvišati. V takih primerih smo precej omejeni z naborom statističnih analiz in rezultate statistik interpretiramo z zadržkom.

Tip spremenljivke in metoda zbiranja meritev zelo zaznamuje izvedbo vzorčenja v naravi in seveda posledično tudi velikost vzorca. Velikost vzorca je eden od ključnih elementov v analizi statističnega sklepanja. V biodiverzitetnih raziskavah se srečujemo s

spremenljivkami vseh tipov: nominalne spremenljivke npr. lokaliteta vzorčenja, ordinalne spremenljivke npr. starostne kategorije, intervalne spremenljivke npr. telesna dolžina, ali število predstavnikov ciljne vrste organizma. Kjer je to mogoče, poskušamo spremenljivko pretvoriti v intervalni tip, npr. lokaliteto zapišemo z navedbo geografskih koordinat, starostno kategorijo pretvorimo v leta ali dni starosti ipd. V *a priori* pristopu, torej v stopnji načrtovanja eksperimenta, lahko v skladu s hipotezo in predvideno močjo testa ter vplivom velikosti učinka ocenimo ciljno vrednost za velikost vzorca. Po drugi strani smo pri realizaciji meritev lahko omejeni s številom vzorcev. Najpogostejši primeri omejitev v biodiverzitetnih raziskavah so prevelik strošek izvedbe vzorčenja (časa, denarja, materiala ...), nedostopnost objekta raziskave in etična nesprijemljivost (odvzem organizmov iz narave).

Poglejmo izračun velikosti vzorca za intervalno spremenljivko X , s povprečjem \bar{x} , standardnim odklonom StD , standardno napako SE in vzorčnimi vrednostmi $x_1, x_2, x_3, \dots, x_n$. Intervalna spremenljivka ima vrednosti na intervalu (min–max) in je praviloma normalno porazdeljena. Interval zaupanja (ang. Confidence Interval) s stopnjo zaupanja $1 - \alpha$, po navadi je to 0,95 oz. 0,05 %, za populacijsko povprečje μ izračunamo v postopku:

$$\bar{x} - t_{\alpha} SE, \bar{x} + t_{\alpha} SE,$$

kjer je t_{α} kritična vrednost t -porazdelitve z $n - 1$ prostostnimi stopnjami. Za velik vzorec $n \geq 30$ lahko vzamemo za $t_{0,05} = 1,96$ in za $t_{0,01} = 2,58$. Oceno velikosti vzorca izvedemo v treh korakih.

Korak 1. **Odločitev o stopnji natančnosti d** . Ali je za našo raziskavo dovolj 10 % natančnost ali imamo zahtevnejše pogoje, npr. 1 % natančnost skladnosti med ocenjenim povprečjem in resničnim povprečjem. Odločitev ni statistična, ampak izhaja iz ekologije in zadeva hipotezo, ki jo preverjamo. Odločitev o stopnji natančnosti d lahko podamo v absolutnem ali relativnem številu, razmerje med relativno napako RN in absolutno napako AN je:

$$RN = \left(\frac{AN}{\bar{x}} \right) * 100$$

Tukaj opozarjamo na različno navajanje natančnosti v besedilnem in grafičnem prikazu. Večina statističnih programov ima v prikazu grafičnih rezultatov vgrajen prikaz 95 % natančnosti, to se dobro sklada s približkom dveh standardnih napak $\bar{x} \pm 2SE$, v

besedilnem zapisu je pogostejša raba $\bar{x} \pm SE$. Pri branju moramo biti pazljivi na ta dva načina podajanja.

Korak 2. **Enačba, ki povezuje velikost vzorca n in natančnost ocene povprečja d .**

Enačba je odvisna od tipa spremenljivke in variabilnosti spremenljivke. Za spremenljivko z normalno porazdelitvijo uporabimo enačbo:

$$d = \frac{t_{\alpha} StD}{\sqrt{n}} \quad \text{oz. preurejeno: } n = \left(\frac{t_{\alpha} StD}{d} \right)^2$$

Korak 3. **Ocena neznanega parametra populacije za rešitev enačbe.** Za rešitev enačbe potrebujemo oceno za standardno deviacijo spremenljivke. Standardno deviacijo lahko ocenimo na različne načine, lahko s predhodnim vzorčenjem podobne populacije (lahko tudi iz literature) ali s pilotno preliminarna študijo (Krebs 2014).

Izračun velikosti vzorca za izračun aritmetične sredine je predstavljen v nadaljevanju v Primeru 18.1.



PRIMER 18.1

Ocena velikosti vzorca za izračun povprečja intervalne spremenljivke

Kondilobazalna dolžina lobanje CBL je standardna mera za podajanje dolžine sesalčje lobanje. Pri analizi neznane populacije gozdne voluharice *Clethrionomys glareolus* nas zanima ocena, kako velik vzorec moramo vključiti v raziskavo, da bo vzorčna aritmetična sredina zelo blizu dejanske vrednosti parametra CBL. Iz literature smo povzeli, da je povprečje dosedanjih raziskav za CBL $\bar{x} = 24,08$ mm in standardna deviacija $StD = 0,939$ mm. Izračun izvedemo v dveh korakih. V prvem definiramo stopnjo natančnosti oz. skladanje med vzorčno in dejansko vrednostjo, v drugem koraku ocenimo variabilnost spremenljivke CBL.

1. Ocena natančnosti d : $RN = 1$ % relativna napaka

Formula za izračun absolutne napake AN :

$$\begin{aligned} AN &= \frac{RN \cdot \bar{x}}{100} \\ AN &= \frac{1 \cdot 24,08}{100} \\ AN &= 0,24 \text{ m} \end{aligned}$$

2. Ocena za velikost vzorca n

Formula za izračun velikosti vzorca n :

$$n = \left(\frac{t_{\alpha} StD}{d} \right)^2$$

Stopnja značilnosti je $\alpha = 0,05$ oz. $\alpha = 0,01$, vrednost $t_{0,05} = 1,96$ in $t_{0,01} = 2,58$.

$$n = \left(\frac{1,96 \cdot 0,939}{0,24} \right)^2$$

$$n = 58,8$$

Zaključek. Za izračun povprečja, ki bi se z dejanskim povprečjem skladalo z verjetnostjo 95 %, moramo v analizo vključiti vzorec vsaj 59 živali ($n = 58,8$); za 99 % skladnost pa moramo vključiti vsaj 102 živali ($n = 101,9$).



18.2 Velikost učinka

Velikost učinka ES (ang. Effect Size) je vrednost, s katero podamo velikost razlike med ničelno in alternativno hipotezo. Z drugimi besedami, velikost učinka je merilo odstopanja od ničelne hipoteze in informacija za zanesljivost sprejema alternativne hipoteze. Oceno velikosti učinka in analizo moči lahko izvedemo pred raziskavo. S tem ocenimo, kako velika naj bo razlika med hipotezama, da lahko sprejmemo alternativno, s tem tudi definiramo velikost vzorca.

Testiranje hipoteze je kljub svojim pomanjkljivostim, najboljši pristop pri sprejemanju odločitev v ekoloških in biodiverzitetnih raziskavah. Za zmanjševanje šibkosti in v podporo testiranju hipotez je prišlo do razvoja mer velikosti učinka. Vpliv oz. velikost učinka izračunamo in podamo z vrednostjo, ki je stopnjevana v tri kategorije, vpliv je majhen (šibak), vpliv je srednje močan ali vpliv je velik (močan). Opisne kategorije so nam v pomoč pri interpretaciji odločitve pri izbiri ene od hipotez. Mere velikosti učinka razdelimo v dve skupini: 1. standardizirane razlike med aritmetičnimi sredinami (kadar preverjamo razlike med povprečji) in 2. mere povezanosti (kadar preverjamo stopnjo korelacije). Pri **standardiziranih razlikah** med aritmetičnimi sredinami prikažemo razdalje med aritmetičnimi sredinami vzorcev v enotah standardne deviacije, npr. Cohenov d koeficient. **Mere povezanosti** temeljijo na deležu razložene variance, najpogosteje jo izrazimo s koeficientom determinacije R^2 .

Razlikujemo več vrst velikosti učinka, ob osnovnih dveh skupinah se nadalje razlikujejo glede na test, pri katerem izvajamo vrednotenje za:

- primerjanje povprečij dveh skupin s t -testom: Cohenov d ,
- primerjanje povprečij več skupin z ANOVA: Effect Size d , Effect Size f ,
- učinek velikosti v razliki med dvema korelacijskima koeficientoma: Cohenov g ,
- učinek velikosti za ne parametrične teste (Mann-Whitney, Kruskal-Wallis, Chi kvadrat test): Eta kvadrat (η^2),
- učinek velikosti za korelacijsko analizo: R^2 in
- učinek velikosti za regresijsko analizo: standardizirana β , R^2 .

Izračunane vrednosti posameznih mer učinka velikosti lahko transformiramo v vrednosti drugih mer s pomočjo preračuna, imenovanega »skupni jezik učinka velikosti« (ang. Common Language Effect Size CLES).

Izoblikovana je praksa, da se v strokovnih in znanstvenih besedilih pri navajanju statističnih izračunov navaja tudi velikost učinka (APA 2012 in APA 2021).

V nadaljevanju je predstavljen izbor metod za izračun osnovnih cenilk učinka velikosti.

18.2.1 Mere standardiziranih razlik

Mera standardizirane razlike med povprečji mora biti, po ničelni hipotezi, ničelne velikosti. Ko mera divergira od nič, narašča razlika med ničelno in alternativno hipotezo.

Cohenov d (ang. Cohen's d) je definiran kot razlika med dvema aritmetičnima sredinama \bar{x}_1 in \bar{x}_2 spremenljivke X , podana v standardnih deviacijah StD . Testiranje razlik med aritmetičnima sredinama dveh vzorcev naredimo s t -testom, izračun t -testa poteka posebej za neodvisna oz. odvisna vzorca. Izračun za neodvisna vzorca pri enosmernem testiranju:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{StD}$$

Izračun za neodvisna vzorca pri dvosmernem testiranju:

$$d = \frac{|\bar{x}_1 - \bar{x}_2|}{StD}$$

In izračun za odvisna vzorca, pri velikosti vzorca, kjer je n število parov:

$$d = \frac{t}{\sqrt{n}}$$

Interpretacija učinka velikosti je prevedena najprej v tri kategorije, ki so bile razširjene v pet kategorij od zelo majhen učinek ($d \approx 0,01$) do izjemno velik učinek ($d \approx 2,00$) (Preglednica 18.2).

Preglednica 18.2 Petstopenjsko rangiranje velikosti učinka za Cohenov d

| Velikost učinka | Cohenov d |
|-----------------|-------------|
| Zelo majhen | 0,01 |
| Majhen | 0,20 |
| Srednji | 0,50 |
| Zelo velik | 0,80 |
| Izjemen | 2,0 |

Indeks f z njim ocenimo velikost učinka v analizi variance, npr. pri ANOVA F -testu. S F -testom primerjamo aritmetične sredine treh ali več vzorcev, število vzorcev označimo s k . Izračun f indeksa izvedemo po enačbi:

$$f = \sqrt{\frac{\sum_{i=1}^k p_i (\bar{x}_i - \bar{\bar{x}})^2}{\sigma^2}}$$

kjer so: $p_i = n_i/N$; n_i – velikost vzorca i ; N – celokupna velikost vzorca; \bar{x}_i – povprečje vzorca i ; $\bar{\bar{x}}$ – povprečje celokupnega vzorca; σ^2 – varianca. Vrednosti indeksa $f \approx 0,1, 0,25$ in $0,4$ predstavljajo majhno (šibko), srednje in veliko (močno) velikost učinka.

18.2.2 Mere povezanosti

Mere povezanosti temeljijo na deležu razložene variance, najpogosteje jo izrazimo s **koeficientom determinacije R^2** . Pri linearni regresijski analizo ga izračunamo s kvadriranjem korelacijskega koeficienta r :

$$R^2 = r^2$$

S koeficientom determinacije R^2 podamo delež razložene variance, oziroma kolikšen delež variance odvisne spremenljivke je razložen z variabilnostjo neodvisne spremenljivke.

Velikost učinka za linearni regresijski model podamo z **indeksom f^2** , izračunamo ga po enačbi:

$$f^2 = \frac{R^2}{1 - R^2}$$

Za izračun velikost učinka obstaja več spletnih aplikacij npr. na spletni strani Psychometrica https://www.psychometrica.de/effect_size.html

18.3 Analiza moči

Analiza moči (ang. Power Analysis) je del statistične analize v raziskovalnem delu, ki jo izvedemo bodisi v začetni stopnji načrtovanja *a priori* bodisi v zaključni fazi vrednotenja *a posteriori*. Z analizo moči vzpostavimo relacijo med vzorčenjem, statističnim sklepanjem in zaključki. Z analizo moči iščemo odgovore na dve temeljni metodološki vprašanji: Kako velik naj bo vzorec? in Kolikšno moč ima statistični test? Statistične parametre računamo na vzorcu, pridobljenem iz populacije. Zaradi napake vzorčenja, statistični izračun odstopa od dejanske vrednosti parametra populacije. Več kot je vzorčenih enot bolj natančna je ocena parametra. Analizo moči naredimo tudi po raziskavi (ang. Retrospective Power Analysis) in z njo na podlagi velikosti vzorca in velikosti učinka ocenimo zanesljivost izračuna. Na analizo moči gledamo predvsem s stališča velikosti vzorca, v pogojih omejene velikosti vzorca ocenimo verjetnost zaznave učinka ob dani stopnji zaupanja. Če je verjetnost za doseganje velikosti učinka nedosegljivo nizka, je na mestu premislek o nadaljevanju eksperimenta oz. vlaganju energije v vzorčenje. V biodiverzitetnih raziskavah smo pogosto omejeni z velikostjo vzorca, omejitve lahko izhajajo iz zahtevnosti vzorčenja, nedostopnosti vzorcev ali etične nesprejemljivosti. V takih primerih moramo sprejeti kompromisno odločitev med še sprejemljivo napako vzorčenja in energijo namenjeno dodatnemu vzorčenju. Pri tej odločitvi si pomagamo z analizo moči, ključni element s katerim lahko vplivamo na moč testa, je velikost vzorca. Na moč statističnega testa vplivajo tri vrednosti: **velikost vzorca, napaka prve vrste** in **velikost učinka**.

Moč statističnega testa pomeni zmožnost prepoznati učinek, kadar je ta prisoten, vezan je na napako druge vrste (β). Splošna statistična praksa je, da je mejna vrednost za napako prve vrste predpisana pri $\alpha = 0,05$ (5 % tveganje za zavrnitev ničelne hipoteze). Pri tem privzamemo, da je napaka prve vrste štiri krat strožja, kot je napaka druge vrste. Z drugimi besedami, napačna zavrnitev ničelne hipoteze, kadar je pravilna, je štiri krat hujša napaka, kot njeno napačno sprejetje, kadar ni pravilna. Mejna vrednost za napako druge vrste je predpisana pri $\beta = 0,20$. Iz vrednosti β lahko izračunamo moč testa, ki jo označimo s P :

$$P = 1 - \beta$$

Moč testa mora presegati 0,80 ($1 - 0,20 = 0,80$), da je test sprejemljiv, to pomeni, da obstaja vsaj 80 % verjetnost za sprejem alternativne hipoteze.



PRIMER 18.2

Velikost učinka, moč testa, velikost vzorca in stopnja značilnosti – odvisna vzorca

Z definiranjem treh vrednosti od štirih: velikost učinka d , moč testa P , velikost vzorca n ali stopnjo značilnosti α , izračunamo četrto. Primer izračuna posameznega parametra v primerjanju dveh odvisnih vzorcev z dvosmernim t -testom. Izračuni so narejeni s paketom »pwr« Champely (2020).

1. Izračun moči testa P , pri testiranju povprečij dveh odvisnih vzorcev s t -testom, P je funkcija n (je število parov), d in α .

Rezultat:

- $n = 30$
- $d = 0,8$
- $\alpha = 0,05$
- **$P = 0,861$**

```
R > package = »pwr«
t.out<-pwr.t.test(n = 30, d = 0.8, sig.level = 0.05)
t.out
plot(t.out)
```

Razlaga. Pri preverjanju razlike med odvisnima vzorcema, v katerem smo preverjali 30 parov, smo predpisali stopnjo značilnosti za napako prve vrste pri 5 % ($\alpha = 0,05$) in velikost učinka zelo močno 0,8, je izračunana moč testa zadostna 0,861 oz. značilna pri $\beta = 0,139$ (mejna vrednost $\beta = 0,2$).

2. Izračun velikosti učinka d , pri testiranju povprečij dveh odvisnih vzorcev s t -testom, d je funkcija n (je število parov), P in α .

Rezultat:

- $n = 30$
- **$d = 0,736$**
- $\alpha = 0,05$
- $P = 0,8$

```
R > package = »pwr«
t.out<-pwr.t.test(p = 0.8, n = 30, sig.level = 0.05)
t.out
```

Razlaga. Pri preverjanju razlike med odvisnima vzorcema, v katerem smo preverjali 30 parov, smo predpisali stopnjo značilnosti za napako prve vrste pri 5 % ($\alpha = 0,05$) in moč testa 0,8 oz. stopnjo značilnosti za napako druge vrste pri $\beta = 0,2$, izračunana velikost učinka je močna do zelo močno 0,736.

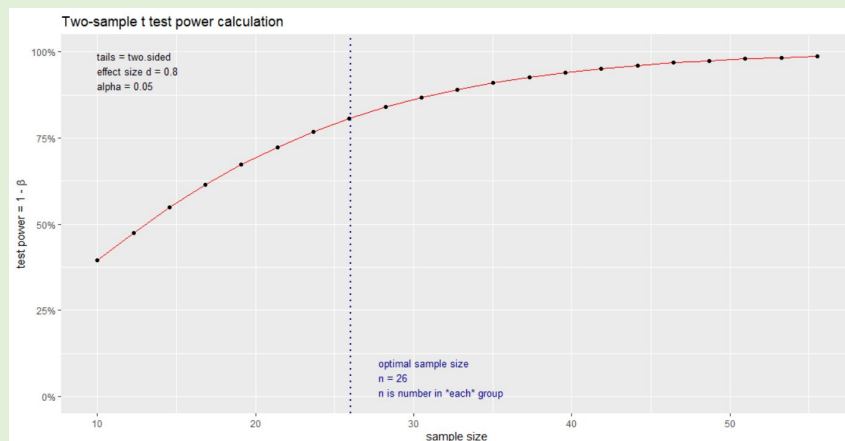
3. Izračun velikosti vzorca n (je število parov) pri testiranju povprečij dveh odvisnih vzorcev s t -testom, n je funkcija d , P in α .

Rezultat:

- $n = 25,5$
- $d = 0,8$
- $\alpha = 0,05$
- $P = 0,8$

```
R > package = »pwr«
t.out<-pwr.t.test(p = 0.8, d = 0.8, sig.level = 0.05)
t.out
```

Razlaga. Pri preverjanju razlike med odvisnima vzorcema, za katere smo predpisali stopnjo značilnosti $\alpha = 0,05$ in $\beta = 0,2$ ter velikost učinka zelo močno 0,8, moramo vključiti vsaj 26 parov ($n = 25,5$) (Slika 18.3).



Slika 18.3 Odnos med velikostjo vzorca in močjo testa ob definirani velikosti učinka in stopnji značilnosti.





PRIMER 18.3

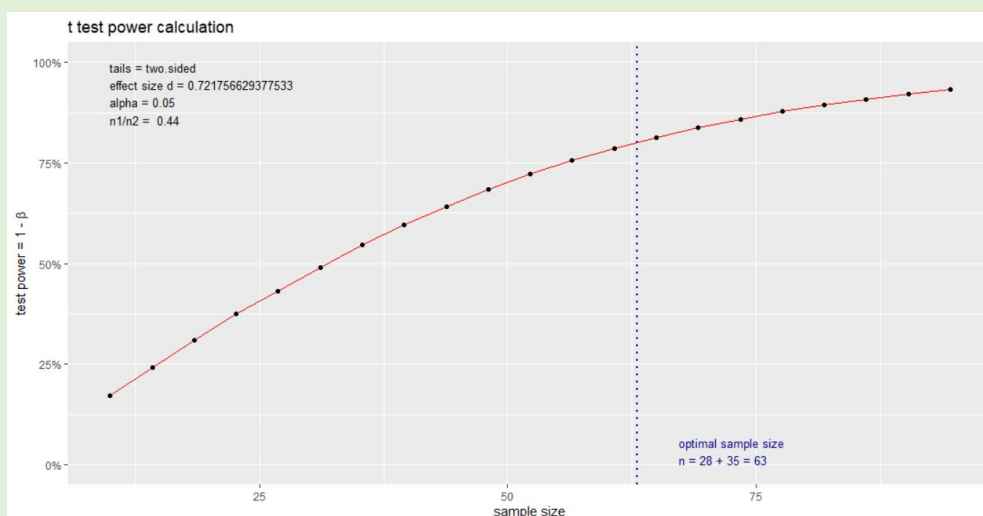
Velikost učinka, moč testa, velikost vzorca in stopnja značilnosti – neodvisna vzorca

Primer izračuna velikosti učinka d pri primerjanju dveh neodvisnih vzorcev z velikostjo vzorca n_1 in n_2 s t -testom.

Rezultat:

- $n_1 = 28$
- $n_2 = 35$
- **$d = 0,722$**
- $\alpha = 0,05$
- $P = 0,8$

Razlaga. Pri preverjanju razlike med neodvisnima vzorcema s številom meritev v prvem vzorcu $n_1 = 28$ in v drugem vzorcu $n_2 = 35$, za katere smo predpisali stopnjo značilnosti $\alpha = 0,05$ in $\beta = 0,2$, je izračunana velikost učinka močna $d = 0,722$ (Slika 18.4).



Slika 18.4 Odnos med velikostjo vzorca in močjo testa ob definirani velikosti učinka in stopnji značilnosti.

```
R > package = »pwr«
tt.out<-pwr.t2n.test(n1 = 28, n2 = 35, p = 0.8, sig.level = 0.05)
tt.out
plot(tt.out)
```



Zaključki

- V znanosti je splošno sprejet koncept statističnega sklepanja s testiranjem ničelne hipoteze. V procesu odločanja zastavimo dve nasprotujoči si trditvi, ničelno hipotezo H_0 , po kateri ne obstaja razlika med dvema populacijama, in alternativno hipotezo H_a , po kateri obstaja razlika med populacijama.
- Pri zastavitvi hipotez ne vemo, katera od hipotez je pravilna in katera predstavlja dejansko stanje v naravi. S pomočjo verjetnosti in statističnega sklepanja poskušamo izluščiti pravilno trditev.
- Pravilo, po katerem na podlagi vzorca ničelno hipotezo H_0 sprejmemo ali ne sprejmemo, se imenuje statistični test.
- Pravilo statističnega testa ne velja vedno, zaradi slučajnega vzorca se zgodi, da sprejmemo napačno hipotezo. Kadar zavrremo pravilno ničelno hipotezo H_0 , smo naredili napako prve vrste. Kadar sprejmemo napačno ničelno hipotezo H_0 , smo naredili napako druge vrste.
- Obe zagrešeni napaki sta v procesu statistične analize nezaželeni, vendar se jima ne moremo v celoti izogniti, zato si prizadevamo, da sta kar se da malo verjetni. Za napako prve vrste predpišemo zgornjo mejo verjetnosti α , ki jo imenujemo stopnja značilnosti in je praviloma pri $\alpha = 0,05$.
- Pri sprejemanju sestavljene hipoteze preverjamo razliko samo za en statistični parameter, npr. povprečje. Izračunamo lahko samo verjetnost za napako prve vrste, napake druge vrste ne moremo določiti. Ničelno hipotezo zavrremo pri $p < 0,05$ ali pa o ničelni hipotezi ne odločimo pri $p > 0,05$. Na podlagi tega pravila se izognemo napaki druge vrste.
- Pri sprejemanju enostavne hipoteze sta znana oba parametra, tako za H_0 kot za H_a . Glede na predpisano stopnjo značilnosti za napako prve vrste $\alpha = 0,05$, ničelno hipotezo zavrremo pri $p < 0,05$ in sprejmemo alternativno hipotezo ali pa glede na $p > 0,05$ in $\beta < 0,2$ sprejmemo ničelno hipotezo. Z β označimo verjetnost za sprejetje ničelne hipoteze, ko je ta napačna. β je povezana z močjo testa P , to je verjetnost za zavrnitev ničelne hipoteze, ko je ta napačna $P = 1 - \beta$.
- Odnosi med verjetnostjo za napako prve vrste, ki jo podamo z α , verjetnostjo za napako druge vrste, ki jo podamo z β in predstavlja moč testa, velikostjo učinka in velikostjo vzorca so taki, da je vsak od njih funkcija preostalih treh. Ko pridobimo tri kriterije, je s tem določen tudi četrti.
- Velikost učinka je vrednost, s katero podamo velikost razlike med ničelno in alternativno hipotezo, je merilo odstopanja od ničelne hipoteze in informacija za zanesljivost sprejema alternativne hipoteze.
- V biodiverzitetni praksi nas zanima predvsem odnos med stopnjo značilnosti α in velikostjo vzorca.

- Presoj o velikosti vzorca za oceno vrstnega bogastva naredimo s krivuljo kopičenja vrst, ki je funkcija vloženega dela v vzorčenju. Delo v vzorčenju merimo kot: število vzorčenj, število vzorčenih lokalitet, število registriranih osebkov, število ur, dni ali let vzorčenj. V primeru, ko akumulacijska krivulja vrst doseže asimptoto, je to pokazatelj, da smo vrstno bogastvo zadostno raziskali.
- V vrednotenje statističnih parametrov moramo vključiti zadostno število meritev, da zagotovimo reprezentativnost vzorca. Potrebno število meritev ocenimo s pomočjo velikosti učinka, močjo testa in stopnje značilnosti.

Literatura

- APA. 2012. Publication Manual of the American Psychological Association. American Psychological Association. Washington DC.
- APA. 2021. Apastyle. Pridobljeno dne 4. 5. 2021, iz: <https://apastyle.apa.org/>.
- Champely S. 2020. pwr: Basic Functions for Power Analysis. R package version 1.3-0, <https://CRAN.R-project.org/package=pwr>.
- Cohen J. 1988. Statistical power analysis for the behavioral sciences. Lawrence Erlbaum Associates Publisher. New York.
- Colwell R.K., Chao A., Gotelli N.J., Lin S., Mao C.X., Chazdon R.L., Longino J.T. 2012. Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *Journal of Plant Ecology*, 5(1): 3–21.
- Faul F., Erdfelder E., Buchner A., Lang A. G. 2009. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41: 1149–1160.
- Krebs C.J. 2014. *Ecological Methodology*. Harper Collins Publisher Inc. New York.
- Lenhard W., Lenhard A. 2016. Calculation of Effect Sizes. Retrieved from: https://www.psychometrica.de/effect_size.html. Dettelbach (Germany): Psychometrica. DOI: 10.13140/RG.2.2.17823.92329.
- Sokal R.J., Rohlf F.J. 2012. *Biometry*. W.H. Freeman Company. New York.



DODATKI

Dodatek 1

Pregled konceptov, vidikov, dimenzij in cenilk biodiverzitet

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

Poglavje 1. Hierarhični koncept biodiverzitete

| Hierarhični nivo | Opis |
|--------------------------|----------------------|
| Genetska diverziteta | Pestrost dednine |
| Vrstna diverziteta | Pestrost vrst |
| Ekosistemska diverziteta | Pestrost ekosistemov |

Poglavje 1. Vidiki in dimenzije biodiverzitete

| Vidik | Opis |
|--------------------------|---|
| Vrstna diverziteta | Pestrost vrst oz. taksonomskih kategorij |
| Funkcionalna diverziteta | Pestrost na nivoju fiziološke, morfološke, ekološke ali etološke organizacije/delovanja organizma |
| Filogenetska diverziteta | Pestrost filogenetskih linij |

| Dimenzija | Opis |
|-------------|---|
| Bogastvo | $q = 0$; Število vrst, znakov ali haplotipov |
| Pestrost | $q = 1$; Pestrost izražena z razmerjem med številom vrst, znakov ali haplotipov in njihovo abundanco |
| Poravnanosť | $q = 2$; Pestrost vrst, znakov ali haplotipov podana v enotah vrst |

Poglavje 3. Prostorska komponenta vrstnega bogastva, samo PA podatki

| Prostorska komponenta | Dimenzija |
|-----------------------------|---|
| Alfa diverziteta - α | Točkovna diverziteta, Lokalni nivo |
| Beta diverziteta - β | Primerjave med pari lokalitet, Vrstni obrat, Ugnezdenost. |
| Zeta diverziteta - ζ | Primerjave v sestavi nabora vrst med mnogo lokalitetami. |
| Gama diverziteta - γ | Nabor vrst, Regionalni nivo |

Poglavje 3. Beta diverziteta, samo PA podatki

| Nivo primerjav | Skupine cenilk vrstnega obrata | Indeks | Razlaga cenilke, Zaloga vrednosti 0-1 |
|---------------------------|--|--|---------------------------------------|
| Primerjave med pari | Mere kontinuitete in izgube | β_{gl} | 0 – različnost 1 – podobnost |
| | Mere gradienta vrstnega bogastva | β_{gl} | 0 – različnost 1 – podobnost |
| | Mere kontinuitete | β_i, β_{sor} | 0 – različnost 1 – podobnost |
| | | $\beta_w, \beta_{-1}, \beta_c, \beta_{wb}, \beta_l, \beta_t, \beta_{hk}, \beta_z, \beta_{me}, \beta_m$ | 0 – podobnost 1 – različnost |
| Mere naraščanja in izgube | $\beta_{co}, \beta_r, \beta_l, \beta_e, \beta_{rs}, \beta_{-2}, \beta_{-3}, \beta_{sim}$ | 0 – podobnost 1 – različnost | |
| Mnogotere primerjave | Prava β diverziteta | β_w | Povprečni obrat |
| | Sørensenov indeks mnogoterih raznolikosti | β_{SOR} | 0 – različnost 1 – podobnost |
| | Simpsonov indeks mnogoterih raznolikosti | β_{SIM} | 0 – podobnost 1 – različnost |
| | Ugnezdenost | β_{NES} | 0 – vrstni obrat 1 – ugnezdenost |

Poglavje 3. Zeta diverziteta

| Nivo | Oznaka | Opis vrednosti | Tip podatkov |
|------------------|---------|---|--------------|
| Zeta diverziteta | ζ | Mnogotera primerjava med vrstnim bogastvom lokalitet. | Prezenca |
| Zeta razmerje | ZR | Stopnja upadanja biodiverzitete. | Prezenca |

Poglavje 3. Temna diverzitete

| Nivo | Ime | Oznaka | Opis vrednosti | Tip podatkov |
|-------------------|---------------------------|-----------------|--|--------------|
| Temna diverziteta | Temna diverziteta | DD _p | Razlika med naborom vrst krajine SP in vrstnim bogastvom lokalitete FC. | Prezenca |
| | Indeks popolnosti združbe | CCI | Kvocient med bogastvom lokalitete in naborom vrst krajine. | Prezenca |
| | Bealov indeks | b _{ij} | Verjetnost za prisotnost vrste na lokaliteti; 0 – odsotna 1 – prisotna | Prezenca |

Poglavje 4. Cenilke vrstne diverzitete

| Nivo | Ime | Oznaka | Opis vrednosti | Tip podatkov |
|------------------------------|--------------------------|--|---|--------------|
| Vrstno bogastvo | Vrstno bogastvo | S | Število vrst. | PA |
| Vrstna pestrost | Indeksi vrstne pestrosti | R1, R2, R3 | Razmerje med številom vrst in deleži posameznih vrst. | Abundance |
| | | H' | | |
| | | λ, 1-D, 1/D | | |
| | Vrstna poravnanos | E ₁ , E ₂ , E ₃ , E ₄ , E ₅ | Cenilke v indeksu (0-1). | Abundance |
| Hillova diverzitetna števila | S, e ^{H'} , 1/D | Cenilke v enotah vrst. | Abundance | |

Poglavje 4. Ocena vrstnega bogastva

| Tip | Cenilka | Kriterij | Tip podatkov |
|-----------|------------|------------------------------------|--------------|
| Incidenca | Chao2 | Edinstvene (=1) in dvojne (=2) | PA |
| | ICE | Pogoste (>10) in redke vrste (<10) | |
| | Jackknife1 | Edinstvene (=1) | |
| | Jackknife2 | Edinstvene (=1) in dvojne (=2) | |
| | Bootstrap | Edinstvene (=1) | |
| Abundanca | Chao1 | Enojci (=1) in dvojci (=2) | Abundanca |
| | ACE | Pogoste (>10) in redke vrste (<10) | |

Poglavje 4. Relativna vrstna abundanca

| Pristop | Porazdelitev | Podatki o abundancah vrst |
|-------------|-----------------------|---|
| Fisherjev | Logaritemska | Surovi |
| Prestonov | Logaritemsko normalna | Organizirani v razrede, oz. v oktave z abundancami: 0, 1, 2-3, 4-7, 8-15, 16-31, 32-63, 64-127 ... |
| Whitakerjev | Naključna | Rangirani od visokih k nizkim vrednostim. |
| | Logaritemska | |
| | Logaritemsko normalna | |
| | Zipf | |
| | Zipf-Mandelbrot | |

Poglavje 5. Funkcionalna pestrost – univariatni pristop

| Cenilka | Kriterij | Tip podatkov |
|-----------|--|--------------|
| Povprečje | Zvezna spremenljivka | PA |
| Mediana | Ordinalna spremenljivka | |
| Frekvenca | Kategorična spremenljivka | |
| CWM | Zvezna, ordinalna, binarna spremenljivka | Abundanca |

Poglavje 5. Funkcionalna pestrost – multivariatni pristop

| Komponenta | Cenilka | Opis | Tip podatkov |
|--------------------------|-------------------------------------|--|-----------------------------------|
| Funkcionalno bogastvo | Funkcionalno bogastvo, FRic | Volumen poligona. | Meritve znakov, glavne koordinate |
| Funkcionalna raznolikost | Funkcionalno razhajanje, FDiv | Odstopanje vrste od centroida uteženo z relativno abundanco. | Meritve znakov, glavne koordinate |
| | Funkcionalna razpršenost, FDis | Povprečna razdalja od centroida. | Matrika razdalj |
| | Raova kvadratna entropija, Q | Vsota razdalj med pari vrst, uteženo z relativno abundanco. | Matrika razdalj |
| | Funkcionalna specializiranost, FSpe | Odstopanje vrste od centroida uteženo z relativno abundanco. | Matrika razdalj |
| Funkcionalna poravnost | Funkcionalna poravnost, FEve | Vsota razdalj MST, utežene z relativno abundanco. | Meritve znakov, glavne koordinate |

Poglavje 6. Filogenetska pestrost

| Komponenta | Cenilka | Opis | Tip podatkov |
|--------------------------|--|--|---|
| Filogenetsko bogastvo | Filogenetska diverziteta PD | Ocena alfa diverzitete, vsota dolžin vej. | Simetrična matrika razdalj in tipologija drevesa. |
| | Filogenetska diverziteta obtežena z abundanco ΔnPD , po Barkerju | Ocena alfa diverzitete, vsota dolžin vej utežena z abundanco. | Simetrična matrika razdalj in tipologija drevesa. |
| | Filogenetski endemizem PE | Ocena alfa diverzitete, vsota dolžin vej. | Simetrična matrika razdalj in tipologija drevesa. |
| | Filogenetsko vrstno bogastvo PSR | Ocena alfa diverzitete, filogenetska razdalja med paroma. | Simetrična matrika razdalj in tipologija drevesa. |
| | Vsota evlucijskih posebnosti ED | Ocena alfa diverzitete, filogenetska izolacija taksonov. | Simetrična matrika razdalj in tipologija drevesa. |
| | Unifrac | Ocena beta diverzitete med nabori vrst, delež skupnih vej. | Simetrična matrika razdalj in tipologija drevesa. |
| Filogenetska raznolikost | Povprečna filogenetska diverziteta avPD | Ocena alfa diverzitete, relativna dolžina vej. | Simetrična matrika razdalj in tipologija drevesa. |
| | Povprečna filogenetska diverziteta obtežena z abundanco avPD _{AB} | Ocena alfa diverzitete, relativna dolžina vej utežena z abundanco. | Simetrična matrika razdalj in tipologija drevesa. |
| | Povprečna razdalja najbližjega taksona MNTD | Ocena alfa diverzitete, razdalja med paroma. | Simetrična matrika razdalj in tipologija drevesa. |
| | Povprečna evlucijska posebnost mean(ED) | Ocena alfa diverzitete, filogenetska izolacija. | Simetrična matrika razdalj in tipologija drevesa. |
| | Povprečna filogenetska razdalja Dpw | Ocena beta diverzitete med nabori vrst, relativna dolžina vej. | Simetrična matrika razdalj in tipologija drevesa. |
| Filogenetska poravnost | Collesov indeks I _C | Ocena alfa diverzitete, topologija drevesa. | Simetrična matrika razdalj in tipologija drevesa. |
| | Varianca razdalj med pari VPD | Ocena alfa diverzitete, razdalja med paroma. | Simetrična matrika razdalj in tipologija drevesa. |
| | Varianca evlucijske posebnosti var(ED) | Ocena alfa diverzitete, filogenetska izolacija. | Simetrična matrika razdalj in tipologija drevesa. |

Poglavje 7. Analiza združbe, asimetrične mere primerjav med dvema združbama

| Tip | Tip koeficientov | Cenilka | Tip podatkov | |
|-----------|------------------|--|-------------------------------|----|
| Q | Podobnost | Koeficient enostavnega ujemanja – S_1 | PA | |
| | | Jaccardov koeficient – S_7 | | |
| | | Sørensenov koeficient – S_8 | Kvantitativni | |
| | | Steinhausov koeficient – S_{17} | | |
| | | Kulczyński koeficient – S_{18} | | |
| | | Modificiran Gowerjev koeficient – S_{19} | Kvantitativni, PA, mešani | |
| | | Hi-kvadrat koeficient podobnosti – S_{21} | Kvantitativni | |
| | Razdalja | Tetivna razdalja – D_3 | Kvantitativni | |
| | | Whittakerjev indeks asociacije – D_9 | | |
| | | Canberra razdalja – D_{10} | | |
| | | Clarkova razdalja – D_{11} | | |
| | | Bray-Curtis razdalja – D_{14} | | |
| | | Hi-kvadrat X^2 – D_{15} | | |
| | | Hi-kvadrat X^2 razdalja – D_{16} | | |
| | | Hellingerjeva razdalja – D_{17} | | |
| | | Razdalja med vrstnimi profili – D_{18} | | |
| | | Modificirani koeficient povprečne razlike – D_{19} | | |
| | R | Podobnost | Jaccardov koeficient – S_7 | PA |
| | | | Sørensenov koeficient – S_8 | |
| Podobnost | | Pearsonov r | Abundanca | |
| | | Spearmanov r | | |

Poglavje 7. Transformacije podatkov

| Tip transformacije | Korekcija | Tip podatkov |
|------------------------------|--------------------|-----------------------|
| Korenjenje | Asimetrija v levo | Abundanca, intervalni |
| Logaritmiranje | Asimetrija v desno | Abundanca, intervalni |
| Tetivna transformacija | Absolutne razlike | Abundanca |
| Hellingerjeva transformacija | Absolutne razlike | Abundanca |
| χ^2 standardizacija | Absolutne razlike | Intervalni |
| Normiranje | Absolutne razlike | Intervalni |

Poglavje 7, 11, 12. Multivariatna obravnava

| Družina | Metoda | Namen | Tip podatkov |
|-------------------|-------------------------------|---|--|
| Ordinacija | Analiza glavnih komponent | Redukcija razsežnosti | Intervalni |
| | Diskriminantna analiza | Redukcija razsežnosti, Klasifikacija objektov | Intervalni |
| | Multidimenzionalno skaliranje | Redukcija razsežnosti | Abundanca |
| | Analiza glavnih koordinat | Redukcija razsežnosti | Mešani |
| Klastering | Združevanje v skupne | Minimalna metoda | Združevanje na osnovi podobnosti ali različnosti |
| | | Maksimalna metoda | |
| | | Metoda povprečja | |
| | | Centroidna metoda | |
| | Wardova metoda | | |
| Cepitev v skupine | Metoda voditeljev | Razporejanje v skupine | Abundanca, Intervalni |

Poglavje 13 in 14. Filogenetske rekonstrukcije

| Pristop | Metodološki pristop | Cenilke | Podatki |
|--------------|---|---|--|
| Fenetski | Tradicionalna morfometrija Geometrijska morfometrija | Združevanje sosedov UPGMA | Nominalne in zvezne morfološke spremenljivke. |
| Kladistični | Monofiletske skupine | Maksimalna varčnost | Nominalne morfološke in molekularne spremenljivke. |
| Filogenetski | Substitucijski modeli | Združevanje sosedov UPGMA Maksimalna verjetje Bayesovo sklepanje | Molekularne spremenljivke. |

Poglavje 13. Fenetika

| Pristop | Metodološki pristop | Cenilke | Podatki |
|----------------------------|--------------------------------|---|---|
| Tradicionalna morfometrija | Kvalitativni morfološki znaki | Jaccardov koeficient Koeficien enostavnega ujemanja Sokal & Sneath S5 koeficient Roger & Tanimoto koeficient Dice in Sorensen koeficient Humannov koeficient Ochiai koeficient Sokal & Sneath S13 koeficient Pearsonov Phi koeficient Gower & Legendre S2 koeficient | Nominalni znaki, Binarni znaki, Kategorični znaki |
| | Mešani tip znakov | Gowerjev splošni koeficient | Nominalni in zvezni znaki |
| | Kvantitativni morfološki znaki | Evklidova razdalja Bray-Curtis razdalja Manhattanova razdalja Mahalanobisova razdalja | Zvezni znaki (enorazsežni) |
| | Fenetsko drevo | UPGMA Združevanje sosedov | Morfometrične razdalje |
| Geometrijska morfometrija | Prokrustova analiza | Prokrustova razdalja | Zvezni znaki (dvorazsežni ali trirazsežni) |
| | Fenetsko drevo | UPGMA Združevanje sosedov | Morfometrične razdalje |

Poglavje 13. Kladistika

| Pristop | Metodološki pristop | Cenilke | Podatki |
|-----------------------|---|---------------------|---|
| Zadnji skupni prednik | Monofiletske skupine Sinapomorfije Homologije | Maksimalna varčnost | Nominalne morfološke in molekularne spremenljivke |

Poglavje 14. Filogenetika

| Pristop | Metodološki pristop | Cenilke | Podatki |
|--------------------------------------|---|--|-----------------------|
| Genetske razdalje | Nukleotidna p-razdalja Jukes-Cantorjeva razdalja Kimura 2 razdalja Tajima-Nei razdalja Tamura 3 razdalja Tamura-Nei razdalja | UPGMA Združevanje sosedov | Nukleotidno zaporedje |
| Modeliranje nukleotidnih substitucij | Jukes-Cantorjev model Kimura 2 model Tajima-Nei model Tamura 3 model Tamura-Nei model | Metoda največje verjetja Bayseovo sklepanje | Nukleotidno zaporedje |

Poglavje 14. Vrednotenje filogenetskih dreves

| Pristop | Metodološki pristop | Cenilke | Podatki |
|----------------|--------------------------------------|--------------------------------------|----------------|
| Samovzorčenje | Bootstrap Jackknife | Delež podpore posamezni razvejitvi | Newick drevesa |
| Drevo soglasja | Strogo soglasje Večinsko soglasje | Razrešeno drevo Nerazrešeno drevo | Newick drevesa |

Poglavje 17. Regresija

| Družina | Metoda | Namen | Tip podatkov |
|----------------------|--|--|--|
| Bivariatni modeli | Enostavna linearna regresija | Model I Metoda navadnih najmanjših kvadratov | Odnos med odvisno in neodvisno variabla. |
| | | Model II | Metoda glavne osi |
| | Metoda standardne glavne osi | | Intervalni |
| | Metoda rangirane glavne osi | | |
| Multivariatni modeli | Multipla linearna regresija | Razlaga variabilnosti odvisne variable z več neodvisnimi variablami. | Intervalni |
| | Polinomska regresija | Razlaga variabilnosti odvisne variable s polinomom neodvisne variable. | Intervalni |
| | Mantel test, Odnos med matrikami razdalj | Jakost povezave med diagonalnima matrikama. | Intervalni, diagonalna matrika |

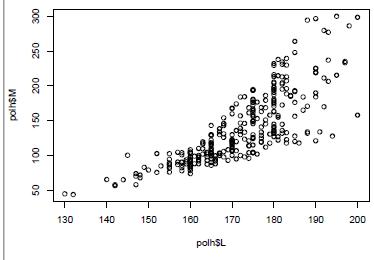
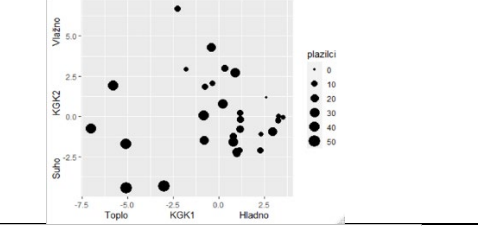
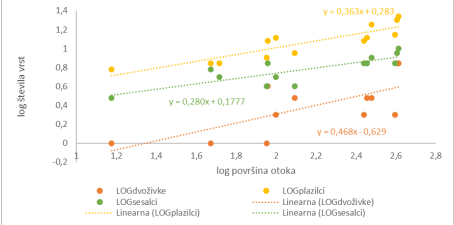
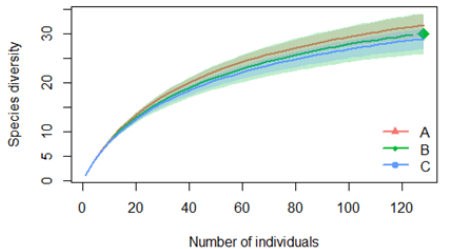
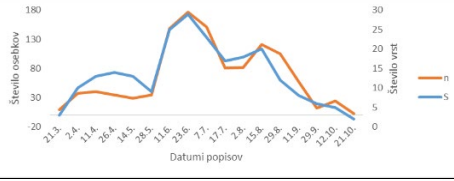
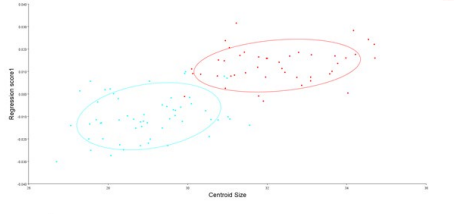
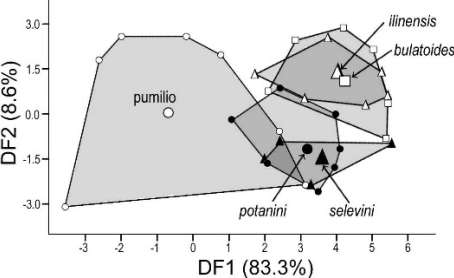
Poglavje 17. Izbor modela v multipli regresiji

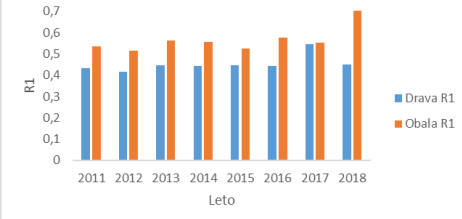
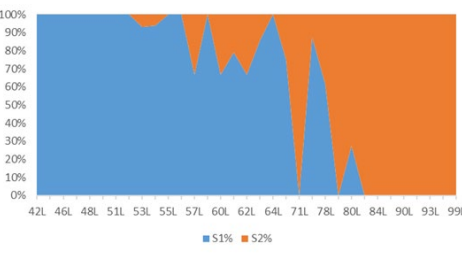
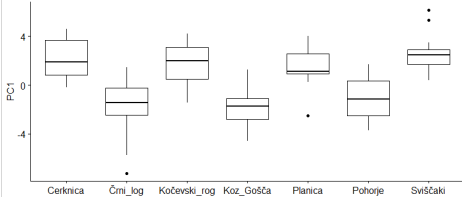
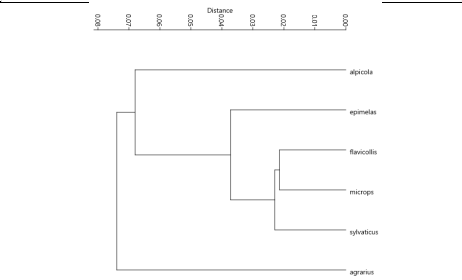
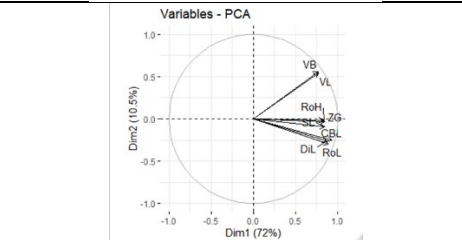
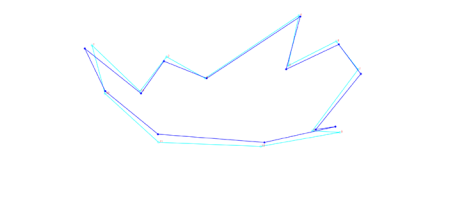
| Pristop | Strategija | Cenilke |
|------------------------------------|--------------------------|--|
| Testiranje ničelne hipoteze | Vključevanje prvih | p , R^2 , R^2_{adj} , VIF |
| | Izključevanje zadnjih | p , R^2 , R^2_{adj} , VIF |
| | Stopenjski model | p , R^2 , R^2_{adj} , VIF |
| Informacijsko – teoretični pristop | Izbor najboljših modelov | R^2_{adj} , AIC, AIC _c , BIC, C_p |

Dodatek 2

Grafične predstavitve statističnih lastnosti

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

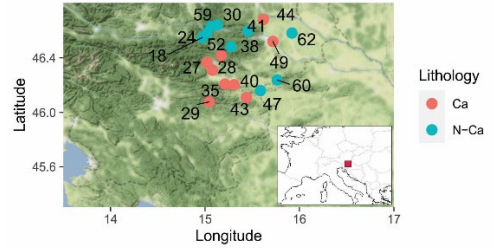
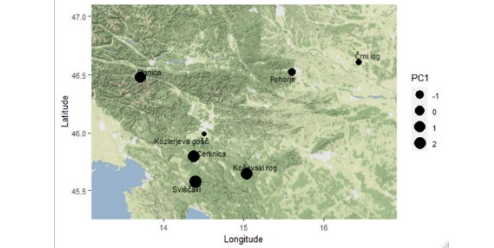

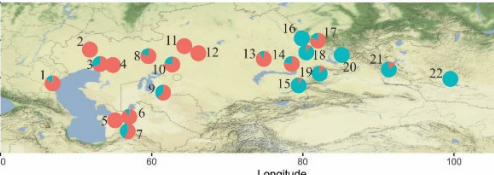
| Vrsta grafa | Opis | Slika |
|--|---|--|
| Razsevni diagram (ang. Biplot, Scatterplot) | Projekcija točk na dve spremenljivki. |  |
| Razsevni diagram, kombiniran z velikostjo točk (ang. Dot Plot) | Projekcija točk na dve spremenljivki, točke poleg položaja vsebujejo dodatno informacijo o velikosti količine. |  |
| Razsevni diagram, kombiniran z regresijsko premico | Projekcija točk na dve spremenljivki z regresijskimi premicami. |  |
| Razsevni diagram, kombiniran z regresijsko krivuljo | Projekcija točk na dve spremenljivki z glajenimi regresijskimi krivuljami. |  |
| Razsevni diagram, kombiniran z ogivo | Projekcija točk na dve spremenljivki, zaporedje točk povezane z ogivo (lomljena črta, ki povezuje točke). |  |
| Razsevni diagram, kombiniran s elipsoidom | Projekcija točk na dve spremenljivki, posamezna skupina je povezana s 95 % elipsoidom. |  |
| Razsevni diagram, kombiniran z minimalnim konveksnim poligonom | Projekcija točk na dve spremenljivki, zunanje točke posamezne skupine so povezane z lomljeno črto; centroid skupine je predstavljen z znakom. |  |

| Vrsta grafa | Opis | Slika |
|-------------------------------|--|--|
| Stolpčni diagram, histogram | S stolpci podala lastnost spremenljivke po kategorijah za dve skupini. |  |
| Ploščinski graf | S ploščino podana lastnost po kategorijah za dve skupini. |  |
| Okvir z ročaj (ang. Box Plot) | Grafična predstavitev opisne statistike (aritmetična sredina, standardna deviacija, minimum, maksimum, osamelci) za eno spremenljivko po skupinah. |  |
| Dendrogram | Podobnost med objekti prikazana v obliki drevesa. |  |
| Vektorji na krožnici | Projekcija korelacijskih vektorjev več odvisnih spremenljivk z dvema neodvisnima spremenljivkama. |  |
| Žični graf | Projekcija točk na dve spremenljivki, ki opisujejo morfološko strukturo; zaporedje točk je povezano z lomljeno črto; svetlo modra je povprečje, temno modra je lastnost fokusne skupine. |  |

Dodatek 3

Kartogrami

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

| Vrsta grafa | Opis | Slika |
|---|---|---|
| Simbolna karta | Zemljevid s točkami ena na mnogo, ki predstavlja lokalitete v prostoru, barva simbola predstavlja njegovo lastnost. |  |
| Simbolna karta | Zemljevid s točkami ena na mnogo, ki predstavlja lokalitete v prostoru, velikost simbola predstavlja njegovo lastnost. |  |
| Koropletna karta s točkami in poligonom | Zemljevid kombiniran s koropletno karto, ki predstavlja areal vrste in točkami, ki predstavljajo konkretne lokalitete. |  |
| Kartodiagram | Zemljevid s tortnimi diagrami, položaj diagrama predstavlja lokaliteto, vrednosti v diagramu pa razmerje v lastnostih pojava. |  |

Dodatek 4

Programska orodja

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

Programska orodja

V učbeniku je navedenih 115 Primerov s katerimi dajemo zgled za pripravo podatkov, analitske postopke, predstavitev rezultatov v obliki besedila, tabel in grafikonov ter vrednotenje rezultatov in oblikovanje zaključkov. Podatke o biodiverziteti smo zbirali in urejali v obliki seznamov, vektorjev in matrik v obliki Excelovih preglednic. Več preglednic smo med seboj povezali v relacijsko matriko v Accessu. Grafične izdelke v katerih smo kombinirali grafe, fotografije in tekst smo urejali v CorelDRAW. Končno obliko besedila, ki ga dopolnjujejo preglednice, diagrami in slike smo izdelali v Wordu.

Statistične izračune in njim pripadajoče grafične priloge smo izvajali z raznoliko programsko opremo. V nadaljevanju so navedeni vsi uporabljeni programi in njihovo kratka predstavitev. Z izjemo komercialnih programov CorelDRAW, MS Word, MS Access in MS Excel je vsa programska oprema namenjena statističnemu vrednotenju prosto dostopna na spletu.

Komercialni programi:

CorelDRAW – Program CorelDRAW je namenjen izdelavi in urejanju izdelkov v vektorski grafiki (vse oblike grafičnih in slikovnih zapisov).

Microsoft Access – MS Access je namenjen izdelavi in upravljanju baz podatkov.

Microsoft Excel – V MS Excel je namenjen izdelavi preglednic, omogoča manipulacijo s podatki (razvrščanje, transponiranje, filtriranje ipd.), izdelavo izračunov, statistične izračune in izris grafov.

Microsoft Word – MS Word je namenjen urejanju besedil.

Prostodostopni programi za statistično vrednotenje:

BEAST – BEAST je platforma s programi za Bayesovo analizo molekularnih zaporedij z uporabo metode markovskih verig. Uporabili smo programe:

BEAST *Bayesian Evolutionary Analysis Sampling Trees* je program za vzorčenje Bayesove posteriorne verjetnosti in izvedbo meritev z metodo Monte Carlo z markovskimi verigami.

BEAUti *Bayesian Evolutionary Analysis Utility* je podporni program za uvoz podatkov in načrtovanje parametrov analize.

TreeAnnotator je program namenjen izbiri tarčnega drevesa.

FigTree je program za grafični prikaz in tehnično ureditev drevesa.

EstimateS – Program EstimateS – *Statistical Estimation of Species Richness and Shared species from Samples* je namenjen statistični analizi biodiverzitetnih podatkov in izračunu biodiverzitetnih cenilk.

MAFFT 7 – Program MAFFT – *Multiple alignment program for amino acid or nucleotide sequences* je aplikacija za poravnavo nukleotidnih sekvenc in njihovo analizo.

MEGA 11 – Program MEGA – *Molecular Evolutionary Genetics Analysis* je program za analizo DNK in proteinskih sekvenc.

MorphoJ – Program MorphoJ je namenjen statističnim analizam in grafikam v geometrijski morfometriji.

PAST4 – Program PAST – *Paleontological STatistics* je program za izvajanje uni- in multivariatnih statističnih analiz in izdelavo grafov.

Tps – *Thin-plate spline* je družina programov za zajemanje meritev, manipulacijo, statistično analizo in grafično predstavitev rezultatov v geometrijski morfometriji. Uporabili smo programe:

tpsDIG program je namenjen zajemanju meritev in urejanju meritev;

tpsRelw program je namenjen Prokrustovi analizi;

tpsUtil program je namenjen manipulaciji s podatki.

R, RStudio in Packages

R *The R Project for Statistical Computing* – Projekt R je prostodostopno programsko okolje za statistično in grafično vrednotenje.

RStudio – *RStudio* je delovno okolje in integralni del R.

Packages – Paketi so dodatki k RStudio in so namenjeni reševanju posameznih nalog. Pri statističnem vrednotenju, izrisu grafov in slik smo uporabili naslednje pakete:

»**ade4**«: *Exploratory and Euclidean Methods in Environmental Sciences Title Analysis of Ecological Data*. Paket »ade4« vsebuje orodja za multivariatno analizo podatkov.

»**betapart**«: *Partitioning Beta Diversity into Turnover and Nestedness Components*. Paket »betapart« vsebuje funkcije za izračun vrstne, funkcionalne in filogenetske beta diverzitete. Iz matrike razdalj izračuna parne in mnogotere cenilke vrstnega obrata in ugnezdenosti.

»**dplyr**«: *A Grammar of Data Manipulation*. Paket »dplyr« je namenjen urejanju in manipuliranju podatkov.

»**ecodist**«: *Dissimilarity-Based Functions for Ecological Analysis*. Paket »ecodist« omogoča izračune razdalj med skupinami, vključuje multivariatne ordinacijske metode in Mantel test.

»**factoextra**«: *Extract and Visualize the Results of Multivariate Data Analyses*. Paket »factoextra« je namenjen vizualizaciji rezultatov multivariatnih metod, npr. analize glavnih komponent in drugih.

»**ggmap**«: *Spatial Visualization with ggplot2*. Paket »ggmap« omogoča vizualizacijo prostorskih podatkov in modelov in dostopa do spletnih podatkovnih baz.

»**ggplot2**«: *Create Elegant Data Visualisations Using the Grammar of Graphics*. Paket »ggplot2« je namenjen izdelavi grafov in združevanju grafike in besedila.

»**FD**«: *Measuring functional diversity (FD) from multiple traits, and other tools for functional ecology*. Paket »FD« omogoča izračun različnih multidimenzionalnih cenilk funkcionalne diverzitete.

»**iNEXT**«: *Interpolation and Extrapolation for Species Diversity*. Paket »iNEXT« omogoča: 1. izračunavanje in izdelavo grafov redčenja (rarefaction) in ekstrapolacije

vrstnega bogastva v obliki Hillovih števil; 2. standardizacijo glede na velikost vzorca in glede na vzorčno pokritost; vsi izračuni so lahko na podatkih o abundancah ali prezencah.

»**lmodel2**«: *Model II Regression*. Paket »lmodel2« omogoča izračune enostavne linearne regresije po modelu II.

»**maps**«: *Draw Geographical Maps*. Paket »maps« je namenjen izdelavi zemljevidov.

»**MASS**«: *Support Functions and Datasets for Venables and Ripley's MASS*. Paket »MASS« je orodje in podatkovna zbirka, ki podpira učbenik statistike avtorjev W.N. Venables in B.D. Ripley »Modern Applied Statistics with S«.

»**MuMIn**«: *Multi-Model Inference*. Paket »MuMIn« je namenjen izdelavi in izbiri regresijskih modelov. Modele avtomatsko rangira na podlagi informacijskega kriterija AICc.

»**phangorn**«: *Phylogenetic Reconstruction and Analysis*. Paket »phangorn« je namenjen izdelavi filogenetskih dreves z metodami največjega verjetja in največje varčnosti.

»**picante**«: *Integrating Phylogenies and Ecology*. Paket »picante« vključuje analizo združb, testiranje ničelnega modela in analizo funkcionalnih znakov. Ekofilogenetski pristop omogoča analizo filogenetske in funkcionalne diverzitete združb, analizo filogenetskega signala, različne oblike uvoza/izvoza podatkov in izdelavo slik.

»**pwr**«: *Basic Functions for Power Analysis*. Paket »pwr« omogoča analizo statistične moči.

»**raster**«: *Geographic Data Analysis and Modeling*. Paket »raster« omogoča branje, pisanje, manipulacijo, analiziranje in modeliranje prostorskih podatkov.

»**rgdal**«: *Bindings for the 'Geospatial' Data Abstraction Library*. Paket »rgdal« omogoča združevanje »geoprostorskih« podatkov, njihovo projekcijo in transformacije.

»**rworldmap**«: *Mapping Global Data*. Paket »rworldmap« omogoča izdelavo zemljevidov posameznih držav.

»**sars**«: *Fit and Compare Species-Area Relationship Models Using Multimodel Inference*. Paket »sars« omogoča izračun in izbiro regresijskih modelov v analizi odnosa med vrstnim bogastvom in površino območja.

»**stats**«: *The R Stats Package*. Paket »stats« vsebuje statistične funkcije.

»**vegan**«: *Community Ecology Package*. Paket »vegan« vključuje orodja za analizo: diverzitete, ordinacijske metode in analizo variabilnosti.

»**zetadiv**«: *Functions to Compute Compositional Turnover Using Zeta Diversity*. Paket »zeta« omogoča analizo vrstnega obrata z istočasno obravnavo mnogoterih primerjav.

Dodatek 5

Aritmetika

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

Razlaga aritmetičnih simbolov:

Seštevanje, plus, +; npr.: $a + b$

Odštevanje, minus, –; npr.: $a - b$

Množenje, krat, * ; v zapisu enačb ali v izpisih posameznih primerov množenja uporabljamo različne znake bodisi zvezdico * bodisi piko · bodisi ničesar; npr.: $a * b$, $a \cdot b$, ab

Deljenje, ulomek, —; za zapis ulomka med besedilom uporabljamo tudi poševno črto npr.:

$$\frac{a}{b}, a/b$$

Logaritmiranje, uporabljamo naravni logaritem z osnovo e : \log_e (e se imenuje Eulerjevo število in ima vrednost 2,718) in desetiški logaritem z osnovo 10: \log_{10} ; v skladu s prakso za zapis naravnega logaritma uporabimo \ln in za zapis desetiškega logaritma uporabimo \log ; npr.: $\ln(a)$, $\log(b)$

Korenjenje, uporabljamo kvadratni koren, ki ga zapišemo: $\sqrt{\quad}$; npr.: \sqrt{a}

Potenciranje, v zapisu posameznih zakonitosti ali odnosa med spremenljivkami, ter pri zapisu zelo velikih ali zelo majhnih števil uporabimo potenčni zapis z eksponentom; npr.: a^2 , e^a , ali s strešico: $a^{\wedge}2$, ali z dvema zvezdicama: $a^{**}2$.

Logični operatorji:

= je enako; npr. a je enko b : $a = b$.

≠ ni enako (je različno); npr. a ni enak b : $a \neq b$.

< je manjše; npr. a je manjše od b : $a < b$.

> je večje; npr. a je večje od b : $a > b$.

≤ je manjše ali enako; npr. a je manjše ali enako b : $a \leq b$.

≥ je večje ali enako; npr. a je večje ali enako b : $a \geq b$.

± prištevanje in odštevanje vrednosti za prikaz razpona; npr. k a prištejemo in odštejemo b : $a \pm b$.

∩ presek oz. produkt dogodkov; npr. produkt dogodkov a in b se zgodi, kadar se hkrati zgodita dogodek a in b : $a \cap b$.

| pogojni dogodek; npr. dogodek a se zgodi ob pogoju, da se je zgodil dogodek b : $a | b$.

Dodatek 6

Osnovni koraki v programskem okolju R

Janžekovič, F. (2023). *Makroekologija: Analiza biodiverzitetnih podatkov*. Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.fnm.2.2023

Vmesnik **RStudio** je uporabniku prijazen način programiranja v **R** okolju. Posamezne programske rešitve so oblikovane v **Pakete**. Za uporabo Paketov je potrebno osnovne znanje programskega jezika. Tukaj navajamo nekaj osnov.

R ima, dobro podprto spletno stran: <https://www.r-project.org/> na kateri so objavljene splošne informacije, novosti, navodila, priročniki in dogodki ter povezave na učbenike statistike in uporabe R. Začetnikom v uporabi R je namenjena spletna stran: http://www.introductoryr.co.uk/R_Resources_for_Beginners.html.

V R okolju ustvarjamo R_objekte na način, da jim priredimo vrednosti. Beseda »vrednost« v programski kodi ima širok nabor razumevanja, z nekaj primeri bo predstavljena v nadaljevanju, vključuje pa različne tipe podatkov (skalarje, vektorje, matrike) ter matematične in logične funkcije. Prirejanje vrednosti izvajamo s puščico »<-« in ukazom Run:

```
R_objekt <- vrednost # R_objekt-u priredim vrednost
```

Zgornji zapis se imenuje programski stavek. Sestavljen je iz programske kode (R_objekt <- vrednost) in komentarja (# R_objekt-u priredim vrednost). Besedilo za znakom # so naši komentarji in zapiski, ki niso del programske kode, nam pa pomagajo razumeti vsebino. Zapiske, ki se začenjajo z # lahko vstavljamo v samostojno vrstico ali za programsko kodo v isti vrstici. Objekte in vrednosti oblikujemo in z njimi manipuliramo glede na namen in postopke, ki jih izvajamo, navedimo nekaj primerov:

```
# Objektu z priredim vrednost 5
z <- 5
ali
s <- 4 # Objektu s priredim vrednost 4
#Objekta z in s uporabim kot vrednosti v seštevanju
z+s
[1] 9
ali
#Objekta z in s priredim v nov objekt
p <- z+s
```

Objekt *p* vključim v matematično operacijo korenjenja:

```
sqrt(p)
[1] 3
```

Izpis rezultata v RStudio se vedno prične z oglatim oklepajem [], v katerem je navedena zaporedna številka vrednosti. Informacija postane smiselna oz. nujna, kadar obravnavamo vektor z velikim številom vrednosti.

Osnovne vrednosti v R so zapisane v obliki vektorja. Vektor lahko ima eno samo vrednost, kot npr. z v predhodnem primeru ali pa ima več vrednosti. Z ukazom `c` več vrednosti združimo v objekt:

```
zdruzba <- c(3, 6, 8, 5, 2, 5) # Abundanca 6 vrst ptic v združbi
zdruzba #Ukaz za izpis vrednosti objekta zdruzba
[1] 3 6 8 5 2 5
```

Vektor z vrednostmi abundance za 6 vrst ptic smo združili v objekt *zdruzba*. To omogoča nadaljnjo manipulacijo, npr. kvadriranje:

```
zdruzba**2
[1] 9 36 72 25 4 25
```

Ali druge statistične operacije:

```
sum(zdruzba) # Izračunaj vsoto abundanc
[1] 29
mean(zdruzba) # Izračunaj povprečje abundanc
[1] 4.833
max(zdruzba) # Poišči najvišjo abundanco
[1] 8
```

Omogoča različne manipulacije s podatki, npr. izbor:

```
zdruzba2 <- zdruzba[3:6] # Izbor podatkov od 3. do 6. vrednosti in zapis v nov
objekta
```

in matematične operacije na izboru:

```
sqrt(zdruzba2)
[1] 2.828427 2.236068 1.414214 2.236068
```

Pogledali smo nekaj operacij s številskimi vrednostmi. Vrednosti so lahko tudi znakovne ali logične:

```
ime <- "mlakarica" # Ustvari objekt ime, katerega vrednost je mlakarica
```

```
ime # Izpiši
[1] "mlakarica"
```

Lahko oblikujemo vektor z imeni:

```
ptice<-c("mlakarica", "sivka", "copasta crnica", "kreheljc", "liska", "recni galeb")
ptice
[1] "mlakarica" "sivka" "copasta crnica" "kreheljc" "liska" "recni galeb"
```

Preverimo število elementov v objektu:

```
length(ptice)
[1] 6
```

Oblikujemo logični vektor, npr. za rase iz nabora ptic

```
race<-c(TRUE, TRUE, TRUE, TRUE, FALSE, FALSE)
ptice[race] # Izpis vrst, ki ustrezajo racam
[1] "mlakarica" "sivka" "copasta crnica" "kreheljc"
```

Abundancam ptic lahko dodamo imena ptic:

```
names(zdruzba)<-ptice
zdruzba # Izpiši
mlakarica  sivka  copasta crnica  kreheljc  liska  recni galeb
           3      6          8           5      2      5
```

Vektor z abundancami pretvorimo v matriko velikosti 6 x 1 (šest vrstic in en stolpec) in dodamo imena za stolpec in za posamezno vrstico:

```
zdruzba_m<-matrix(zdruzba) # Vektor pretvori v matriko
zdruzba_m # Izpis matrike
 [1]
[1,] 3
[2,] 6
[3,] 8
[4,] 5
[5,] 2
[6,] 5
colnames(zdruzba_m)<-c("Abundanca") # Dodeli ime za stolpec
rownames(zdruzba_m)<-ptice # Dodeli ime za vrstice iz obstoječega objekta
```

```

zdruzba_m      # Izhajanje matrike
  Abundanca
mlakarica      3
sivka           6
copasta crnica 8
kreheljc       5
liska          2
recni galeb     5

```

Uvoz/izvoz podatkov v R lahko izvedemo na različne načine. Za uvoz podatkov predlagamo oblikovanje .csv ali .txt datoteke in uvoz preko konzole z ukazom `Import data`.

Za izvoz podatkov uporabimo programsko kodo:

```

write.table(zdruzba_m, file = "zdr.txt", sep = ",") # Izvoz .txt datoteke
write.csv(zdruzba_m, file = "zdr.csv") # Izvoz .csv datoteke

```

Nov paket namestimo in ga aktiviramo z ukazom:

```

install.packages("vegan") # Namesti paket »vegan«
library(vegan)           # Aktiviraj paket »vegan«

```

Do pomoči lahko pristopamo skozi konzolo in zavihek `Help`. Druga možnost iskanja pomoči je skozi ukazni stavek, kjer sta možni dve obliki, npr. za logaritmiranje:

```

help(log) # Pomoč za logaritmiranje
?log     # Pomoč za logaritmiranje

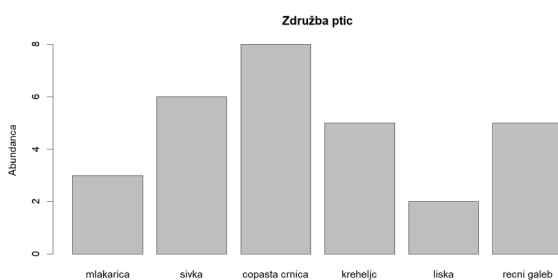
```

R okolje ima izjemen nabor grafičnih možnosti, npr. stolpični graf:

```

barplot(zdruzba, ylab = "Abundanca", main = "Združba ptic") # Izdelaj histogram

```



Stvarno kazalo

A

ACE, cenilka vrstnega bogastva, **52**, 53, 55-58, 486
 AIC, Akaikov informacijski kriterij, 45, 98, 192, 200, 203, 357, 430, 445, **446**, 448, 456, 491
 AIC_C, prilagojeni Akaikov informacijski kriterij, 357, 363, **447**, 448, 454, 491
 Akaikov informacijski kriterij, 45, 98, 192, 200, 203, 357, 430, 445, **446**, 448, 456, 491
 akumulacijska krivulja vrst, 34, **51**, 58, 464, 469, 481
 alfa diverziteta, 34, **36**, 40, 41, 45, 50, 51, 61-64, 66, 139-142, 485, 487
 alometrija, 245, **248**, 249, 273, 294, 297, **308**, 340, anageneza, 348
 analiza glavnih komponent, 20, 185, 181, 185, 187, 216, 243-245, **247**; 255-257, 264, 265, 273, 278, **280**, 282, 298, 309, 313, 488
 analiza glavnih koordinat, 113,116-**118**, 119, 129, 130, **185-187**, 488
 analiza moči, 476
 ANOVA, 247
 aritmetična sredina, 216, 245, **246**, 280, 308, 472
 asimetričnost porazdelitve, **246**
 asociacijski koeficient, 299

B

Bayesov informacijski kriterij, 357, 363, 445, **447**, 491
 Bayesov izrek, 367, **368**,
 Bayesovo sklepanje, 296, 344, 352, 353, 357, 366, **367**, 370, 371, 373, 382, 384, 385, 388, 489
 Bealov indeks, 32, **63-65**, 67, 486
 beta diverziteta temelječa na povprečni filogenetski razdalji, **142**, 487
 beta diverziteta, 34, **37-41**, 43, 44, 50, 51, 66, 139-142, 235, 236, 485, 487
 β , beta diverziteta, 34, **37-41**, 43, 44, 50, 51, 66, 139-142, 235, 236, 485, 487
 BI, Bayesovo sklepanje, 296, 344, 352, 353, 357, 366, **367**, 370, 371, 373, 382, 384, 385, 388, 489
 BIC, Bayesov informacijski kriterij, 357, 363, 445, **447**, 491
 biodiverziteta, 6, **8**, 10, 13, 32, 33, 35, 51, 54, 61, 66, 70, 71, 74, 84, 85, 89, 100, 105, 108, 138, 139, 144, 147, 151, 160, 178, 197, 206-210, 214, 220, 224, 228-230, 238, 240, 295, 345, 346, 348, 392-395, 403, 408, 409, 411-414, 417-419, 421-423, 426, 431, 442, 448, 458, 460, 464, 465, 468-471, 473, 476, 480, 485, 498
 bioklimatska pravila, 270
 Bonferronijev popravek, 289, **290**
 bootstrap, **54**, 56, 345, 366, **377**, 378, 381, 486, 490
 Bray-Curtisova razdalja, **168**, 183, 184, 187, 306, **307**, 488, 489
 Burnabyjeva metoda, 309

C

Canberrova razdalja, **170**, 488
 cenilka fenotipske plastičnosti, **253**
 cenilka okoljsko standardizirane plastičnosti, **254**
 cenilka poenostavljene relativne razdalje v plastičnosti, **254**
 cenilka relativne razdalje v plastičnosti, **253**
 cenilke mnogoterih primerjav, **40**, 41, 485
 cenilke parnih primerjav, **37-39**, 485
 Chao1, cenilka vrstnega bogastva, **52**, 55, 57, 486
 Chao2, cenilka vrstnega bogastva, **53**, 56, 57, 486
 Clarkova razdalja, **170**, 488
 Cohenov d, **474**
 Collesov indeks, **142**, 487
 CV koeficient variacije, 253
 CWM, tehtano povprečje združbe, **110**, 486

D

D, indeks dominanc **93**, 94
 D, Simpsonov diverzitetni indeks vrstnega bogastva, **81**, 486
 dendrogram, 154, 180, 190, 373, 374, **376**
 dimenzija bogastva, 6, **8**, 9, 13, 104, 107, 134, 138-142, 151, 485
 dimenzija poravnosti, 6, **9**, 13, 104, 107, 134, 138-142, 151, 485
 dimenzija raznolikosti, 6, **9**, 13, 104, 107, 134, 138-142, 151, 485
 diskriminantna analiza, 243, 278, **287**, 288, 290, 291, 488
 drevo najkrajšega razvejanja, 131, 133, 487
 drevo soglasja, 345, 378, 490
 drevo življenja, 345, 346

E

E₁, indeks poravnosti, **82**, 83, 486
 E₂, indeks poravnosti, **82**, 83, 486
 E₃, indeks poravnosti, **82**, 83, 486
 E₄, indeks poravnosti, **82**, 83, 486
 E₅, indeks poravnosti, **82**, 83, 486
 endosimbioza, 346
 enostavna linearna regresija, **432**, 434, 437, 460, 490
 enostavno naključno vzorčenje, 409, **413**, 426
 ES, velikost učinka, 464, **473-477**, 479, 480
 Evklidova razdalja, 109, 113, **114**, 122, 123, 126, 130, 131, 177, 178, 185, 187, **306**, 307, 489
 evolucijska pravila, 271

F

FDis, funkcionalna razpršenost, 104, 109, **113**, 115, 116, 119, 120, 134, 487

FDiv, funkcionalno razhajanje, 104, 109, **126-129**, 134, 487
 fenetika, **296**, 349, 489
 fenogram, 296, 297, 338, **376**, 388
 fenotip, **249-251**, 256, 266, 273, 297, 347, 349
 fenotipska plastičnost, **250**, 251, 253, 254, 256, 257, 259, 260, 262-266, 268, 273, 295, 297
 fenotipska variabilnost, 244, **249**, 250, 253, 259, 273, 308
 FEve, funkcionalna poravnost, 104, **106-109**, **131**, 132, 134, 487
 filofenetika, **297**, 340, 350
 filogenetika, **295**, 296, 344, **345**, 348-350, 388, 490
 filogenetska pestrost obtežena z abundanco, **141**, 487
 filogenetska pestrost, 9, **138-144**, 151, 345, 487
 filogenetska poravnost, 9, 487
 filogenetska rekonstrukcija, 138, 139, 147, 295-298, 301, 308, 310, 316, 322, 329, 337, 340, 344, 345, 347-349, 352, 357, 366, 368, 370, 371, 373, 377, 378, 386, 388, 489
 filogenetski endemizem, **141**, 487
 filogenetski signal, 296, 309, **385**, 386, 388
 filogenetski vidik pestrosti, 6, **8**, 9, 13
 filogenetsko bogastvo, 9, 487
 filogenetsko drevo, 9, 11, 139, 143, 294-**297**, 329, 330, 340, **345**, 362, 369, **371**, 372, 374, 377, 386
 filogenetsko vrstno bogastvo, **141**, 487
 filogeneza, 295, 332, 346
 filogenija, 294, 295, **345**,
 filogram, 296, 345, **376**, 377, 388
 filtriranje vrst, 61, 62
 Fisherjeve logaritemske porazdelitve, **91-95**, 100, 486
 FRic, funkcionalno bogastvo, 9, 104, 106, **107-109**, **129**, 130, 134, 487
 FSpe, funkcionalna specializiranost, 104, 109, **122**, 134, 487
 F-test, 245, 247, 253, 255, 259, 311, 446
 funkcionalna pestrost, 9, **104-107**, 109, 134, 486, 487
 funkcionalna poravnost, 104, **106-109**, **131**, 132, 134, 487
 funkcionalna razpršenost, 104, 109, **113**, 115, 116, 119, 120, 134, 487
 funkcionalna specializiranost, 104, 109, **122**, 134, 487
 funkcionalni vidik pestrosti, 6, **8**, 9, 13
 funkcionalno bogastvo, 9, 104, 106, **107-109**, **129**, 130, 134, 487
 funkcionalno razhajanje, 104, 109, **126-129**, 134, 487

G

gama diverziteta, 34, **36**, 40, 41, 51, 61-63, 66, 485
 generalizirani časovno reverzibilni model, **356**, 357, 380, 383, 490
 genetska razdalja, 351, 487
 genska banka, 311, **360**, 379
 geometrijska morfometrija, 248, 255, 267, 268, 279, 295, 306, 310, **315-319**, 340, 489
 Gini-Simpsonov indeks, **73**
 Gowerjev splošni koeficient podobnosti, **117**, 118, **299**, 303-305, 488, 489
 Gowerjeva razdalja, 113, **118-120**, 123, 130, 131
 gradient vrstnega bogastva vzdolž geografske širine, 207, 210
 GTR, generalizirani časovno reverzibilni model, **356**, 357, 380, 383, 490

H

H', Shannonov diverzitetni indeks vrstnega bogastva, **81**, 486
 Hellingerjeva razdalja, **170**, 488
 Hellingerjeva transformacija, **179**, 488
 hierarhični test razmerja verjetij, **357**
 hierarhično združevanje v skupine, **181**, 189
 Hi-kvadrat X_2 koeficient podobnosti, 164, **165**, 488

Hi-kvadrat X^2 razdalja, **171**, 488
 Hi-kvadrat X^2 , **170**, 488
 Hillova diverzitetna števila, **80-84**, 87-90, 100, 486
 Hillovo diverzitetno število vrstnega bogastva, **81**, 486
 Hillovo število Shannonove vrstne diverzitete, **81**, 486
 Hillovo število Simpsonove recipročne diverzitete, **81**, 486
 homologija, 295, 296, 332, 333, 340, 489
 homologni geni, 349
 homoplazija, 332, 333,
 horizontalni prenos genov, 346, 349
 Huisman-Olff-Frescojevi modeli, 231

I

ICE, cenilka vrstnega bogastva, **53-55**, 486
 indeks dominanc **93**, 94
 indeks popolnosti združbe, 32, **63**, 64, 67, 486
 indeks ugnedene mnogotere raznolikosti, **41**
 indeks ugnedene raznolikosti, **41**

J

Jaccardov koeficient, **162**, 163, 167, **173**, 174, 183, **300-302**, 322, 327, 329, 488, 489
 jackknife, **378**, 490
 jackknife1, cenilka vrstnega bogastva, **54**, 56, 486
 jackknife2, cenilka vrstnega bogastva, **54**, 56, 486
 JC, Jukes-Cantorjeva razdalja, 320, 344, 352, **355**, 358, 366, 379, 388, 490
 Jukes-Cantorjev model, 320, 352, **355**, 363, 490
 Jukes-Cantorjeva razdalja, 320, 344, 352, **355**, 358, 366, 379, 388, 490

K

K2, Kimura 2 razdalja, 320, **355**, 490
 Kimura 2 razdalja, 320, **355**, 490
 Kimura 2-parametrični model, **355**, 357, 490
 klad, 332
 kladistika, 331, 348, 350, 489
 kladogeneza, 345, 348, 371
 kladogram, 294, 298, 332-334, 338-340, 344, 350, **376**, 377, 388
 klasično multidimenzionalno skaliranje, 186
 klastrska analiza, 181
 klina, 244, **252**, 273
 koeficient determinacije, 186, 187, 193, 200, 203, 233, 436, 440, 444, 446, 451, 456, 473, **475**,
 koeficient enostavnega ujemanja, **162**, 488
 koeficient variacije, **253**
 koeficienta inflacije variance, 444, 451
 kolinearnost, **443**, 448, 450, 451, 455
 Kulczynskijev koeficient, **164**, 488

L

Linnejeva taksonomija, **147**

M

Mahalanobisova razdalja, 290, 291, **307**, 489

makroekologija, 6, 7, 8, 12, 13, 17, 18, 83, 104, 107, 108, 134, 139, 206, 207, 208, 223, 244, 270, 295, 385
maksimum, **246**
Manhattanova razdalja, 306, **307**, 489
MANOVA, 247
Mantelov test, 232, 235, 431, **458**, 459, 461, 490
Margalefov indeks, **72**
MCMC, Monte Carlo z markovskimi verigami, **369-371**, 383
MCMCMC, Metropolis Monte Carlo z markovskimi verigami, 370, **371**, 383
MDE, učinek osrednjega območja, **210**, **231**, 238-240
Menhinickov indeks, **72**
metoda glavne osi, 430, **438**, 441, 460, 490
metoda komolca, 217, 276, 281, **284**, 285, 292
metoda najmanjših kvadratov, 430, **435**, 460, 490
metoda največje varčnosti, 334, **336**, 338, 374, 373
metoda največjega verjetja, 143, 344, **361**, 362, 366, 373, 377, 379, 381, 385, 387, 388, 437, 490
metoda navadnih najmanjših kvadratov, **435**, 439, 441
metoda rangirane glavne osi, **438**, 490
metoda standardne glavne osi, **438**, 490
metoda voditeljev, 276, 278, **283**, 285, 286, 292, 488
metoda združevanja sosedov, 294, 296, **319**, 322, 326, 340, 344, 359, 373, 377, 379, 385, 388, 489, 490
metrično multidimensionalno skaliranje, 186
Metropolis Monte Carlo z markovskimi verigami, 370, **371**, 383
minimum, **246**
ML, metoda največjega verjetja, 143, 344, **361**, 362, 366, 373, 377, 379, 381, 385, 387, 388, 437, 490
moderna sinteza, 347
modificiran Gowerjev koeficient, 164, **166**, 488
modificirani koeficient povprečne razlike, **171**, 488
monitoring, **393-395**, 401, 403, 409, 421
monofiletska skupina, 32, 65, 67, 294-296, **332-337**, 340, 344, 348, 376, 388, 489
Monte Carlo z markovskimi verigami, **369-371**, 383
morfometrija, 111, 115, 245, 252, 254, 255, 267, 273, 280, 386, 431, 458, 469
morfoprostor, **251**, 281, 289, 291,
MP, metoda največje varčnosti, 334, **336**, 338, 374, 373
MST, drevo najkrajšega razvejanja, 131, 132, 487
multidimensionalno skaliranje, 185, 189
multikolinearnost, **443**
multipla linearna regresija, **442**, 443, 448, 450, 451, 453, 490
multipli determinacijski koeficient, **446**
mutacija, 17, 210, 220, 295, 332, 344, 347, 348, 351-353, 375, 376, 388

N

načelo FAIR, **422**
naravni izbor, 347, 348
naravoslovna zbirka, **425**
ne-metrično multidimensionalno skaliranje, **186**
NJ, metoda združevanja sosedov, 294, 296, **319**, 322, 326, 340, 344, 359, 373, 377, 379, 385, 388, 489, 490
nukleotidna p-razdalja, 354, 490
numerična taksonomija, 276, **277**, 292

O

odnos med vrstnim bogastvom in površino ozemlja, SAR, **183**
odprta znanost, **420**

Odumov indeks, **72**
opisna statistika, 246
oslonilne točke, 255, 267, 268, 279, 282, 310, 315-317
osrednja točka ugodja, **231**

P

PCA, analiza glavnih komponent, 20, 185, 181, 185, 187, 216, 243-245, **247**; 255-257, 264, 265, 273, 278, **280**, 282, 298, 309, 313, 488
PCoA, analiza glavnih koordinat, 113,116-**118**, 119, 129, 130, **185-187**, 488
PD, filogenetska pestrost, **140**, 487
PE, filogenetski endemizem **141**, 487
Pearsonov korelacijski koeficient, 173
permutacijski test, 270, 276, **284**, 285, 292
Podanijeva razdalja, 130, 131
pogojna verjetnost, **367**
polinomska regresija, 222, 430, **456**, 457, 460, 490
politomija, 138, 139, 151, 338, 339, 344, 348, 377-379
popolno taksonomsko razlikovanje, **146**
povprečna evolucijska posebnost, **142**, 487
povprečna filogenetska pestrost obtežena z abundanco, **142**, 487
povprečna filogenetska pestrost, **142**, 487
povprečna razdalja najbližjega taksona, **142**, 487
povprečna taksonomska pestrost, **145**
povprečno taksonomsko razlikovanje, **146**
prava diverziteta β , **40**
prečno preverjanje, 290
preskus značilnosti, 466, 468
Prestonova logaritemsko normalna porazdelitev, **92**, 93, 96, 100
prikrita diverziteta, 34, **65**
prilagojeni Akaikov informacijski kriterij, 357, 363, **447**, 448, 454, 491
prilagojeni multipli determinacijski koeficient, 200, 445, **446**; 451, 452
Prokrustova analiza, **315**, 316, 489
Prokrustova razdalja, **307**, 310, 315, 316, 318, 319, 489
PSR, filogenetsko vrstno bogastvo, **141**, 487

Q

Q analiza, 156
Q, Raova kvadratna entropija, 104, 109, **123**, 124, 134, 487

R

R analiza, 156, 172
R1, Margalefov indeks, **72**
R², koeficient determinacije, 186, 187, 193, 200, 203, 233, 436, 440, 444, 446, 451, 456, 473, **475**,
R2, Menhinickov indeks, **72**
R², multipli determinacijski koeficient, **446**
R²_a, prilagojeni multipli determinacijski koeficient 200, 445, **446**; 451, 452
R3, Odumov indeks, **72**
radialno drevo, 344, **376**, 377, 388
Raova kvadratna entropija, 104, 109, **123**, 124, 134, 487
razdalja med vrstnimi profili, **170**, 488
reakcijska norma, 251
redčenje, 32, 34, **57**, 66, 84
relativna vrstna abundanca, **91**

S

S, diverzitetni indeks vrstnega bogastva, **81**, 486
 SAR, odnos med vrstnim bogastvom in površino ozemlja, **183**
 Shannonov indeks, **73, 74**, 80-83, 100, 486
 Shepardov diagram, **185**
 SIE, učinek malega otoka, 194
 silhuetna metoda, **283**, 285
 Simpsonov indeks diverzitete, 70 **73**, 78, 79, 100, 486
 Simpsonov indeks mnogoterih raznolikosti, **40**, 41, 43, 485
 Simpsonov indeks raznolikosti, 39, **40**, 43,
 Simpsonov indeks, 70-**72**, 78, 79, 83, 100, 145, 486
 Simpsonov recipročni indeks, 70 **73**, 79, 81, 82, 88, 89, 100, 486
 sinapomorfija, 332, 333, 335, 339, 344, 376, 377, 388, 489
 sistematično vzorčenje, 409, 413, **414**, 417, 426
 skupna verjetnost, **367**
 Sørensenov indeks mnogoterih raznolikosti, **40**, 41, 43, 485
 Sørensenov koeficient, **162**, 163, **173**, 174, 488
 Spearmanov koeficient korelacije, 173
 speciacija, 208, 224, **348**, 349
 sploščenost porazdelitve, **246**
 standardizacija, **307**, 308, **438**, 488
 standardna deviacija, 216, 245, **246**, 280, 308
 statistični test, 232, 433, 465-468, 476, 480
 Steinhaušov koeficient, **164**, 167, 168, 488
 stopnja značilnosti, 466
 stratificirano naključno vzorčenje, 409, 413, **414**, 426

T

T3, Tamura 3 razdalja, **355**, 490
 Tajima-Nei model, **355**, 357, 490
 Tajima-Nei razdalja, **355**, 490
 taksonomija, **277**, 292
 taksonomska pestrost, 139, **144-146**, 151
 taksonomska širina, **146**
 Tamura 3 razdalja, **355**, 490
 Tamura 3-parametrični model, **355**, 357, 490
 Tamura-Nei model, **356**, 357, 490
 Tamura-Nei razdalja, 320, **356**, 490
 tehtano povprečje združbe, **110**, 486
 temna diverziteta, 33, 34, 51, **61**, **62**, 64-67, 486
 test razmerja verjetij, **357**
 testiranje ničelne hipoteze, **465**,
 tetivna razdalja, **169**, 171, 488
 tetivna transformacija, **179**, 488
 TN, Tajima-Nei razdalja, **355**, 490
 TN, Tamura-Nei razdalja, 320, **356**, 490
 tradicionalna morfometrija, 248, 255, 295, 306, **308-313**, 340, 489
 transformacija s korenjenjem, **178**, 488
 transformacija s logaritmiranjem, **178**, 488
 transverzija, 320, 352-357
 tranzicija, 320, 352-357

U

učinek malega otoka, SIE, 194
 učinek osrednjega območja, **210**, **231**, 238-240
 ugnezenost, 34, **40**, 41, 43, 44, 485
 unifrac, **141**, 487
 UPGMA, metoda neutženih aritmetičnih sredin, 183, 294, 296,
327, 329, 340, 344, 352, 357, 373, 374, 377, 388, 489,
 490

V

varčnost, 294, 332, 333, **446**, 448, 451, 456, 457, 460
 variabilnost v taksonomskem razlikovanju, **146**
 varianca evolucijske posebnosti, **142**, 487
 varianca razdalj med pari, **142**, 487
 varianca, 183, 216, 246, 247, 258, 265, 280-282, 284, 288, 289,
 296, 306, 307, 430, 436, 438, 442, 446, 470, 473, 475
 velikost telesa, **17-28**, 111, 112, 243-246, 248-251, 256-273, 294,
 296, 297, 308-310, 312, 313, 315, 316, 340
 velikost učinka, ES, 464, **473-477**, 479, 480
 velikost vzorca, 245, **246**, 434, 465, 469, 470, 472, 473, 476, 477,
 479
 verjetje, 350, **362**, 364, 365
 VIF, koeficienta inflacije variance, 444, 451
 vrstna pestrost, 9, 33, 70, **71**, 100, 219, 230, 486
 vrstna poravnost, 9, 70, **71**, **82**, 83, 100, 486
 vrstni obrat, 34, **37-41**, 44, 45, 47, 51, 66, 141, 235, 485
 vrstni vidik pestrosti, 6, **8**, 9, 13
 vrstno bogastvo, 9, **32-35**, 41, 50, 52, 53, 61, 64, 66, 71, 74, 81,
 83, 88, 89, 100, 140, 146, 148, 192, 194-196, 198, 203,
 204, 207, 209-214, 216, 219-222, 224, 228, 229, 231,
 233, 234, 237, 238, 240, 271, 417, 436, 448, 450-453,
 464, 469, 481, 486
 vsota evolucijskih posebnosti, **141**, 487
 vzorčenje s presojo, 409, **414**, 426
 vzorčna pokritost, 84, 86, 87, 89-91, 100
 vzorec, 9, 33, 57, 73, 146, 155, 408, 409, **413**, 426, 465, 466, 468,
 470, 471

W

Whittakerjev indeks asociacije, **170**, 488
 Whittakerjeva porazdelitev rangiranih abundanc, **92**, 93, 98, 100

Z

z standardizacija, **307**, 308, **438**, 488
 zadnji univerzalni skupni prednik, LUCA, 345,346
 zeta diverziteta, 34, **44-51**, 66, 485
 zeta razmerje, **46-50**, 485

MAKROEKOLOGIJA: ANALIZA BIODIVERZITETNIH PODATKOV

FRANC JANŽEKOVIČ

Univerza v Mariboru, Fakulteta za naravoslovje in matematiko, Maribor, Slovenija
franc.janzekovic@um.si

Povzetek Makroekologija, analiza biodiverzitetnih podatkov je univerzitetni učbenik, namenjen usvajanju znanja s področja ved o biotski pestrosti. Vsebina obsega vsa področja biodiverzitete, s katero se srečuje sodobna biološka, ekološka in naravovarstvena stroka. Predstavljena je metodologija raziskovanja v biodiverziteti, ki sledi konceptu razdelitve biodiverzitete na tri vidike: na vrstno, funkcionalno in filogenetsko diverziteto. Pri vrstni in funkcionalni diverziteti so predstavljene cenilke bogastva in pestrosti ter metode ocenjevanja vrstnega bogastva. Predstavljen je prostorski koncept diverzitete s cenilkami alfa, beta, gama in zeta diverzitete. Filogenetska diverziteta obsega vsebine filogenetskih rekonstrukcij na morfoloških in genetskih znakih. Predstavljene so cenilke filogenetske pestrosti, taksonomske diverzitete in analiza filogenetskega signala. Vsebina analize ekoloških združb obsega Q in R pristop vrednotenja in izračuna cenilk, s katerimi raziskujemo odnose med združbami. Obravnavani so biodiverzitetni vzorci in procesi v prostoru in času: odnos med diverzitetno in velikostjo območja, odnos med diverzitetno in geografsko širino, odnos med diverzitetno in nadmorsko višino in odziv diverzitete vzdolž ekoloških gradientov. Metodologija morfometričnih analiz vsebuje tradicionalni pristop in pristop z geometrijsko morfometrijo za namen taksonomskih analiz in analiziranja fenotipske plastičnosti. Predstavljene so splošne metodološke vsebine, ki segajo na področje vzorčenja in izvajanju monitoringa, ter vsebina o regresijah in statističnem sklepanju, pomembne za proučevanje biodiverzitete.

Ključne besede:
makroekologija,
biodiverziteta,
genetska diverziteta,
vrstna diverziteta,
funkcionalna
diverziteta

MACROECOLOGY: ANALYSIS OF BIODIVERSITY DATA

FRANC JANŽEKOVIČ

University of Maribor, Faculty of Natural Sciences and Mathematics, Maribor, Slovenia
franc.janzekovic@um.si

Abstract Macroecology, Biodiversity Data Analysis is a university textbook aimed at acquiring knowledge in the field of biotic diversity sciences. The content includes all areas of biodiversity encountered by the modern biological, ecological and nature conservation profession. The methodology of biodiversity research is presented, which follows the concept of dividing biodiversity into three aspects: species, functional and phylogenetic diversity. In the case of species and functional diversity, richness and variety evaluations and methods of species richness assessment are presented. The spatial concept of diversity with alpha, beta, gamma and zeta diversity estimators is presented. Phylogenetic diversity comprises the contents of phylogenetic reconstructions based on morphological and genetic characters. Estimators of phylogenetic diversity, taxonomic diversity and phylogenetic signal analysis are presented. The content of the analysis of ecological communities includes the Q and R approach of evaluation and calculation of indicator with which we investigate the relationships between communities. Biodiversity patterns and processes in space and time are discussed: the relationship between diversity and area size, the relationship between diversity and latitude, the relationship between diversity and altitude, and the response of diversity along ecological gradients. The methodology of morphometric analyses includes a traditional approach and an approach with geometric morphometrics for the purpose of taxonomic analysis and analysing phenotypic plasticity. General methodological content, which extends to the field of sampling and monitoring, as well as content on regressions and statistical inference, important for the study of biodiversity, are presented.

Keywords:

macroecology,
biodiversity,
phylogenetic
diversity,
species diversity,
functional diversity



Franc Janžekovič je redni profesor za področje zoologije in zaposlen na Univerzi v Mariboru. Na študijskih programih biologije in ekologije predava vsebine o vretenčarjih in o biodiverziteti, s poudarkom na metodologiji vrednotenja biodiverzitete. Raziskovalno deluje na področju proučevanja biodiverzitete živali predvsem ptičev in malih sesalcev, morfometričnih raziskavah glodalcev in na področju zooarheologiji. Posveča se razumevanju odnosa med plenom in plenilcem na modelnem sistemu med sovami in malimi sesalci. Raziskuje fenotipsko plastičnost glodalcev in členonožcev, kjer razlaga morfologijo živali z abiotskimi ekološkimi dejavniki. Vključen je v vrednotenje biodiverzitete podzemnih habitatov. Sodeluje v arheoloških raziskavah koliščarske kulture, predvsem pri vrednotenju in pomenu ptic v ekonomiji koliščarjev.



Maša Miklavc zaključuje osnovno šolo, kmalu bo stara 15 let. Njena želja je, da se vpiše v klasično gimnazijo, kjer pričakuje, da bo več družboslovnih predmetov, nato pa načrtuje vpis na likovno akademijo. Že od nekdaj rada riše, hodi na izbirne predmete likovne umetnosti in v popoldansko likovno šolo. Prosti čas nameni risanju v različnih tehnikah.



Univerza v Mariboru

Fakulteta za naravoslovje
in matematiko