

# STATISTIKA

z raziskovalnimi metodami



Maja Rožman

Polona Tominc



Univerzitetna založba  
Univerze v Mariboru





Univerza v Mariboru

Ekonomsko-poslovna fakulteta

# Statistika z raziskovalnimi metodami

## Avtorici

Maja Rožman

Polona Tominc

November 2024

<b>Naslov</b> <i>Title</i>	<b>Statistika z raziskovalnimi metodami</b> <i>Statistics With Research Methods</i>
<b>Avtorici</b> <i>Authors</i>	Maja Rožman (Univerza v Mariboru, Ekonomsko-poslovna fakulteta)  Polona Tominc (Univerza v Mariboru, Ekonomsko-poslovna fakulteta)
<b>Recenzija</b> <i>Review</i>	Vesna Čančer (Univerza v Mariboru, Ekonomsko-poslovna fakulteta)  Dijana Oreški (Univerza v Zagrebu, Fakulteta organizacije in informatike)
<b>Lektoriranje</b> <i>Language editing</i>	Alenka Plos (Univerza v Mariboru, Ekonomsko-poslovna fakulteta)
<b>Tehnični urednik</b> <i>Technical editor</i>	Jan Perša (Univerza v Mariboru, Univerzitetna založba)
<b>Grafične priloge</b> <i>Graphics material</i>	Viri so lastni, razen če ni navedeno drugače. Rožman, Tominc (avtorici), 2024
<b>Oblikovanje ovitka</b> <i>Cover designer</i>	Jan Perša (Univerza v Mariboru, Univerzitetna založba)
<b>Grafika na ovitku</b> <i>Cover graphic</i>	Entrepreneur Diagram Curve, avtor: geralt, pixabay.com, 2024 A Group of People Discussing Charts, avtorica: Mikael Blomkvist, pexels.com, 2020 Close-Up Photo of Accounting Documents, avtor: Kaboompics.com, pexels.com, 2021
<b>Založnik</b> <i>Published by</i>	<b>Univerza v Mariboru</b> <b>Univerzitetna založba</b> Slomškov trg 15, 2000 Maribor, Slovenija <a href="https://press.um.si">https://press.um.si</a> , <a href="mailto:zalozba@um.si">zalozba@um.si</a>
<b>Izdajatelj</b> <i>Issued by</i>	<b>Univerza v Mariboru</b> <b>Ekonomsko-poslovna fakulteta</b> Razlagova ulica 14, 2000 Maribor, Slovenija <a href="https://www.epf.um.si">https://www.epf.um.si</a> , <a href="mailto:epf@um.si">epf@um.si</a>
<b>Izdaja</b> <i>Edition</i>	Prva izdaja
<b>Izdano</b> <i>Published at</i>	Maribor, november 2024
<b>Vrsta publikacije</b> <i>Publication type</i>	E-knjiga
<b>Dostopno na</b> <i>Available at</i>	<a href="https://press.um.si/index.php/ump/catalog/book/7epf24">https://press.um.si/index.php/ump/catalog/book/7epf24</a>





© Univerza v Mariboru, Univerzitetna založba

/ University of Maribor, University Press

**Besedilo** / *Text* © Rožman, Tominc (avtorici), 2024

To delo je objavljeno pod licenco Creative Commons Priznanje avtorstva-Nekomercialno-Deljenje pod enakimi pogoji 4.0 Mednarodna. / *This work is released under a Creative Commons Attribution-Noncommercial-Share Alike 4.0 International license.*

Uporabnikom je dovoljeno reproduciranje, distribuiranje, dajanje v najem, javno priobčitev in predelavo avtorskega dela, če navedejo avtorja in širijo avtorsko delo/predelavo naprej pod istimi pogoji. Za nova dela, ki bodo nastala s predelavo, ni dovoljena komercialna uporaba.

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-sa/4.0/>

CIP - Kataložni zapis o publikaciji  
Univerzitetna knjižnica Maribor

311.1:004.42 (0.034.2)

ROŽMAN, Maja, 1988-

Statistika z raziskovalnimi metodami [Elektronski vir] / avtorici Maja Rožman, Polona Tominc. - 1. izd. - E-publikacija. - Maribor : Univerza v Mariboru, Univerzitetna založba, 2024

Način dostopa (URL): <https://press.um.si/index.php/ump/catalog/book/7epf24>

ISBN 978-961-286-929-8 (Pdf)

doi: 10.18690/um.epf.7.2024

COBISS.SI-ID 215399171

**ISBN** 978-961-286-929-8 (pdf)

**DOI** <https://doi.org/10.18690/um.epf.7.2024>

**Cena**  
*Price* Brezplačni izvod

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

**Citiranje** Rožman, M., Tominc, P. (2024). *Statistika z raziskovalnimi metodami*. Univerza v  
*Attribution* Mariboru, Univerzitetna založba. doi: 10.18690/um.epf.7.2024



# Kazalo

<b>1</b>	<b>Uvod.....</b>	<b>1</b>
<b>2</b>	<b>Naloge za seminarske vaje.....</b>	<b>5</b>
2.1	Formulacija raziskovalnega problema .....	5
2.2	Urejanje in prikazovanje podatkov.....	11
2.3	Deskriptivna statistika .....	24
2.4	Relativna števila: indeksna števila, povprečna vrednost v časovni vrsti.....	45
2.5	Metode zbiranja podatkov .....	59
2.5.1	Sekundarni in primarni viri.....	59
2.5.2	Opazovanje.....	60
2.5.3	Intervju .....	60
2.5.4	Vprašalnik .....	61
2.6	Normalna porazdelitev .....	71
2.7	Osnove vzorčenja in osnove preizkušanja domnev.....	84
2.8	Regresijska analiza.....	106
2.9	Napovedovanje vrednosti v časovni vrsti z uporabo trenda in sezonske komponente .....	124
<b>3</b>	<b>Naloge za laboratorijske vaje.....</b>	<b>133</b>
3.1	Deskriptivna statistika in vzorčni pristop.....	133
3.2	Normalna porazdelitev .....	143
3.2.1	Preverba domneve o normalni porazdelitvi obravnavane spremenljivke.....	143
3.2.2	Kolmogorov-Smirnov test in Shapiro-Wilk W test .....	143
3.3	Univariatni statistični testi.....	147
3.3.1	Parametrični test za odvisna vzorca: t-test za odvisna vzorca.....	148
3.3.2	Parametrični test za neodvisna vzorca: t-test za neodvisna vzorca.....	151
3.3.3	Parametrični test za več kot 2 neodvisna vzorca: ANOVA.....	155
3.3.4	Neparametrični test: $\chi^2$ -test za analizo povezanosti dveh nominalnih spremenljivk.....	158
3.4	Faktorska analiza.....	159
3.5	Enostavna linearna regresija .....	167
3.6	Multipla regresijska analiza .....	172
<b>4</b>	<b>Rešitve računskih nalog .....</b>	<b>177</b>
4.1	Urejanje in prikazovanje podatkov.....	177
4.2	Deskriptivna statistika .....	185
4.3	Relativna števila: indeksna števila, povprečna vrednost v časovni vrsti.....	202
4.4	Normalna porazdelitev .....	207
4.5	Osnove vzorčenja in osnove preizkušanja domnev.....	217
4.6	Regresijska analiza.....	235
4.7	Napovedovanje vrednosti v časovni vrsti z uporabo trenda in sezonske komponente .....	245

Ploščine $H(z)$ za standardizirano normalno porazdelitev.....	251
Kritične vrednosti za $t$ porazdelitev.....	252
Obrazci.....	253
Literatura in viri.....	265



# 1 Uvod

V današnjem svetu, kjer se poslovno okolje nenehno spreminja, je sposobnost razumevanja in uporabe statistike bolj pomembna kot kdaj koli prej. Od napredka tehnologije do naraščajočih podatkovnih baz, od podnebnih sprememb do družbenih trendov – zmožnost analize in interpretacije podatkov je ključna za uspeh pri sprejemanju utemeljenih poslovnih odločitev. Statistično znanje omogoča ne le prepoznavati trende in gibanja na trgu, ampak tudi sprejemati odločitve, ki temeljijo na analizi podatkov. S pomočjo statistike lahko ugotovimo na primer, v kolikšni meri ima povpraševanje po določenem izdelku ali storitvi sezonski značaj, kakšna je razporeditev potrošnikov po starostnih skupinah ali geografskih regijah ter kakšen vpliv imajo različni poslovni ukrepi na prodajo. S statističnimi tehnikami lahko analiziramo ekonomske kazalnike, kot so rast BDP, inflacija in brezposelnost, ter ocenimo, kako vplivajo na poslovno strategijo in odločanje podjetij. Statistika pomaga razumeti, kateri proizvodi ali storitve so bolj priljubljeni in kako oblikovati učinkovite marketinške strategije. S tem znanjem lahko sprejemamo odločitve, ki temeljijo na dejstvih in podatkih. Postati kompetentni v statistiki pomeni biti pripravljeni na prihodnost, pomeni razumevanje dinamike sprememb in sposobnost prilagoditi svoja orodja in tehnike tako, da so najučinkovitejši.

Gradivo *Statistika z raziskovalnimi metodami* je namenjeno študentom gospodarskega inženirstva za pomoč pri razumevanju in uporabi različnih statističnih tehnik ter raziskovalnih metod, kar študentom omogoča analizo podatkov, interpretacijo rezultatov in reševanje raziskovalnih problemov v akademskih ter poklicnih okoljih. Cilj tega gradiva je zagotoviti celovito razumevanje tako osnovnih kot naprednih statističnih konceptov.

Gradivo vključuje rešene primere raznolikih nalog, razlage in ilustracije, ki vodijo študente skozi različne faze raziskovalnega procesa. V naslednjih poglavjih so opisani ključni koncepti, potrebni za razumevanje in uporabo statističnih metod. Gradivo je sestavljeno iz nalog za seminarske in laboratorijske vaje. Praktični del gradiva vsebuje naloge za laboratorijske vaje z uporabo statistične programske opreme SPSS. Te vaje pomagajo utrditi znanje, pridobljeno v predhodnih poglavjih, ter ilustrirajo praktično razsežnost uporabe statističnih metod.

Poglavja vključujejo naslednje vsebine:

- *Formulacija raziskovalnega problema*: to poglavje govori o tem, kako identificirati in jasno formulirati raziskovalni problem. Sposobnost pravilne formulacije problema je bistvena za uspešno raziskavo, saj določa smer in obseg študije.
- *Urejanje in prikazovanje podatkov*: v tem poglavju obravnavamo različne metode za zbiranje, urejanje in prikazovanje podatkov ter za jasen in razumljiv prikaz podatkovnih nizov. Obravnavamo, kako uporabiti različne grafične prikaze, kot so stolpčni in tortni diagrami, histogrami in linijski grafi.
- *Deskriptivna statistika*: v tem poglavju obravnavamo kazalnike za kvantitativni opis podatkovnega niza, kot so aritmetična sredina, modus, mediana, varianca in standardni odklon idr. Te metode pomagajo razumeti osnovne značilnosti nabora podatkov in oblikovati temelje za nadaljnje analize.
- *Relativna števila, indeksna števila in povprečna vrednost v časovni vrsti*: v tem poglavju prikazujemo, kako primerjati podatke skozi čas ali med različnimi geografskimi območji. Indeksna števila so koristna orodja za analizo dinamike in trendov, povprečna vrednost v časovni vrsti pa omogoča tudi ocenjevanje vrednosti spremenljivke v prihodnjih časovnih enotah.
- *Metode zbiranja primarnih podatkov*: v tem poglavju obravnavamo različne metode za zbiranje podatkov, kot so opazovanje, intervjuji in vprašalniki, z namenom pridobiti kakovostno podatkovno osnovo za poslovno odločanje.
- *Normalna porazdelitev*: normalna porazdelitev je osnova številnih statističnih metod. V tem poglavju obravnavamo njene značilnosti, kako jo prepoznati in zakaj je ključna za analizo podatkov. V drugem delu gradiva, ki se nanaša na laboratorijske vaje, prikazujemo tudi Kolmogorov-Smirnov test in Shapiro-Wilk  $W$  test, ki se uporabljata za preverjanje hipoteze o normalnosti porazdelitve obravnavane spremenljivke.
- *Osnove vzorčenja in preizkušanja domnev*: v tem poglavju prikazujemo različne tehnike vzorčenja, ki so bistvene za pridobivanje reprezentativnih vzorčnih podatkov.

Prikazujemo tudi osnove preizkušanja domnev, ki omogočajo sklepanje o populaciji na podlagi vzorca.

- *Univariatni statistični testi*: služijo za preverjanje domnev o razlikah v povprečni vrednosti spremenljivke v dveh ali več odvisnih ali neodvisnih skupinah, kar predstavlja pogosto vprašanje, na katerega moramo odgovoriti v procesu sprejemanja poslovnih odločitev.
- *Regresijska analiza*: regresijska analiza je kvantitativna statistična metoda, ki omogoča preučevanje odnosov (poveznosti in odvisnosti) med spremenljivkami. Regresijska analiza predstavlja zelo močno orodje v različnih raziskovalnih situacijah.
- *Faktorska analiza*: je nepogrešljivo orodje pri analizi večdimenzionalnih spremenljivk, s katerimi pogosto opisujemo značilnosti statističnih enot na področju ekonomskih in poslovnih ved oziroma v družboslovju na splošno.
- *Napovedovanje vrednosti v časovni vrsti z uporabo trenda in sezonske komponente*: v tem poglavju analiziramo, kako napovedovati vrednosti spremenljivke v prihodnjih časovnih enotah s pomočjo trenda in sezonske komponente.

Uporaba tega gradiva omogoča pridobiti celovito razumevanje statistike in raziskovalnih metod. Študenti se naučijo oblikovati raziskovalne probleme, zbirati in analizirati podatke ter uporabljati različne statistične tehnike. Sposobnost uporabe statističnih metod in raziskovalnih tehnik je vse bolj cenjena v različnih disciplinah, od ekonomije in trženja do zdravstva in družbenih ved. Zahvaljujoč tem veščinam so diplomanti gospodarskega inženirstva boljše pripravljene na izzive, ki jih prinaša današnji hitrorastoči podatkovni svet.



## 2 Naloge za seminarske vaje

### 2.1 Formulacija raziskovalnega problema

Znanstveno raziskovanje vključuje sistematičen proces, ki je usmerjen v objektivnost in zbiranje obsežnih informacij za analizo, s čimer raziskovalcu omogoča, da pride do utemeljenih zaključkov. Raziskava se začne z jasno formulacijo raziskovalnega problema. To je ključni korak, ki določa smer in obseg raziskave (Martin and Bridgmon, 2012). Brez jasno definiranega raziskovalnega problema raziskava tvega, da bo neusmerjena, neučinkovita in morda ne bo prinesla smiselnih ali uporabnih rezultatov. Raziskovalni problem je jasno in natančno formulirano vprašanje ali niz vprašanj, ki jih raziskovalec namerava razrešiti z analizo empiričnih podatkov. Ta problem mora biti dovolj specifičen, da omogoča osredotočen pristop, hkrati pa dovolj širok, da zajame relevantne vidike raziskane tematike (Pardede, 2018; Boudah, 2019).

Prvi korak v procesu raziskovanja vključuje prepoznavanje problema ali formulacijo raziskovalnega vprašanja. Ko je problem jasno identificiran, mora raziskovalec pridobiti poglobljeno razumevanje preučevane teme. To doseže s pregledom relevantne literature. Ta korak zagotavlja temeljno znanje o področju problema in raziskovalcu omogoča seznanitev z dosedanjimi študijami, njihovimi metodologijami in ugotovitvami (Boncz, 2015). Na primer, pri raziskovanju uporabe umetne inteligence za izboljšanje uporabniške izkušnje strank bi pregled literature lahko razkril pomembne študije o vplivu algoritmov strojnega učenja na personalizacijo storitev in povečanje zadovoljstva strank. Dodatno bi raziskovalec lahko našel članke, ki obravnavajo etične dileme in varnostne izzive, povezane

z zbiranjem in analizo podatkov strank. Te informacije omogočajo raziskovalcu celovit vpogled v problematiko, pomagajo prepoznati ključne izzive in možnosti, ki jih umetna inteligenca ponuja, ter oblikovati učinkovite strategije za njeno implementacijo.

Naslednji korak se nanaša na pojasnitev problema. Pogosto je problem, identificiran v prvem koraku postopka, preveč obsežen ali splošen. V tem koraku raziskovalec natančneje opredeli in zoži obseg raziskave. To je mogoče storiti šele po temeljitem pregledu literature, saj spoznanja, pridobljena med pregledom, raziskovalcu pomagajo pri razjasnitvi in specifikaciji raziskovalnega projekta (Creswel, 2014). Na primer, v kontekstu umetne inteligence bi lahko raziskovalec začetno identificiral problem kot »izboljšanje strankinih izkušenj s pomočjo umetne inteligence«. Ta tema je obsežna in lahko vključuje različne vidike, kot so personalizacija storitev, avtomatizacija procesov, varovanje zasebnosti in etične dileme. Vseh teh področij ni mogoče obravnavati v eni sami raziskavi; zato je nujno, da raziskovalec problem bolj specifično opredeli. Po pregledu literature se raziskovalec odloči, da bo raziskava osredotočena na ugotavljanje, kako lahko algoritmi strojnega učenja personalizirajo uporabniško izkušnjo v spletnih trgovinah. Ta cilj je veliko bolj omejen in raziskovalno ciljno usmerjen kot prvotno široko zastavljen problem. Nato sledi jasna opredelitev izrazov in konceptov. Izrazi in koncepti, ki se uporabljajo v izjavi o namenu študije ali v opisu raziskave, morajo biti jasno opredeljeni, saj to poveča razumevanje in zmanjšuje možnost nesporazumov med bralci raziskave. Izrazi in pojmi pogosto nosijo različne pomene, odvisno od konteksta in bralca. Da bi preprečili zmedo glede pomena uporabljenih izrazov in fraz, jih mora raziskovalec natančno definirati. V kontekstu umetne inteligence bi tako lahko koncept učinkovitosti algoritma vključeval več dimenzij, kot so hitrost, natančnost in poraba virov. Za potrebe raziskave je treba te dimenzije jasno razločiti in opredeliti. V raziskavi, ki se osredotoča na izboljšanje strankinih izkušenj z uporabo algoritmov strojnega učenja, bi lahko natančnost algoritma opredelili kot odstotek pravilno identificiranih preferenc strank na podlagi njihovih preteklih nakupov. Ta specifična opredelitev omogoča raziskovalcu, da natančneje meri in ocenjuje uspešnost algoritma, kar olajša zbiranje relevantnih podatkov in razlago rezultatov raziskave. Hkrati jasna opredelitev terminov zagotavlja, da so koncepti razumljivi za bralca.

Raziskovalec prav tako mora določiti populacijo. Na primer, če raziskovalec namerava preučevati določeno skupino ljudi v skupnosti, lahko izbere določeno starostno skupino, spol, ljudi z določenega geografskega območja ali pripadnike določene etnične skupine. Obstoječi raziskovalni problem in namen študije pomagata raziskovalcu identificirati specifično populacijo, ki bo vključena v študijo (Boncz, 2015). V kontekstu umetne inteligence, na primer, če je cilj raziskave analiza vpliva algoritmov strojnega učenja na

personalizacijo uporabniških izkušenj v spletnem trgovskem okolju, bi bila določena populacija lahko spletni kupci določene trgovine ali platforme. Opredelitev te populacije omogoča raziskovalcu, da zoži študijo z obsežne na bolj obvladljivo populacijo.

Dejanska izvedba študije začne z zbiranjem podatkov. Ta korak je ključen za pridobivanje informacij, ki so potrebne za odgovor na zastavljeno raziskovalno vprašanje. Zbiranje podatkov lahko poteka na različne načine – preko anket, vprašalnikov, intervjujev, opazovanj, ali zbiranja podatkov iz obstoječe literature (Boncz, 2015). V že omenjenem kontekstu umetne inteligence, bi lahko, če raziskava proučuje učinkovitost algoritmov strojnega učenja pri optimizaciji uporabniških izkušenj v spletni trgovini, raziskovalec zbiral podatke o uporabniškem vedenju, kot so čas, preživet na spletni strani, pogostost nakupov in vrste kupljenih izdelkov. Poleg tega se lahko zbirajo tudi podatki o interakcijah uporabnikov z različnimi funkcijami strani, ki jih poganja umetna inteligenca, kot so personalizirana priporočila. Zbrani podatki omogočajo raziskovalcu prehod na zadnji korak procesa, to je analizo podatkov.

V nadaljevanju prikazujemo primer oblikovanja raziskovalnega problema (primer 1). Rešite še primer 2 in primer 3.

V okviru formulacije raziskovalnega problema si pomagajte s pregledom ustreznih člankov v bazi SCOPUS.

Za izbrano temo v razčlembi znotraj področij:

- trajnostni razvoj in
- izgorelost

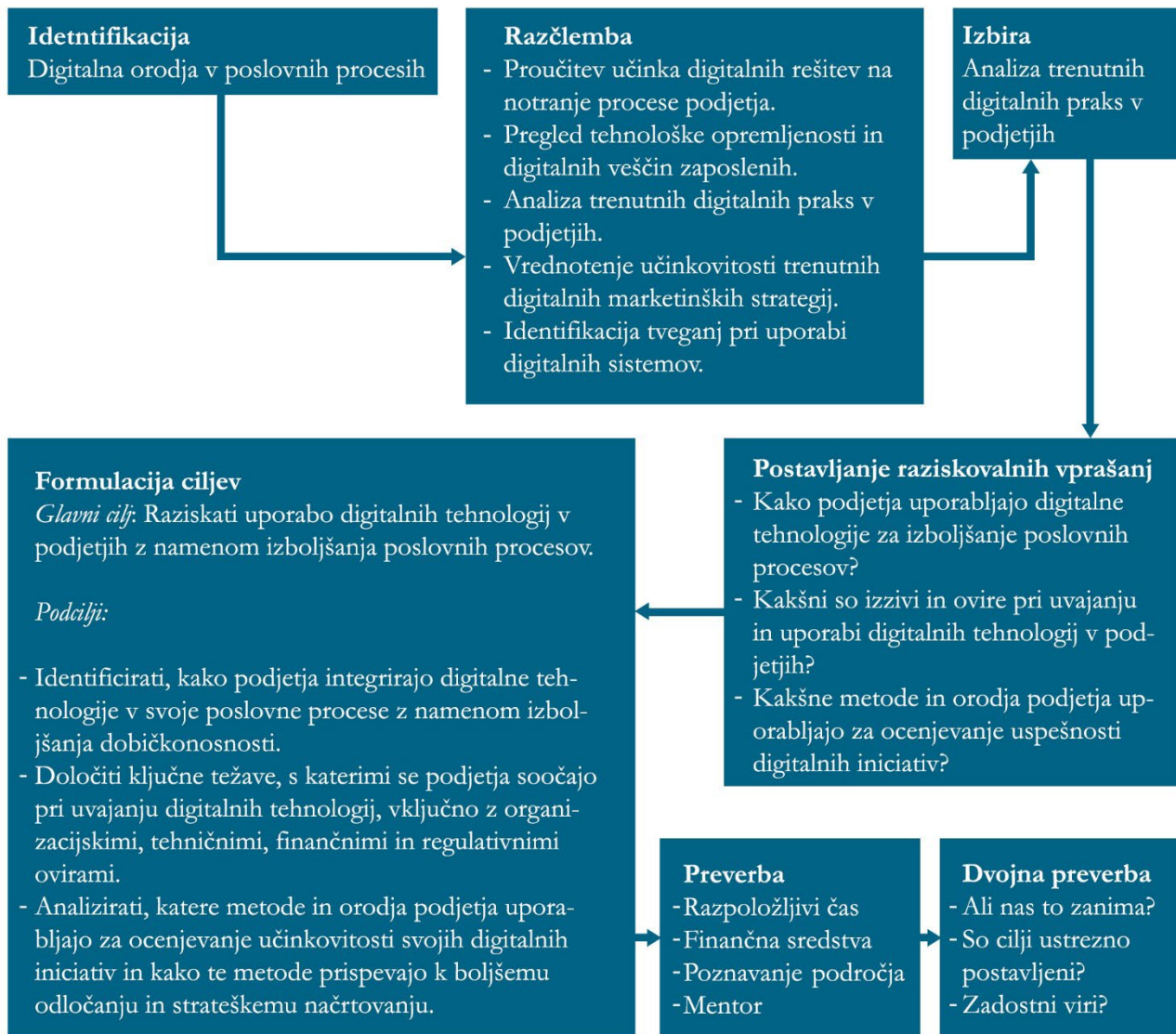
poiščite pet po vaši presoji najbolj relevantnih člankov. Članki naj bodo iz najmanj štirih revij, vsaj dva članka naj bosta objavljena po letu 2020.

Za vsakega od najdenih člankov opredelite:

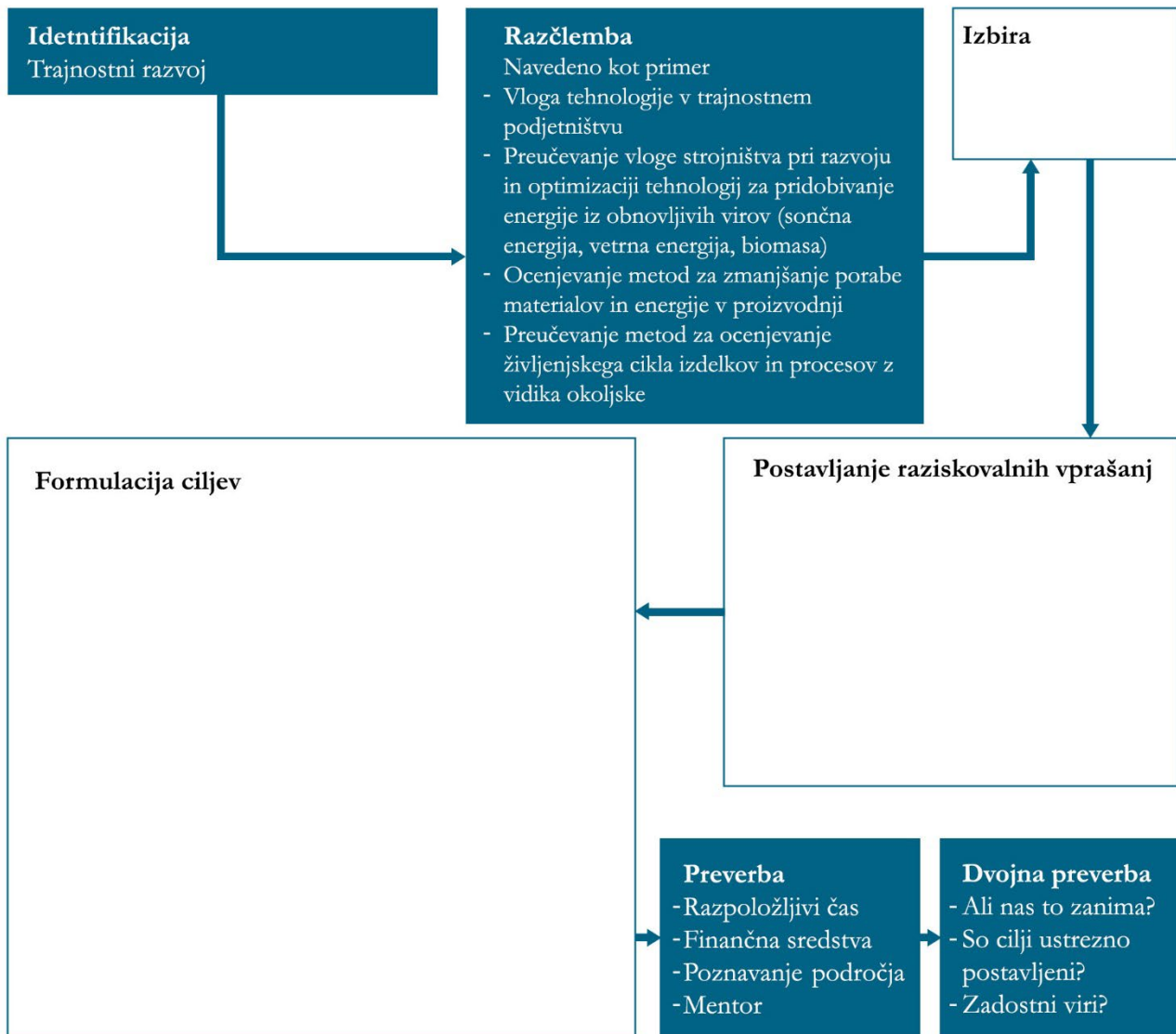
- a) Kakšen je cilj raziskave?
- b) Kako so zbrani podatki?
- c) Kako so podatki obdelani?
- d) Kakšen analitičen pristop je bil uporabljen?
- e) Kako je raziskava utemeljena?



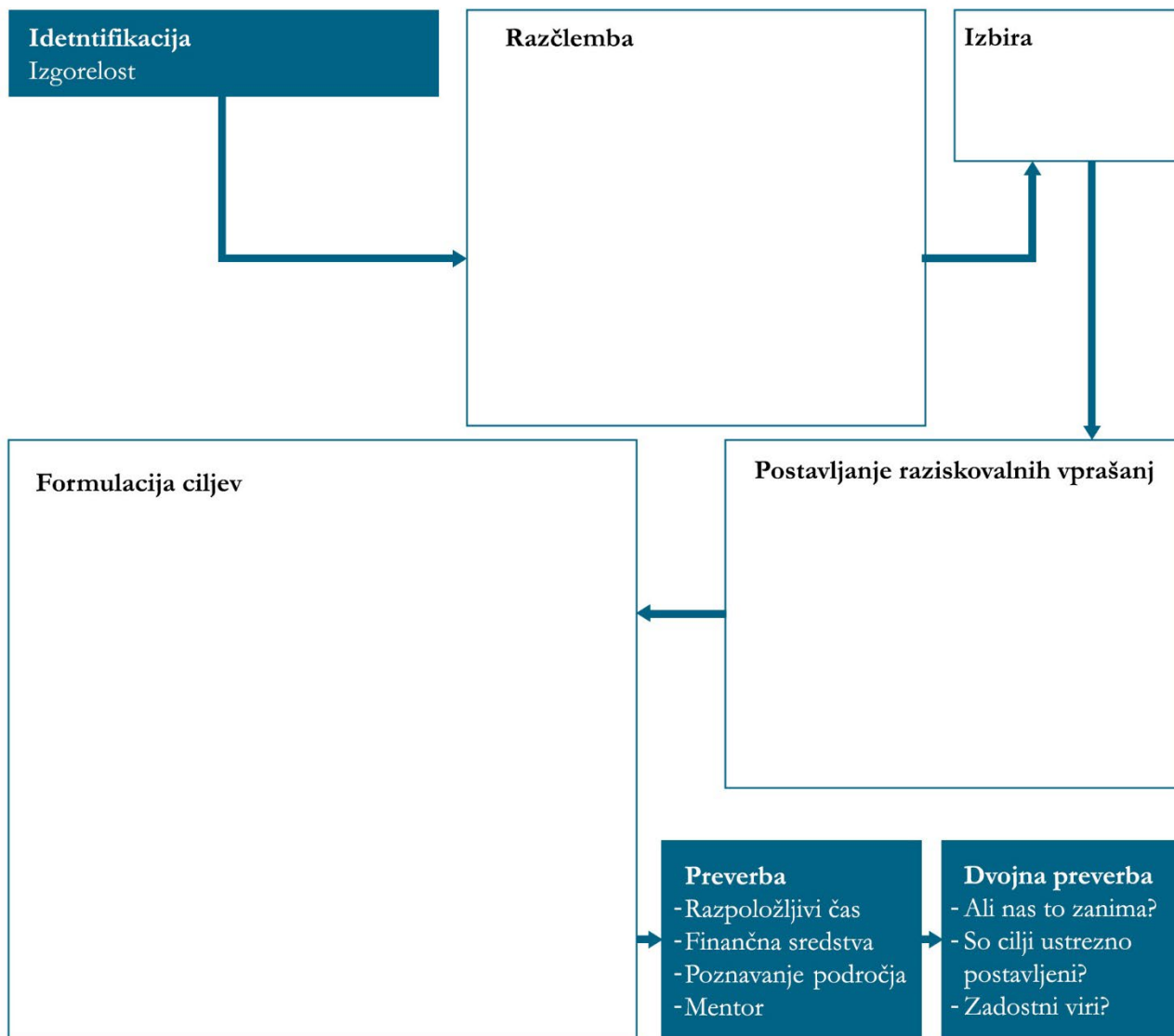
## Primer 1



## Primer 2



## Primer 3



---

---

---

---

---

---

---

---

## 2.2 Urejanje in prikazovanje podatkov

Statistika je multidisciplinarna znanost, ki se osredotoča na kvantitativno analizo naravnih in družbenih pojavov. S pomočjo raznolikih statističnih metod raziskuje vzorce in zakonitosti, ki se pojavljajo v teh fenomenih, kar omogoča boljše razumevanje in napovedovanje trendov (Ghauri idr., 2020).

*Statistična množica* ( $N$ ), uporabljamo tudi izraz populacija, je skupina vseh elementov, ki jih preučujemo v določeni raziskavi. Vsak posamezni element te množice imenujemo statistična enota. Ta množica mora biti jasno opredeljena s specifičnimi značilnostmi, ki določajo, kdo ali kaj spada v to skupino. Statistična množica je lahko zelo široka ali zelo specifična, odvisno od ciljev raziskave (Moore idr., 2016). *Vzorec* ( $n$ ) predstavlja del statistične množice, a kljub temu mora odražati njene značilnosti, da bi lahko zaključke posplošili na celotno statistično množico (Levy in Lemeshow, 2008). Ključnega pomena je reprezentativnost vzorca, saj zagotavlja, da so ugotovitve, pridobljene iz vzorca, veljavne za statistično množico. Vzorec mora biti slučajen, kar pomeni, da ima vsaka statistična enota v statistični množici znano in neničelno verjetnost, da bo izbrana (Tominc in Kramberger, 2007; Lind idr., 2021).

Primer, ki se nanaša na statistično množico in vzorec:

*Statistična množica:* vsa podjetja, ki so registrirana v določeni industriji v Sloveniji.

Ta statistična množica vključuje vsa podjetja, ki delujejo v specifični industrijski panogi, na primer v avtomobilski industriji, in so uradno registrirana in delujejo znotraj geografskih meja Slovenije. Raziskave, ki se osredotočajo na takšno množico, bi lahko preučevale različne vidike, kot so ekonomska učinkovitost, zaposlovanje, proizvodne metode, tehnološke inovacije ipd.

*Primer vzorca iz te množice:* naključno izbranih 100 podjetij iz te množice.

Iz celotne statistične množice vseh podjetij v avtomobilski industriji v Sloveniji se izbere 100 podjetij z uporabo metode slučajnega vzorčenja. Ta vzorec bi moral biti reprezentativen za celotno statistično množico. Raziskovalci bi lahko na tem vzorcu opravili analize, kot so merjenje in primerjava produktivnosti, inovativnosti ali finančne uspešnosti, s ciljem posplošitve zaključkov raziskave na celotno statistično množico.

Tako vzorčenje omogoča, da se izvedejo praktične raziskave brez potrebe po zbiranju podatkov o značilnostih vsakega elementa v celotni statistični množici, kar je pogosto logistično nemogoče ali finančno neizvedljivo. Izbrani vzorec je ključen za zagotavljanje, da so ugotovitve iz raziskave verodostojne in aplikativne na širši skupini.

*Statistična enota* je vsak posamezen element statistične množice.

*Statistična spremenljivka* je lastnost, ki jo preučujemo pri posamezni statistični enoti.

Vrste statističnih spremenljivk:

*Številске spremenljivke*: te lahko zavzemajo numerične vrednosti. Razdelijo se na:

- zvezne spremenljivke: spremenljivke, ki lahko zavzemajo katerokoli vrednost na intervalu oziroma številski premici (npr. teža, višina, temperatura);
- nezvezne ali diskretne spremenljivke: spremenljivke, ki zavzemajo samo določene, najpogosteje celoštevilčne vrednosti (npr. število otrok v družini, število avtomobilov v gospodinjstvu).

*Opisne spremenljivke*: za opisne spremenljivke velja, da lahko njihove vrednosti izražamo le z besedami (na primer spol, rojstni kraj, izobrazba).

Statistični podatki so temeljni elementi za analizo in razumevanje različnih fenomenov, ki jih lahko kategoriziramo v tri glavne vrste: časovne, krajevne in stvarne (Tominc in Kramberger, 2007). Vsaka izmed teh kategorij ima svoje specifične značilnosti:

- *Časovne statistične vrste* se osredotočajo na opazovanje sprememb spremenljivk skozi čas. Dinamika sprememb je ključna značilnost te vrste, saj omogoča sledenje in analizo trendov. Za grafični prikaz časovne statistične vrste uporabimo linijski grafik, saj omogoča jasno in intuitivno vizualizacijo podatkov, kot so na primer, rast ali padec prodaje, spreminjanje borznih indeksov skozi različne časovne periode ipd.
- *Krajevne statistične vrste* analizirajo podatke glede na geografsko lokacijo. Ti podatki so ključni za proučevanje regionalnih razlik, distribucije virov, demografskih trendov ali epidemioloških vzorcev. Grafični prikazi za krajevne statistične vrste vključujejo stolpčne grafike, ki prikazujejo, kako se spremenljivke razlikujejo med različnimi lokacijami oz. krajevnimi območji.

- *Stvarne statistične vrste* zajemajo analizo konkretnih predmetov ali pojavov, neodvisno od časa ali prostora. V to kategorijo spadajo na primer analize proizvodnih procesov, pregledi kakovosti izdelkov, ocene učinkovitosti storitev ipd. Pogosto uporabljeni grafični prikazi za stvarne statistične vrste vključujejo histograme in točkovne diagrame, ki pomagajo pri analizi porazdelitev in razmerij med različnimi stvarnimi spremenljivkami.

Vrednosti številske spremenljivke lahko uredimo v razrede frekvenčne porazdelitve, ki jih grafično prikazujemo s *frekvenčnimi histogrami* ali *poligoni*. Najprej pa je potrebno še ugotoviti, ali imamo opravka s porazdelitvijo, ki ima enako ali različno široke razrede. V primeru, ko imamo nezvezne meje, naredimo še popravek za zveznost mej (na levi strani odštejemo vrednosti od 0,5 in na desni strani prištejemo vrednostim 0,5). *Frekvenčni histogram* je grafikon frekvenčne porazdelitve numeričnih spremenljivk v pravokotnem koordinatnem sistemu s pravokotniki, ki se dotikajo drug drugega. Širina pravokotnika odraža širino razreda, medtem ko višina pravokotnika običajno odraža frekvenco razreda. Vendar navedeno lahko storimo le v primeru, ko imamo enako široke razrede. Če so razredi različno široki, je potrebno izračunati še gostoto frekvence, saj je višina pravokotnika določena z gostoto frekvence. *Frekvenčni poligon* je linijski diagram. Na abscisno os narišemo sredine razredov in nad njimi točke v višini frekvenc razredov (ali gostote frekvence) ter točke povežemo z daljicami. Če analiziramo vrednosti opisne spremenljivke, pa le-te uredimo v skupine na osnovi možnih vrednosti opisne spremenljivke (Tominc in Kramberger, 2007).

*Frekvenca* ( $f_k$ ) nam pove, kako pogosto se pojavlja vrednost spremenljivke, ki je po vrednosti znotraj mej  $k$ -tega razreda (med spodnjo in zgornjo mejo  $k$ -tega razreda). Na primer: bila je izvedena raziskava, v kateri so kupci ocenili svoje zadovoljstvo z novim izdelkom na lestvici od 1 do 10. Rezultati so bili nato razdeljeni v frekvenčne razrede, da bi bolje razumeli splošno zadovoljstvo strank. Podatki so bili razporejeni tako: 5 kupcev je dalo oceno od 9 do pod 10 (visoko zadovoljstvo), 15 kupcev je dalo oceno od 7 do pod 9 (srednje zadovoljstvo), 20 kupcev je dalo oceno od 5 do pod 7 (zmerno zadovoljstvo), 10 kupcev je dalo oceno od 1 do pod 5 (nizko zadovoljstvo). V tem primeru število kupcev, katerih ocene zadovoljstva padejo znotraj meja posameznega razreda, predstavlja frekvenco tega razreda. Ta pristop omogoča, da podjetje učinkovito analizira strankino dožemanje izdelka in identificira področja za izboljšave glede na skupne povratne informacije.

*Kumulativna frekvenca* ( $F_k$ ) za posamezni frekvenčni razred pove, koliko statističnih enot ima vrednost spremenljivke enako ali nižje od zgornje meje danega razreda. *Relativna frekvenca* izraža razmerje med številom statističnih enot, ki spadajo v določen frekvenčni razred, in skupnim številom vseh statističnih enot v podatkovnem nizu. Ta delež ali odstotek nam pove, kako pomemben ali prevladujoč je določen razred v okviru celotnega podatkovnega niza. Grafično relativno frekvenco najpogosteje prikazujemo s strukturnim stolpcem, v katerem prikažemo strukturne odstotke. Drugi grafični prikazi so še strukturni krog, strukturni polkrog, strukturni kvadrat.



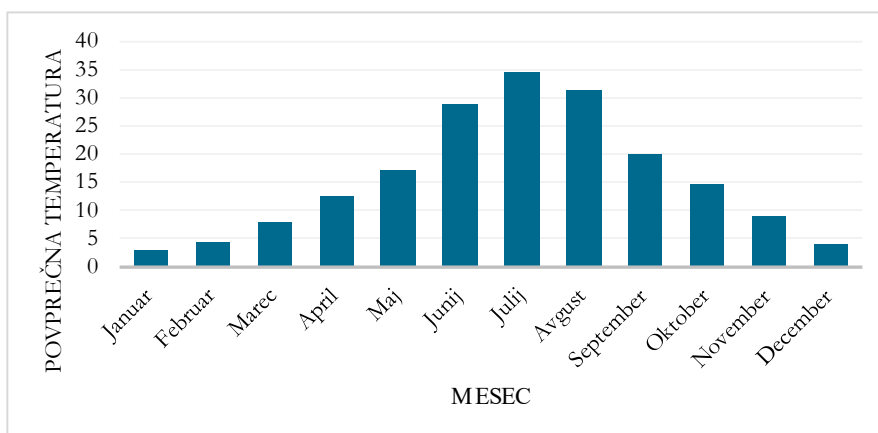
## Primer rešene naloge

V preglednici so podani podatki o povprečni temperaturi za določen kraj v letu 2022 po posameznih mesecih:

Mesec	Povprečna temperatura (°C)
Januar	3,2
Februar	4,5
Marec	8,1
April	12,6
Maj	17,2
Junij	26,8
Julij	34,5
Avgust	28,3
September	20,0
Oktober	14,7
November	8,9
December	5,1

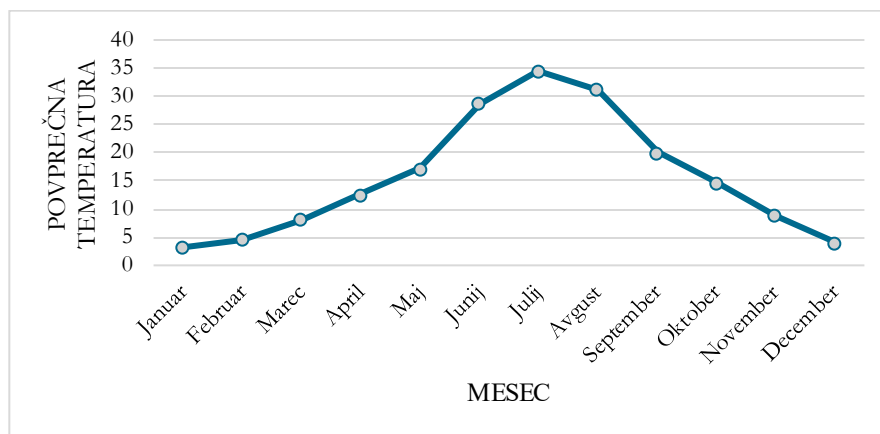
Statistično vrsto za povprečno mesečno temperaturo za določen kraj v letu 2022 prikažite grafično.

Prikaz s stolpci:





Linijski grafikon:



Prikazana je časovna statistična vrsta, saj so vrednosti spremenljivke urejene po mesecih.



## Primer rešene naloge

Študente dodiplomskega študijskega programa na Ekonomsko-poslovni fakulteti v Mariboru smo vprašali, koliko ur na teden posvetijo učenju pri določenem predmetu. Podatki so podani v preglednici:

Čas učenja v urah	Število študentov
Od 5 do pod 10	20
Od 10 do pod 15	35
Od 15 do pod 20	22
Od 20 do pod 25	18
Od 25 do pod 30	13
<b>Skupaj</b>	<b>108</b>

**a) Opredelite statistično enoto ter spremenljivko.**

Statistična enota je vsak posamezen element statistične množice, kar pomeni, da je v našem primeru statistična enota en študent dodiplomskega študijskega programa na Ekonomsko-poslovni fakulteti v Mariboru. Statistična spremenljivka opisuje lastnost statistične enote in je v našem primeru čas učenja pri določenem predmetu v urah (številska, zvezna spremenljivka).

**b) Izračunajte kumulativne člene frekvenčne porazdelitve.**

*Kumulativni členi frekvenčne porazdelitve*

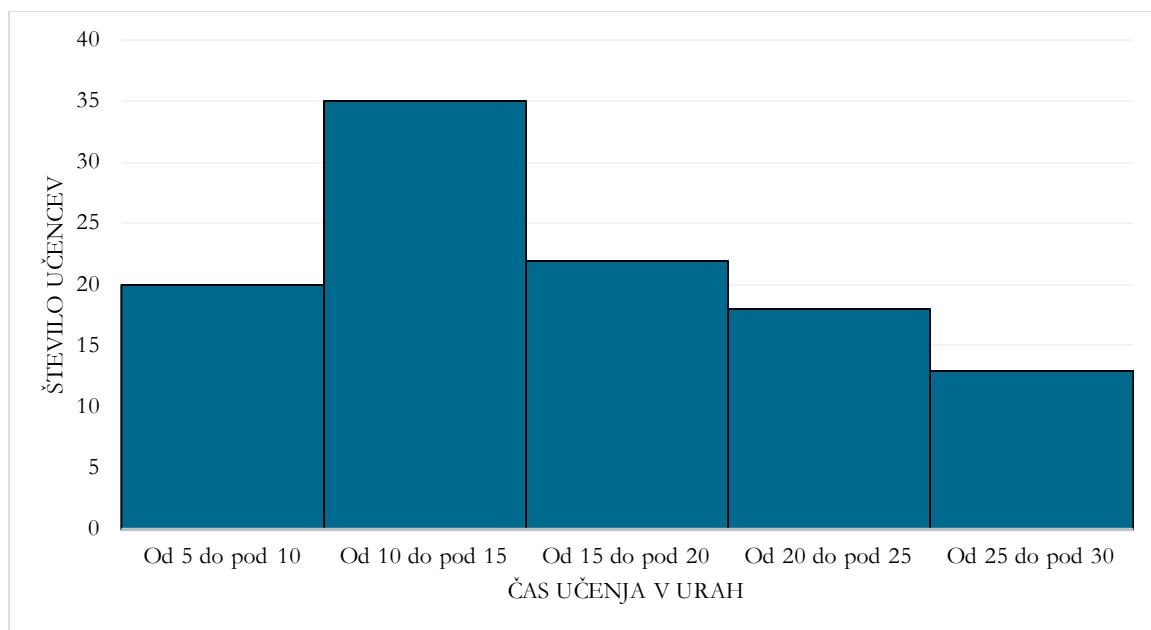
$$F_1 = f_1, \quad F_k = F_{k-1} + f_k \quad \text{za } k = 2, 3, \dots, r \quad r = \text{število razredov v frekvenčni porazdelitvi}$$

Čas učenja v urah	Število študentov ( $f_k$ )	Kumulativna frekvenčna porazdelitev ( $F_k$ )
Od 5 do pod 10	20	20
Od 10 do pod 15	35	20 + 35 = 55
Od 15 do pod 20	22	55 + 22 = 77
Od 20 do pod 25	18	77 + 18 = 95
Od 25 do pod 30	13	95 + 13 = 108
<b>Skupaj</b>	108	

**c) Frekvenčno porazdelitev prikažite grafično.**

Grafični prikaz (frekvenčni histogram):

Frekvenčna porazdelitev z enako širokimi razredi: v frekvenčnem histogramu na y osi prikažemo frekvence ( $f_k$ ).





## Naloga 1

V preglednici so podatki o številu objavljenih znanstvenih člankov po raziskovalnih področjih in vrsti publikacije v določenem raziskovalnem inštitutu v tekočem letu:

Področje raziskav	Članki v konferenčnih zbornikih	Članki v znanstvenih revijah
Biotehnologija	89	156
Računalništvo	312	205
Fizika	215	198
Sociologija	128	276

- Opreделите statistično enoto in statistične spremenljivke.
- Izpišite statistično vrsto objavljenih člankov v konferenčnih zbornikih po raziskovalnih področjih. Kako imenujemo takšno statistično vrsto? Statistično vrsto prikažite grafično.
- Grafično prikažite relativno strukturo objavljenih člankov v znanstvenih revijah po raziskovalnih področjih.



## Naloga 2

V podjetju, ki prodaja avtomobile, so zbrali podatke o številu prodanih avtomobilov v prvem četrtletju leta 2023. V podjetju ponujajo avtomobile v različnih cenovnih razredih. Podatki o prodanem številu avtomobilov glede na cenovni razred (v tisočih dolarjev) so naslednji:

Cenovni razred (v 1.000 \$)	Število prodanih avtomobilov
Od 21 do 30	36
Od 31 do 40	30
Od 41 do 50	25
Od 51 do 60	14
Od 61 do 70	6
<b>Skupaj</b>	<b>111</b>

- Opreделите statistično množico, statistično enoto in spremenljivko ter njene značilnosti.
- Določite spodnje in zgornje meje razredov ter določite širino razredov.
- Prikažite strukturo prodanih avtomobilov glede na cenovni razred (v 1.000 \$) v strukturnem stolpcu in strukturnem krogu.
- Frekvenčno porazdelitev prikažite grafično.

**Naloga 3**

V srednje velikem podjetju A je vodja kadrovske službe zbral podatke o starosti zaposlenih, da bi bolje razumel starostno strukturo zaposlenih. Starost je zabeležena v letih:

Starost zaposlenih	Zaposleni
Od 20 do pod 30	16
Od 30 do pod 40	20
Od 40 do pod 50	28
Od 50 do pod 60	23
Od 60 do pod 70	11
<b>Skupaj</b>	<b>98</b>

- Opreделите statistično množico, statistično enoto ter spremenljivko.
- Izračunajte kumulativne člene frekvenčne porazdelitve.
- Koliko odstotkov zaposlenih je starih od 30 do 40 let?
- Koliko odstotkov zaposlenih je starih do 60 let?
- Frekvenčno porazdelitev ter kumulativno frekvenčno porazdelitev prikažite grafično.



## Naloga 4

Za 115 zaposlenih, ki smo jih opazovali glede na število dni dopusta, ki so ga vzeli v preteklem letu (diskretna spremenljivka) v podjetju  $X$ , so na razpolago naslednji podatki:

$$y_{1,\min} = 5 \text{ dni} \qquad y_{6,\max} = 35 \text{ dni} \qquad r = 6 \qquad i = 5$$

$$F_1 = 6 \qquad F_2 = 23 \qquad F_3 = 61 \qquad F_4 = 94 \qquad F_5 = 106 \qquad F_6 = 115$$

- Opreделите statistično enoto in spremenljivko ter njene značilnosti.
- Na osnovi danih podatkov sestavite frekvenčno porazdelitev in jo grafično prikažite.
- Koliko odstotkov zaposlenih je imelo od 15 do 20 dni dopusta?
- Koliko odstotkov zaposlenih je imelo do 25 dni dopusta?



## Naloga 5

Razpolagamo s podatki o produktivnosti zaposlenih (na lestvici od 1 do 60) za 40 zaposlenih:

10	12	14	15	15	16	17	17	18	19
20	20	22	23	23	24	25	26	27	28
30	30	31	32	33	34	35	36	37	38
39	40	42	43	45	49	51	54	55	57

- Opreделите statistično množico, statistično enoto ter spremenljivko, njene značilnosti in zalogo vrednosti spremenljivke.
- Sestavite frekvenčno porazdelitev pri pogojih:  $y_{1,\min} = y_{\min}$ ;  $i_{1-6} = 7$ ,  $r = 6$ , meje razredov so podane nezvezno.
- Frekvenčno porazdelitev ter kumulativno frekvenčno porazdelitev prikažite grafično.





## Naloga 6

Razpolagamo s podatki o številu dni bolniških zaostankov zaposlenih v podjetju:

1	1	1	1	2	2	3	3	3
4	5	5	5	6	7	8	9	9
9	10	11	12	12	13	14	15	15
16	17	18	19	20	21			

- Opreделите statistično množico, statistično enoto ter spremenljivko in njene značilnosti
- Sestavite frekvenčno porazdelitev pri pogojih:  $y_{1,\min} = y_{\min}$ ;  $i_{1-3} = 2$ ,  $i_{4-5} = 5$ ,  $r = 5$ , meje razredov so podane nezvezno.
- Frekvenčno porazdelitev ter kumulativno frekvenčno porazdelitev prikažite grafično.
- Koliko odstotkov zaposlenih je imelo od 7 do 9 dni bolniških izostankov?
- Koliko odstotkov zaposlenih je imelo do 15 dni bolniških izostankov?



## Naloga 7

Podani so podatki za 430 podjetij v Sloveniji glede na število zaposlenih:

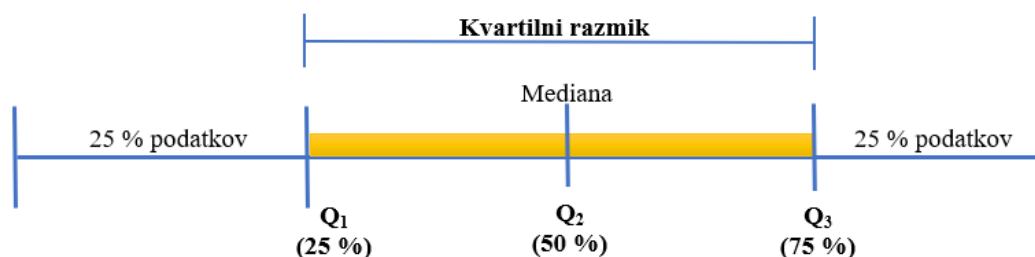
Število zaposlenih	Število podjetij
Od 1 do 40	90
Od 41 do 80	123
Od 81 do 120	145
Od 121 do 160	72
<b>Skupaj</b>	<b>430</b>

- Določite spodnje in zgornje meje razredov ter širino razredov. Narišite poligon.
- Prikažite strukturo podjetij glede na število zaposlenih v strukturnem stolpcu in krogu.

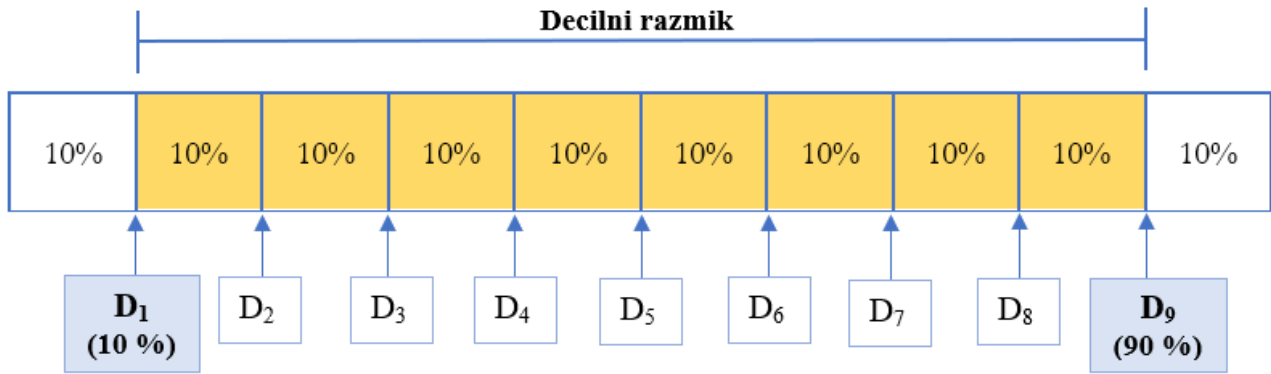
## 2.3 Deskriptivna statistika

*Deskriptivna statistika ali opisna statistika* kvantitativno opisuje ali povzema značilnosti nabora podatkov. Opisne statistične analize (angl. Descriptive statistics) so metode, s katerimi sistematično raziskujemo strukturo in značilnosti zbranih ali opazovanih podatkov (Kaliyadan in Kulkarni, 2019). Namen teh analiz je jasno in natančno predstaviti informacije, kot so velikost vzorca, odstotki, frekvence, minimalne in maksimalne vrednosti, variacijski razmik, povprečje, mediana, modus, standardni odklon, varianca. Opisne statistične analize ne temeljijo na statističnem sklepanju in ne posplošujejo ugotovitev z vzorca na celotno statistično množico, ampak služijo zgolj za opisovanje zbranih podatkov. Te analize so še posebej koristne v začetnih fazah raziskave za ugotavljanje osnovnih lastnosti podatkov. Med najpomembnejše **mere osrednje tendence** uvrščamo *aritmetično sredino*, ki predstavlja povprečno vrednost nabora podatkov, in *modus*, ki je najpogostejša vrednost v naboru podatkov, ter *mediano*, srednjo vrednost spremenljivke, od katere ima 50 % enot manjše ali enake vrednosti, 50 % enot pa večje vrednosti. Te mere nam pomagajo razumeti osrednjo vrednost distribucije podatkov in so ključne pri opisni statistiki (Holmes idr., 2018; Kaliyadan in Kulkarni, 2019).

Najpomembnejše **mere variabilnosti** zajemajo *variacijski razmik*, *kvartilni in decilni razmik*, *varianco* in *standardni odklon* ter *koeficient variabilnosti*. *Variacijski razmik* je razlika med največjo in najmanjšo vrednostjo spremenljivke. *Kvartilni razmik* (slika 1) je mera variabilnosti, ki označuje razpon srednje velikih vrednosti (50 % podatkov na sredini ranžirne vrste ali frekvenčne porazdelitve). To je razlika med tretjim kvartilom ( $Q_3$ ) in prvim kvartilom ( $Q_1$ ). Prvi kvartil ( $Q_1$ ; 25 %) je vrednost, pod katero je 25 % podatkov. Mediana je  $Q_2$ . Tretji kvartil ( $Q_3$ ; 75 %) je vrednost, pod katero se nahaja 75 % podatkov. Kvartilni razmik zajema 50 % podatkov na sredini ranžirne vrste (slika 1). *Decilni razmik* (slika 2) pa je razlika med devetim ( $D_9$ ) in prvim decilom ( $D_1$ ) (srednjih 80 % vrednosti). *Varianca* je povprečje kvadratov odklonov od aritmetične sredine. Ker je enota mere kvadrat enote mere osnovne spremenljivke, uporabljamo običajno kvadratni koren variance  $\sigma = \sqrt{\sigma^2}$ , ki ga imenujemo *standardni odklon* (Barde in Barde, 2012; Thukral idr, 2019; Frost, 2020).



Slika 1: Kvartilni razmik

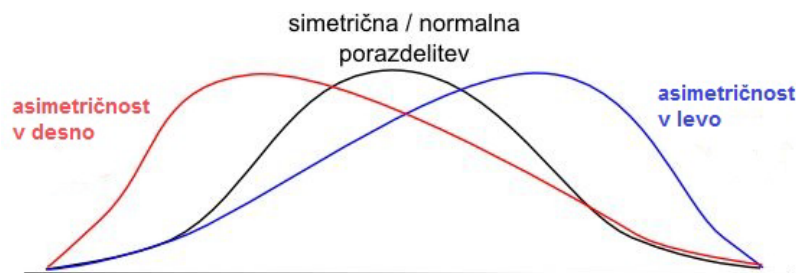


Slika 2: Decilni razmik

Med mere asimetrije in sploščenosti uvrščamo *koeficient asimetrije* in *koeficient sploščenosti*.

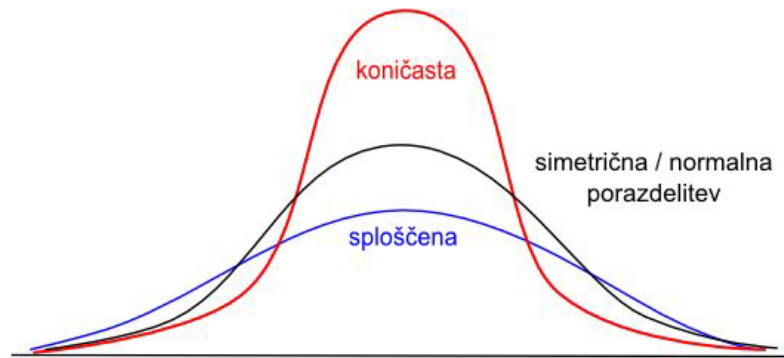
*Asimetrične porazdelitve* (slika 3) so lahko *asimetrične v desno* (pozitivna asimetrična porazdelitev), zanje je značilna večja gostitev pri manjših vrednostih spremenljivke, ali *asimetrične v levo* (negativna asimetrična porazdelitev) in je zanje značilna večja gostitev vrednosti pri večjih vrednostih spremenljivke. Koeficient asimetrije je manjši od 0, če je za porazdelitev spremenljivke značilna asimetrija v levo, pri asimetriji v desno je koeficient asimetrije večji od 0. Koeficient asimetrije pri večini empiričnih porazdelitev lahko zavzame vrednost med  $-3$  in  $+3$  (Artenjak, 2003).

*Sploščenost porazdelitve* (slika 4) primerjamo z normalno porazdelitvijo, za katero rečemo, da je normalno sploščena. Če je porazdelitev bolj koničasta od normalne porazdelitve, rečemo, da je *porazdelitev koničasta*. Če je porazdelitev bolj sploščena od normalne, rečemo, da je *porazdelitev sploščena*. Za koeficient sploščenosti je značilno, da kadar je le-ta večji od 0, nakazuje na koničasto porazdelitev in v primeru, ko je koeficient sploščenosti manjši od 0, na sploščeno porazdelitev. Pri teoretični normalni porazdelitvi, ki jo bomo obravnavali v nadaljevanju, sta tako koeficienta asimetričnosti in sploščenosti enaka 0 (Freedman idr., 2007; Evans idr., 2010).



Slika 3: Asimetričnost porazdelitve

Vir: Freedman idr., 2007; Evans idr., 2010



**Slika 4: Sploščenost porazdelitve**

Vir: Freedman idr., 2007; Evans idr., 2010

Poznamo še *inferenčno statistiko*, ki se uporablja za sklepanje o značilnostih populacije na podlagi slučajnega vzorca podatkov. Vključuje metode, kot so intervali zaupanja, analizo variance (ANOVA), t-test, Hi-hvadrat test, regresijska analiza ipd. Med inferenčno statistiko sodijo tiste analize, s katerimi ocenjujemo parametre in preverjamo domneve.



### Primer rešene naloge

V podjetju X so zbrali podatke o razdalji od stalnega bivališča do delovnega mesta (v km) svojih zaposlenih. Podani so dnevni podatki o prevoženih kilometrih za šest zaposlenih v podjetju X, ki se vozijo na svoje delovno mesto: 33 km, 8 km, 12 km, 20 km, 24 km, 5 km.

- Izračunajte in pojasnite povprečno vrednost za dnevno prevožene kilometre zaposlenih, ki se vozijo na svoje delovno mesto.
- Izračunajte in pojasnite standardni odklon za dnevno prevožene kilometre zaposlenih, ki se vozijo na svoje delovno mesto.
- Pojasnite mero variabilnosti, ki upošteva variabilnost za 80 % zaposlenih, ki se glede na dnevno prevožene kilometre na delovno mesto razvrščajo na sredino ranžirne vrste.

a)

Za izračun povprečne vrednosti dnevno prevoženih kilometrov zaposlenih, ki se vozijo na svoje delovno mesto, uporabimo enačbo (*aritmetična sredina iz nerazvrščenih vrednosti*):

$$\bar{y} = \frac{1}{N}(y_1 + y_2 + \dots + y_N) = \frac{1}{N} \sum_{i=1}^N y_i$$

$$\bar{y} = \frac{1}{6} \cdot (33 + 8 + 12 + 20 + 24 + 5) = 17 \text{ km}$$

Odgovor: Povprečna vrednost dnevno prevoženih kilometrov šestih zaposlenih, ki se vozijo na svoje delovno mesto, znaša 17 km.

b)

Izračunamo varianco po enačbi (*varianca iz nerazvrščenih vrednosti*):

$$VAR = \sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2$$

$$\sigma^2 = \frac{1}{6} \cdot [(33 - 17)^2 + (8 - 17)^2 + (12 - 17)^2 + (20 - 17)^2 + (24 - 17)^2 + (5 - 17)^2] = 94 \text{ km}^2$$

Izračunamo standardni odklon:

$$SD = \sigma = \sqrt{VAR} = \sqrt{\sigma^2},$$

$$\sigma = \sqrt{94} = 9,7 \text{ km}$$

Standardni odklon je kvadratni koren iz varince in znaša 9,7 km.

c)

Podatke uredimo v ranžirno vrsto:

<b>R<sub>i</sub></b>	1	2	3	4	5	6
<b>y<sub>i</sub></b>	5	8	12	20	24	33

Decilni razmik:  $D = D_9 - D_1$

$D_1$ : 10 %

Uporabimo skupino enačb z naslovom *kvantili iz nerazvrščenih vrednosti*:

$$P_i = 0,1$$

$$R_i = N \cdot P_i + 0,5$$

$$R_i = 6 \cdot 0,1 + 0,5 = 1,1$$

$$R_0 \leq R_i < R_1$$

$$R_0 = 1 \leq R_i = 1,1 < R_1 = 2$$

$$y_0 \leq y_i < y_1$$

$$y_0 = 5 \leq y_i < y_1 = 8$$

$$y_i = y_0 + \frac{R_i - R_0}{R_1 - R_0} \times (y_1 - y_0)$$

$$y_i = 5 + \frac{1,1 - 1}{2 - 1} \cdot (8 - 5) = 5,3 \text{ km}$$

Odgovor: 10 % zaposlenih dnevno prevozi na svoje delovno mesto 5,3 km ali manj.

$$D_9: 90 \%$$

$$P_i = 0,9$$

$$R_i = 6 \cdot 0,9 + 0,5 = 5,9$$

$$R_0 \leq R_i < R_1$$

$$R_0 = 5 \leq R_i = 5,9 < R_1 = 6$$

$$y_0 \leq y_i < y_1$$

$$y_0 = 24 \leq y_i < y_1 = 33$$

$$y_i = 24 + \frac{5,9 - 5}{6 - 5} \cdot (33 - 24) = 32,1 \text{ km}$$

Odgovor: 90 % zaposlenih dnevno prevozi na svoje delovno mesto 32,1 km ali manj.

$$D = 32,1 - 5,3 = 26,8 \text{ km}$$

Odgovor: 80 % zaposlenih, ki se glede na dnevno prevožene kilometre na delovno mesto razvrščajo na sredino ranžirne vrste, se med seboj razlikujejo za največ 26,8 prevoženih km.



### Primer rešene naloge

Za pet različnih trgovin imamo podatke o odstotku povečanja prodaje za izdelek X v zadnjem mesecu, ki so ga dosegli s spremembami v trženjski strategiji: 6,2 %; 4,8 %; 7,5 %; 3,1 % in 5,9 %.

- Opreделите statistično enoto, statistično množico in statistično spremenljivko.
- Izračunajte delež standardnega odklona v aritmetični sredini.
- Pojasnite mero variabilnosti, ki upošteva variabilnost za 50 % trgovin, ki se glede na odstotek povečanje prodaje razvrščajo na sredino ranžirne vrste.



b)

Izračunamo koeficient variabilnosti v odstotku:

$$KV\% = \frac{\sigma}{\bar{y}} \times 100$$

Za izračun povprečne vrednosti prodaje (v %) za pet trgovin uporabimo enačbo (aritmetična sredina iz nerazvrščenih vrednosti):

$$\bar{y} = \frac{1}{N}(y_1 + y_2 + \dots + y_N) = \frac{1}{N} \sum_{i=1}^N y_i$$

$$\bar{y} = \frac{1}{5} \cdot (6,2 + 4,8 + 7,5 + 3,1 + 5,9) = 5,5\% \text{ povečanje prodaje}$$

Odgovor: Povprečna vrednost povečane prodaje petih trgovin v zadnjem mesecu znaša 5,5 %.

Izračunamo varianco iz nerazvrščenih vrednosti po enačbi:

$$VAR = \sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2$$

$$\sigma^2 = \frac{1}{5} \cdot [(6,2 - 5,5)^2 + (4,8 - 5,5)^2 + (7,5 - 5,5)^2 + (3,1 - 5,5)^2 + (5,9 - 5,5)^2] = 2,18 \text{ (\% prodaje)}^2$$

Izračunamo standardni odklon po enačbi:

$$SD = \sigma = \sqrt{VAR} = \sqrt{\sigma^2}$$

$$\sigma = \sqrt{2,18} = 1,48 \text{ \% prodaje}$$

$$KV\% = \frac{1,48}{5,5} \cdot 100 = 26,91 \text{ \%}$$

Odgovor: Delež standardnega odklona v aritmetični sredini znaša 26,91 %.

c)

<b>R<sub>i</sub></b>	1	2	3	4	5
<b>y<sub>i</sub></b>	3,1	4,8	5,9	6,2	7,5

Uporabimo enačbo za kvartilni razmik  $Q = Q_3 - Q_1$  (kjer je  $Q_3 = 75\%$  in  $Q_1 = 25\%$ ), vendar je potrebno najprej izračunati:

relativni rang  $P_i$  je znan:  $P_i = 0,25$  (kjer je  $Q_1 = 25\%$ ), zato pri izračunu uporabimo enačbe (*kvantili iz nerazvrščenih vrednosti*):

$$R_i = N \times P_i + 0,5$$

$$R_0 \leq R_i < R_1$$

$$y_i = y_0 + \frac{R_i - R_0}{R_1 - R_0} \times (y_1 - y_0)$$

Izračunamo rang  $R_i$ :

$$R_i = 5 \cdot 0,25 + 0,5 = 1,8$$

Nato določimo vrednosti:

$$R_0 \leq R_i < R_1$$

$$R_0 = 1 \leq R_i = 1,8 < R_1 = 2$$

$$y_0 \leq y_i < y_1$$

$$y_0 = 3,1 \leq y_i < y_1 = 4,8$$

Nato izračunamo vrednost  $y_i$ :

$$y_i = 3,1 + \frac{1,8 - 1}{2 - 1} \cdot (4,8 - 3,1) = 4,46 \text{ \% prodaje}$$

Odgovor: 25 % trgovin je imelo 4,46 % prodaje ali manj, 75 % poslovalnic pa več kot 4,46 % prodaje.

$P_i = 0,75$  (kjer je  $Q_3 = 75\%$ ),

$$R_i = 5 \cdot 0,75 + 0,5 = 4,3$$

$$R_0 \leq R_i < R_1$$

$$R_0 = 4 \leq R_i = 4,3 < R_1 = 5$$

$$y_0 \leq y_i < y_1$$

$$y_0 = 6,2 \leq y_i < y_1 = 7,5$$

Nato izračunamo vrednost  $y_i$ :

$$y_i = 6,2 + \frac{4,3 - 4}{5 - 4} \cdot (7,5 - 6,2) = 6,59 \% \text{ prodaje}$$

Odgovor: 75 % trgovin je imelo 6,59 % prodaje ali manj, 25 % poslovalnic pa več kot 6,59 % prodaje.

Izračun kvartilnega razmika:

$$Q = Q_3 - Q_1 = 6,59 - 4,46 = 2,13\% \text{ točk}$$

Odgovor: 50 % trgovin, ki glede na odstotek povečanja prodaje ležijo na sredini ranžirne vrste, se razlikuje za največ 2,13 odstotnih točk.



### Primer rešene naloge

Podani so podatki za devet zaposlenih o tedenskem številu nadur v podjetju X za leto 2022:

Zaposleni	A	B	C	D	E	F	G	H	I
Tedensko število nadur	8	2	7	5	10	1	12	4	6

- Izračunajte in pojasnite koeficient asimetrije na podlagi mediane. Prav tako pojasnite, kolikšno je tedensko število nadur 50 % zaposlenih, ki so imeli najmanj nadur.
- Pojasnite mero variabilnosti, ki upošteva variabilnost za 80 % zaposlenih, ki se glede na tedensko število nadur razvrščajo na sredino ranžirne vrste.
- Izračunajte, kolikšno število nadur so imeli tisti zaposleni, ki spadajo med 30 % zaposlenih z najmanj nadurami v podjetju.

a)

$R_i$	1	2	3	4	5	6	7	8	9
$y_i$	1	2	4	5	6	7	8	10	12

Uporabimo enačbo za koeficient asimetrije na podlagi mediane:

$$KA_{Me} = \frac{3(\bar{y} - Me)}{\sigma}$$

Izračun mediane:

$$P_i = 0,5$$

$$R_i = N \cdot P_i + 0,5 = 9 \cdot 0,5 + 0,5 = 5$$

$$Me = 6 \text{ nadur}$$

Odgovor: 50 % zaposlenih je imelo tedensko število nadur 6 ali manj, 50 % zaposlenih pa več kot 6 nadur.

$$\bar{y} = \frac{1}{9} \cdot (1 + 2 + \dots + 12) = 6,1 \text{ nadur}$$

$$VAR = \sigma^2 = \frac{1}{N} \sum_{i=1}^N y_i^2 - \bar{y}^2,$$

$$\sigma^2 = \frac{1}{9} \cdot [(1^2 + 2^2 + \dots + 12^2)] - 6,1^2 = \frac{1}{9} \cdot 439 - 37,21 = 11,57 \text{ (nadir)}^2$$

$$\sigma = \sqrt{11,57} = 3,40 \text{ nadur}$$

Izračun koeficienta asimetrije na podlagi mediane:

$$KA_{Me} = \frac{3(\bar{y} - Me)}{\sigma}$$

$$KA_{Me} = \frac{3(6,1 - 6)}{3,40} = 0,088$$

Porazdelitev je asimetrična v desno.

*Ponovitev:*

$\bar{y} > Me$  pomeni, da je porazdelitev asimetrična v desno (več kot polovica vrednosti je manjših od aritmetične sredine).

$\bar{y} < Me$  pomeni, da je porazdelitev asimetrična v levo (manj kot polovica vrednosti je manjših od aritmetične sredine).

$\bar{y} = Me$  pomeni, da je porazdelitev simetrična.

b)

Uporabimo enačbo za decilni razmik:  $D = D_9 - D_1$  (kjer je  $D_9 = 90\%$  in  $D_1 = 10\%$ ), vendar je potrebno najprej izračunati:

Relativni rang  $P_i$  je znan:  $P_i = 0,9$  (kjer je  $D_9 = 90\%$ ), zato pri izračunu uporabimo enačbe (*kvantili iz nerazvrščenih vrednosti*):

$$R_i = N \times P_i + 0,5$$

$$R_0 \leq R_i < R_1$$

$$y_i = y_0 + \frac{R_i - R_0}{R_1 - R_0} \times (y_1 - y_0)$$

Izračunamo rang  $R_i$ :

$$R_i = 9 \cdot 0,9 + 0,5 = 8,6$$

Nato določimo vrednosti:

$$R_0 \leq R_i < R_1$$

$$R_0 = 8 \leq R_i = 8,6 < R_1 = 9$$

$$y_0 \leq y_i < y_1$$

$$y_0 = 10 \leq y_i < y_1 = 12$$

Nato izračunamo vrednost  $y_i$ :

$$y_i = 10 + \frac{8,6 - 8}{9 - 8} \cdot (12 - 10) = 11,2 \text{ nadur}$$

Odgovor: 90 % zaposlenih je imelo 11,2 nadur ali manj, 10 % zaposlenih pa več kot 11,2 nadur.

$P_i = 0,1$  (kjer je  $D_1 = 10\%$ )

$$R_i = 9 \cdot 0,1 + 0,5 = 1,4$$

$$R_0 \leq R_i < R_1$$

$$R_0 = 1 \leq R_i = 1,4 < R_1 = 2$$

$$y_0 \leq y_i < y_1$$

$$y_0 = 1 \leq y_i < y_1 = 2$$

Nato izračunamo vrednost  $y_i$ :

$$y_i = 1 + \frac{1,4 - 1}{2 - 1} \cdot (2 - 1) = 1,4 \text{ nadur}$$

Odg. 10 % zaposlenih je imelo 1,4 nadur ali manj, 90 % zaposlenih pa več kot 1,4 nadur.

Decilni razmik:

$$D = D_9 - D_1 = 11,2 - 1,4 = 9,8 \text{ nadur}$$

Odgovor: 80% zaposlenih, ki glede na število nadur ležijo na sredini ranžirne vrste, se razlikuje za največ 9,8 nadur.

c)

$$D_3 = 30 \%$$

$$P_i = 0,3$$

$$R_i = 9 \cdot 0,3 + 0,5 = 3,2$$

$$R_0 \leq R_i < R_1$$

$$R_0 = 3 \leq R_i = 3,2 < R_1 = 4$$

$$y_0 \leq y_i < y_1$$

$$y_0 = 4 \leq y_i < y_1 = 5$$

Nato izračunamo vrednost  $y_i$ :

$$y_i = 4 + \frac{3,2 - 3}{4 - 3} \cdot (5 - 4) = 4,2 \text{ nadur}$$

Odgovor: Zaposleni, ki spadajo v 30 % zaposlenih z najmanj nadurami v podjetju, so imeli 4,2 nadur ali manj.



## Naloga 8

Za deset srednjih šol imamo podatke o odstotku izboljšanja uspešnosti učencev pri določenem predmetu po uvedbi novega učnega programa:

1,8 %   4,5 %   3,2 %   6,1 %   2,5 %   3 %   4,8 %   5,2 %   2,9 %   7 %

- a) Opredelite statistično enoto, statistično množico, statistično spremenljivko.
- b) Izračunajte variacijski razmik.
- c) Izračunajte povprečen odstotek uspešnosti učencev pri določenem predmetu po uvedbi novega učnega programa.
- d) Izračunajte in pojasnite koeficient variabilnosti v %.
- e) Izračunajte in pojasnite kvartilni razmik.



## Naloga 9

V dveh organizacijah smo opazovali zaposlene glede na čas (v minutah), ki so ga porabili za izdelavo enega izdelka. Podatki so:

Organizacija A:  $N = 12$        $y_i = 26, 38, 45, 22, 33, 29, 34, 41, 40, 39, 43, 30$  minut

Organizacija B:  $N = 730$

Organizacija B	
Poraba časa v minutah	Število zaposlenih
Nad 22 do 26	76
Nad 26 do 30	123
Nad 30 do 34	235
Nad 34 do 38	162
Nad 38 do 42	98
Nad 42 do 46	36
<b>Skupaj</b>	<b>730</b>

- a) Za organizacijo A izračunajte in pojasnite:
- Odstotek zaposlenih, ki so za izdelek porabili manj kot 36 in več kot 40 minut.
  - Koliko časa je za en izdelek porabilo 30 % zaposlenih z največjo (oziroma najmanjšo) porabo časa za en izdelek?
  - Decilni razmik.
  - Povprečno porabljeni čas za en izdelek.
  - Čas za en izdelek, ki ga je porabilo 50 % zaposlenih z najmanjšo porabo časa za en izdelek.
- b) Za obe organizaciji izračunajte in pojasnite:
- Variacijski razmik.
  - Varianco in standardni odklon.
  - Koeficient variabilnosti.







## Naloga 10

Proučevali smo 260 zdravnikov splošne medicine glede na število opravljenih pregledov v določenem časovnem obdobju. Podatki so prikazani v naslednji tabeli:

Število opravljenih pregledov	Število zdravnikov
Od 31 do 60	35
Od 61 do 90	52
Od 91 do 120	74
Od 121 do 150	41
Od 151 do 180	32
Od 181 do 210	26
<b>Skupaj</b>	<b>260</b>

- Navedite statistično množico, enoto, spremenljivko in njene značilnosti.
- Grafično ocenite asimetrijo gornje porazdelitve.
- Izračunajte delež standardnega odklona v aritmetični sredini.

**Naloga 11**

V določeni trgovini so v 45 zaporednih dneh beležili naslednje število prodanih izdelkov:

Število prodanih izdelkov	Število dni
Od 1 do 10	6
Od 11 do 20	14
Od 21 do 30	15
Od 31 do 40	5
Od 41 do 50	5

- Izračunajte variacijski razmik.
- Grafično prikažite frekvenčno porazdelitev.
- Izračunajte in pojasnite aritmetično sredino.
- Izračunajte in pojasnite standardni odklon.
- Izračunajte in pojasnite koeficient variabilnosti v odstotku.



## Naloga 12

Vrednost naložb startup podjetij je bila v opazovanem obdobju sledeča (v d.e):

75      87      150      230      380      450      500

- Opreделите statistično množico, statistično enoto in statistično spremenljivko
- Kolikšno vrednost naložb je imelo 25 % startup podjetij z najnižjo vrednostjo naložb in kolikšno vrednost naložb je imelo 25 % startup podjetij z najvišjo vrednostjo naložb?
- Izračunajte mero variabilnosti, ki upošteva variabilnost za 50 % startup podjetij, ki se glede na vrednosti naložb razvrščajo na sredino ranžirne vrste.



## Naloga 13

V okviru ekološke iniciative so v podjetju analizirali zmanjšanje emisij CO<sub>2</sub> (v tonah na leto) zaradi uporabe javnega prevoza za devet zaposlenih. Podatki so sledeči:

9      12      16      41      60      80      100      120      150

- a) Kakšno zmanjšanje emisij CO<sub>2</sub> je bilo doseženo s strani 50 % zaposlenih, ki so dosegli najnižje vrednosti?
- b) V okoljevarstveni organizaciji A je povprečno zmanjšanje emisij CO<sub>2</sub> za 456 članov 7.530 ton na leto, varianca pa 7.022.500 ton<sup>2</sup> na leto. V katerem okolju (podjetje ali okoljevarstvena organizacija) je zmanjšanje emisij CO<sub>2</sub> med posamezniki bolj raznoliko?



## Naloga 14

Devet vlagateljev je v enem mesecu vložilo naslednje zneske (v d.e.) v trajnostne finančne produkte, kot so zeleni skladi ali obveznice. Zneski v d.e. so naslednji:

75      98      160      180      195      220      246      260      290

- a) Opredelite statistično enoto in statistično spremenljivko.
- b) Izračunajte in pojasnite koeficient asimetrije na podlagi mediane ter vse parametre, ki ste jih pri tem izračunali.

**Naloga 15**

V sedmih investicijskih projektih podjetja Y je bila v opazovanem letu naslednja donosnost: 4 %, 12 %, 7 %, 18 %, 2 %, 16 %, 10 %.

- a) Izračunajte, koliko odstotkov investicijskih projektov je imelo donosnost manjšo od 9 %.
- b) Izračunajte, koliko odstotkov investicijskih projektov je imelo donosnost večjo od 14 %.
- c) Izračunajte donosnost v 50 % investicijskih projektih z največjo donosnostjo.
- d) Izračunajte mero variabilnosti, ki upošteva donosnost v 80 % investicijskih projektov, ki se glede na odstotek donosnosti razvrščajo na sredino ranžirne vrste.
- e) Izračunajte povprečen odstotek donosnosti za sedem investicijskih projektov v podjetju Y.
- f) Izračunajte standardni odklon za povprečen odstotek donosnosti za sedem investicijskih projektov v podjetju Y.
- g) V konkurenčnem podjetju Z je povprečno število donosnosti njihovih investicijskih projektov 12 %, standardni odklon pa znaša 6,3 %. V katerem podjetju se investicijski projekti glede na odstotek donosnosti med seboj bolj razlikujejo?





## 2.4 Relativna števila: indeksna števila, povprečna vrednost v časovni vrsti

*Relativna števila* so izračunana z medsebojno vsebinsko smiselno primerjavo dveh podatkov. Lahko primerjamo posamezen podatek z drugim podatkom v isti statistični vrsti, na primer primerjava značilnosti dveh geografskih območij, dveh časovnih obdobj, ali pa z drugim podatkom v drugi statistični vrsti, kot na primer prodaja in stroški. Vrednost razlike med dvema vrednostima spremenljivke je absolutna razlika. Če pa podatka primerjamo z izračunom njunega razmerja, opisujemo odnos med pojavoma na relativni način – relativna števila. Glede na vsebinske povezave med primerjanimi podatki, poznamo naslednje vrste relativnih števil:

- strukturna ali razčlenitvena števila, to je primerjava vrednosti, ki se nanaša na del, z vrednostjo, ki se nanaša na celoto opazovanega pojava,
- statistični koeficienti, ko primerjamo podatke dveh statističnih vrst (raznovrstni podatki),
- indeksi, primerjava členov znotraj ene statistične vrste (istovrstni podatki).

*Indeksi* so relativna števila, ki omogočajo medsebojno primerjanje dveh ali več podobnih podatkov o določenem pojavu. Za izračun indeksov se lahko uporabijo različne vrste podatkov, vključno z absolutnimi podatki, koeficienti, strukturnimi odstotki in drugimi kazalci. Vsi primerjani podatki morajo biti izraženi v istih merilnih enotah. V primeru, da primerjamo med seboj le dva podatka, govorimo o enostavnih indeksih. Z indeksi dobimo zelo nazorno sliko o velikosti relativnih sprememb pojava v času oz. o velikosti relativnih razlik za pojav v prostoru (Ralph idr., 2015).

Če je vrednost indeksa enaka 100, sta podatka, ki ju primerjamo, enaka. Če je vrednost indeksa večja od 100, je primerjani podatek večji od osnove. Če pa je primerjani podatek manjši od osnove, je vrednost indeksa manjša od 100. Indekse pogosto komentiramo na podlagi relativne razlike oz. stopnje rasti, izražene v %, ki jo dobimo tako, da od izračunanega indeksa odštejemo 100. Če primerjamo dva indeksa, rezultat izrazimo v odstotnih točkah. Indeks 110,8 na primer pomeni, da je preučevani podatek za 10,8 % večji od osnove, indeks 94,5 pa pomeni, da je preučevana vrednost za 5,5 % manjša od osnove. Če indeks 110,8 primerjamo z indeksom 94,5, je razlika med njima 16,3 odstotne točke (Mišić, 2022).

Na primer, indeksi nudijo jasen vpogled v obseg relativnih sprememb nekega pojava skozi čas. Indekse delimo na (Tominc, 2016):

- krajevne indekse,
- stvarne indekse,
- časovne indekse: indeksi s stalno osnovo in indeksi s premično osnovo.

Indekse lahko izračunamo na podlagi primerjave dveh podobnih podatkov iz različnih geografskih regij, pri čemer govorimo o *krajevnih indeksih*. Ko primerjamo dva istovrstna podatka, ki se nanašata na različna časovna obdobja, govorimo o *časovnih indeksih*. Lahko izračunamo indekse za specifične pojave, ki niso geografsko ali časovno opredeljeni, in v tem primeru govorimo o *stvarnih indeksih*.

Pri časovnih indeksih spremljamo pojav v času. Iz časovnih indeksov lahko izračunamo indekse s stalno ali indekse s premično osnovo. Indekse s stalno osnovo ali bazne indekse izračunamo tako, da posamezen podatek v časovni vrsti primerjamo vedno z istim podatkom, ki si ga izberemo za bazo. Indekse s premično osnovo pa izračunamo tako, da v isti časovni vrsti spreminjamo podatek oziroma osnovo primerjave (npr. mesec, leto). Med indeksi s premično osnovo so najbolj znani in uporabljeni verižni indeksi; o teh govorimo, kadar za osnovo vedno vzamemo predhodni podatek v časovni vrsti (Mišič, 2022).

*Povprečno vrednost v časovni statistični vrsti* opredelimo s povprečno stopnjo rasti, ki se izraža v odstotkih (na primer, povprečna stopnja rasti BDP v Sloveniji v preteklih desetih letih). Povprečno stopnjo rasti v časovni statistični vrsti izračunamo bodisi iz *povprečnega koeficienta dinamike* bodisi iz *povprečnega verižnega indeksa*. Pri tem izhajamo iz postopka izračuna geometrijske sredine (in ne aritmetične sredine). Povprečno stopnjo rasti opazovane spremenljivke v preteklem obdobju lahko uporabimo tudi za ocenjevanje vrednosti spremenljivke oziroma napovedovanje vrednosti v prihodnjih časovnih enotah (Tominc, 2016).



## Primer rešene naloge

V preglednici so podatki o prodaji knjig študentom v knjigarni X v petih letih:

Leto	2020	2021	2022	2023
Prodaja knjige	215	300	280	310

- a) Za koliko odstotkov se je število prodanih knjig študentom v letu 2021 razlikovalo od števila v predhodnem letu?

b) Ocenite število prodanih knjig tujim študentom v letu 2025 z upoštevanjem povprečne stopnje rasti.

a)

Izračunati je potrebno verižni indeks za leto 2021, v %:

$$V_t = 100 \times \frac{Y_t}{Y_{t-1}} \quad \text{za } t = 2, 3, \dots, T$$

$$V_{2021} = \frac{300}{215} \cdot 100 = 139,53 \%$$

Število prodanih knjig študentom v letu 2021 se je povečalo za 39,53 % glede na leto 2020.

b)

Uporabimo enačbo za *povprečni koeficient dinamike*:

$$K = \sqrt[r]{\frac{Y_T}{Y_1}}$$

$$K = \sqrt[3]{\frac{310}{215}} = 1,1297$$

$$S = (K - 1) \cdot 100 = (1,1297 - 1) \cdot 100 = 12,97 \%$$

Število prodanih knjig študentom v knjigarni X se je v štirih letih povečevalo povprečno za 12,97 % na leto.

$$Y_{2025} = 310 \cdot K^2 = 310 \cdot 1,1297^2 = 395,6$$

Za leto 2025 na osnovi povprečne stopnje rasti ocenjujemo, da bo število prodanih knjig študentom enako 395,6 oziroma 396 knjig.



## Primer rešene naloge

V raziskovalnem inštitutu so opravili analizo števila objavljenih znanstvenih člankov in gradiv v zadnjih šestih letih. Število objav je bilo izraženo z indeksi s stalno osnovo v letu 2019. Podatki o številu objav v obliki indeksov so podani v naslednji preglednici:

Leto	2019	2020	2021	2022	2023	2024
$I_{t/2019}$	100	102	115	70	90	117

- Izračunajte število objavljenih znanstvenih člankov in gradiv po letih, če je bilo v letu 2021 objavljenih 60 znanstvenih člankov in gradiv.
- Izračunajte in vsebinsko pojasnite povprečno letno stopnjo rasti števila objavljenih znanstvenih člankov in gradiv v teh šestih letih.

a)

Leto	$I_{t/2019}$	Koeficient dinamike $K_t$	$Y_t$
2019	100	/	$\frac{53,1}{1,02} = 52,06$
2020	102	1,02	$\frac{60}{1,13} = 53,1$
2021	115	1,13	60
2022	70	0,61	$60 \cdot 0,61 = 36,6$
2023	90	1,29	$36,6 \cdot 1,29 = 47,21$
2024	117	1,3	$47,21 \cdot 1,3 = 61,37$

Za koeficient dinamike uporabimo enačbo:

$$K_t = \frac{y_t}{y_{t-1}} \quad K_1 = / \quad t = 2, \dots, t$$

Na primer:

$$K_{t/2020} = \frac{102}{100} = 1,02$$

$$K_{t/2021} = \frac{115}{102} = 1,13$$

b)

Uporabimo enačbo za povprečni koeficient dinamike:

$$K = \sqrt[T-1]{\frac{I_{T/o}}{I_{1/o}}}$$

$$K = \sqrt[6-1]{\frac{117}{100}} = \sqrt[5]{1,17} = 1,032$$

Uporabimo enačbo za povprečno stopnjo rasti:

$$S = (K - 1) \cdot 100$$

$$S = (1,032 - 1) \cdot 100 = 3,2 \%$$

Odgovor: Povprečna letna stopnja rasti števila objavljenih znanstvenih člankov in gradiv v obdobju šestih let znaša 3,2%. To pomeni, da je raziskovalni inštitut v povprečju dosegal 3,2-odstotno letno rast v številu objavljenih znanstvenih člankov in gradiv.



## Naloga 16

V podjetju Y so analizirali trend prodaje svojega glavnega produkta v zadnjih petih letih. Prodaja je bila izražena z indeksi s stalno osnovo v letu 2018. Podatki o prodaji v obliki indeksov so podani v naslednji preglednici:

Leto	2018	2019	2020	2021	2022
$I_t/2018$	100	105	95	110	120

- Izračunajte vrednost prodaje produkta po letih, če je bila v letu 2018 prodaja 200 enot.
- Izračunajte in vsebinsko pojasnite povprečno letno stopnjo rasti prodaje produkta v teh petih letih.

**Naloga 17**

Podatki o letnih stopnjah rasti proizvodnje električne energije iz vetrne energije na opazovanem področju v zadnjih devetih letih so naslednji (v GWh):

Leto	1	2	3	4	5	6	7
S <sub>v</sub> %	+3,5	+15,0	+1,0	-3,0	+4,5	-2,5	+6,0

- a) Pojasnite največjo pozitivno in največjo negativno stopnjo rasti opazovanega pojava.
- b) Zapišite in pojasnite relativne spremembe v proizvodnji električne energije iz vetrne energije v opazovanih letih:
  - z vrsto verižnih indeksov,
  - z vrsto letnih koeficientov dinamike,
  - z vrsto indeksov s stalno osnovo v letu 4.
- c) Če je bila proizvodnja električne energije iz vetrne energije v letu 4 enaka 1200 GWh, izračunajte proizvodnjo električne energije iz vetrne energije v preostalih letih.



## Naloga 18

Prodaja v letu 2018 je bila 1550 količinskih enot, v letu 2024 pa 890 količinskih enot. Predvidite prodajo v letu 2032 na osnovi povprečne stopnje rasti.



**Naloga 19**

V preglednici so podatki o številu novo registriranih električnih avtomobilov v nekem mestu v zadnjih sedmih letih:

Leto	1	2	3	4	5	6	7
Št. novo registriranih električnih avtomobilov	112	150	200	250	300	350	400

- Za koliko odstotkov se je število novo registriranih električnih avtomobilov v letu 5 razlikovalo od števila v predhodnem letu? Kako imenujemo izračunano vrednost?
- Koliko odstotkov števila novo registriranih električnih avtomobilov iz leta 1 predstavlja število novo registriranih avtomobilov v letu 7? Kako imenujemo izračunano vrednost?
- Statistično vrsto prikažite grafično.

**Naloga 20**

V preglednici so podatki o stopnji rasti naložb nekega naložbenega sklada v zadnjih štirih letih:

Leto	1	2	3	4
S <sub>t</sub> %	/	+8,4	+5,7	-2,3

- Analizirajte, kako so se relativno spreminjale vrednosti naložb glede na leto 2.
- Izračunajte, kolikšne so bile vrednosti naložb v obravnavanih letih, če je bila vrednost naložbe v letu 3 enaka 210.000 €.
- Izračunajte in pojasnite povprečno stopnjo rasti vrednosti naložb v obravnavani 4-letni časovni vrsti.
- Napovejte povprečno vrednost naložb obravnavanega naložbenega sklada v letu 7, če upoštevate povprečno stopnjo rasti v opazovanem obdobju.

**Naloga 21**

Dana je indeksna vrsta in podatek za leto 2023 v osnovni časovni vrsti:

Leto	2021	2022	2023	2024	2025
$V_t$	/	130	90	110	125
$Y_t$			1630		

- Izračunajte manjkajoče člene osnovne časovne vrste ter povprečni koeficient dinamike in povprečno stopnjo rasti.
- Napovejte vrednost spremenljivke v letu 2028.



## Naloga 22

V preglednici so podatki o spreminjanju obsega prodaje finančnih produktov v banki Z, v obliki indeksnega števila od leta 2018 do leta 2023:

Leto	2020	2021	2022	2023	2024	2025
$V_t$	/	70	100	85	95	110

- Kako imenujemo statistično vrsto?
- Vsebinsko pojasnite indeksno število za leto 2023.
- Izračunajte število prodanih izdelkov po letih, če je bilo leta 2022 število prodanih izdelkov enako 415. Časovno vrsto prikažite grafično.

**Naloga 23**

V preglednici so podatki o številu raziskovalnih projektov na področju umetne inteligence, ki jih je raziskovalni inštitut začel vsako leto skozi pet zaporednih let.

<b>Leto</b>	1	2	3	4	5
<b>Projekti</b>	5	8	10	13	16

- Kako imenujemo statistično vrsto v preglednici? Statistično vrsto grafično prikažite.
- Ocenite število raziskovalnih projektov v 8. zaporednem letu z upoštevanjem povprečne stopnje rasti.

**Naloga 24**

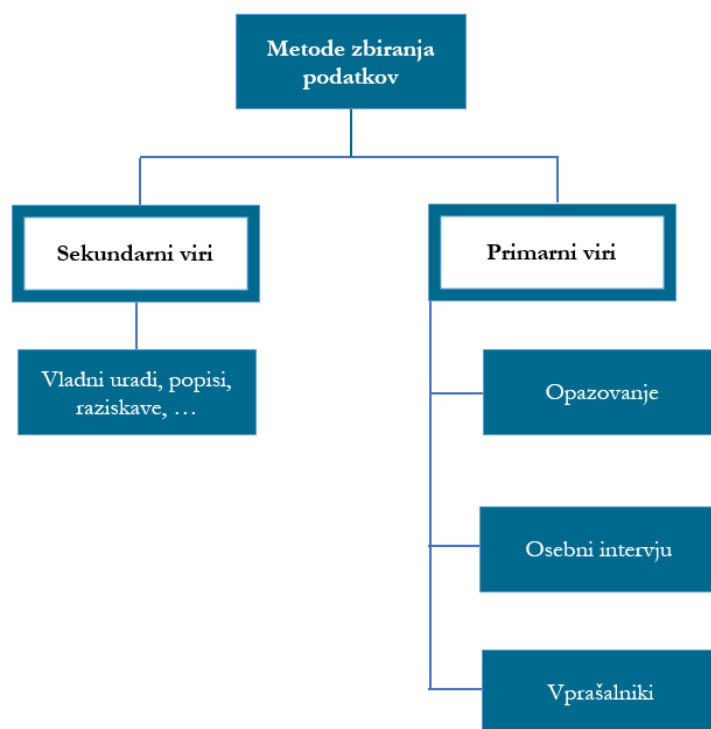
V opazovani organizaciji so podatki o številu zaposlenih in vrednosti proizvodnje po mesecih zapisani v naslednji časovni vrsti:

Mesec	I	II	III	IV	V
Vrednost proizvodnje v 10 <sup>6</sup> EUR	652	730	840	752	/
Število zaposlenih na začetku meseca	214	240	226	208	200

Kolikšna je bila povprečna mesečna vrednost proizvodnje v opazovani organizaciji na 10 zaposlenih?

## 2.5 Metode zbiranja podatkov

Zbiranje podatkov je temeljni korak v procesu raziskovanja, ki omogoča znanstvenikom, tržnim analitikom, družboslovcem in mnogim drugim strokovnjakom, da zgradijo trdne temelje za svoje raziskave in analize. Le z zanesljivimi podatki lahko oblikovalci politik oblikujejo učinkovite intervencije, podjetja izboljšajo svoje izdelke in storitve ter oblikujejo zanesljive informacije, potrebne za poslovno odločanje. Zbiranje podatkov je tako več kot le tehnična dejavnost; je ključno orodje za pridobivanje informacij, ki jih v poslovnem svetu neobhodno potrebujemo. Obstajajo različne metode zbiranja podatkov, ki se izberejo glede na raziskovalni cilj, postavljene domneve, naravo podatkov in razpoložljive vire. Slika 5 predstavlja metode zbiranja podatkov, ki so opisani v podpoglavju tega poglavja.



Slika 5: Metode zbiranja podatkov

### 2.5.1 Sekundarni in primarni viri

Sekundarni viri vključujejo uporabo podatkov, ki so bili že zbrani za druge namene in so na voljo preko različnih virov. To so podatki, ki jih raziskovalec pridobi iz obstoječih virov (Guest idr., 2012). Na primer, raziskovalec lahko analizira podatke iz preteklih raziskav, vladnih poročil, statističnih uradov (najpomembnejši je v Sloveniji Urad RS za statistiko), podjetniških zapisov ipd.

Primarni viri zbiranja podatkov vključujejo neposredno pridobivanje podatkov od virov, ki so neposredno povezani z raziskovalno temo. To so originalni podatki, zbrani specifično za trenutno raziskavo, kar pomeni, da raziskovalec sam oblikuje metode in instrumente za zbiranje podatkov. Pod primarne viri zbiranja podatkov uvrščamo opazovanje, intervju, anketni vprašalnik.

### 2.5.2 Opazovanje

Opazovanje je namenski in sistematičen način gledanja, poslušanja in zapisovanja opažanj o lastnostih opazovanega fenomena ali o medsebojnem vplivanju in delovanju različnih dejavnikov: opazovanje delovnega procesa, opazovanje reakcij ljudi ipd. Na splošno pa velja, da takrat, kadar nas bolj zanima obnašanje kot pa zaznave ljudi ali pa kadar so opazovani subjekti tako močno vključeni v opazovano dogajanje, da ne morejo hkrati tudi podajati informacij o dogajanju, je opazovanje najprimernejši način pridobivanja podatkov (Tominc, 2008).

Opazovanje lahko poteka s sodelovanjem, kar pomeni, da je opazovalec aktivno vključen v dejavnosti skupine, ki jo proučuje. Ta pristop omogoča raziskovalcu, da iz prve roke doživi in razume socialne interakcije, kulturo in vedenje znotraj skupine. Udeleženo opazovanje je še posebej koristno v antropoloških, socioloških in psiholoških raziskavah, kjer globlje razumevanje konteksta in subjektivnih izkušenj udeležencev igra ključno vlogo. Vendar pa ta metoda prinaša tudi etične izzive, saj lahko prisotnost raziskovalca vpliva na naravno vedenje skupine, hkrati pa raziskovalec mora paziti, da ohrani profesionalno distanco in objektivnost (Russell Bernard, 2011; Iacono idr., 2009).

Pri opazovanju brez sodelovanja opazovalec ne sodeluje neposredno v dejavnosti skupine. Ta pristop zmanjšuje možnost, da bi prisotnost opazovalca vplivala na obnašanje opazovanih, kar omogoča bolj objektivno zbiranje podatkov. Neudeleženo opazovanje je pogosto uporabljeno v psiholoških eksperimentih, tržnih raziskavah in drugih situacijah, kjer je pomembno zaznavanje nepristranskih vedenjskih vzorcev. Kljub temu da ta metoda omogoča večjo objektivnost, pa lahko pomanjkanje neposredne interakcije omeji globino razumevanja konteksta in subtilnih socialnih dinamik (Given, 2008; Altschuld, 2009).

### 2.5.3 Intervju

Intervju je kvalitativna raziskovalna metoda, ki se uporablja za zbiranje primarnih podatkov. Vključuje spraševanje enega ali več ljudi o temi, ki jo želi raziskovalec raziskati.



Ta metoda omogoča raziskovalcem, da pridobijo podrobne informacije, ki morda niso na voljo z drugimi raziskovalnimi metodami.

Intervju omogoča raziskovalcu, da zbere podatke neposredno od posameznikov, kar zagotavlja kontekstualno bogastvo in globino, ki je pogosto nedosegljiva z drugimi metodami. Udeleženci lahko podajo podrobne opise svojih izkušenj, občutkov in mnenj, ki razkrivajo subjektivne perspektive, ki so ključne za razumevanje socialnih in psiholoških pojavov. Pri izvedbi intervjujev je pomembno, da raziskovalec vzpostavi zaupanje in udobje z udeleženci, saj to povečuje verodostojnost in globino pridobljenih informacij. Etika intervjuvanja zahteva tudi jasno komunikacijo o namenu raziskave, uporabi podatkov in pravici udeležencev do prekinitve sodelovanja kadarkoli (Flick, 2018; Denzin in Lincoln, 2018).

Intervjuji so lahko *strukturirani* ali *nestrukturirani*. Nestrukturirani so tisti, ki so prilagodljivi in kjer ima izpraševalec več možnosti in svobode pri zastavljanju vprašanj, lahko sproti odloča o vrstnem redu vprašanj in sproti izbira besede pri postavljanju posameznih vprašanj. Med tovrstne intervjuje sodijo različni poglobljeni osebni intervjuji ali intervjuji, ki potekajo v skupini (fokusna skupina). Pri intervjuju vprašanja torej bere anketar, ki vprašanja lahko tudi dodatno razloži, če je potrebno, odgovore vprašanega pa zapisuje anketar (Tominc, 2008). Strukturirani intervjuji so zelo formalizirani, pri čemer raziskovalec uporablja vnaprej določen niz vprašanj, ki jih postavlja v nespremenjeni obliki vsem udeležencem. Ta pristop omogoča lažje primerjanje odgovorov med različnimi udeleženci in je primeren za kvantitativne analize, kjer je potrebna visoka stopnja standardizacije (Flick, 2018).

#### 2.5.4 Vprašalnik

Vprašalnik je merski instrument, orodje, ki se uporablja za zbiranje podatkov in pridobivanje informacij od ciljne populacije. Vprašalnik je lahko sestavljen iz zaprtega ali odprtega tipa vprašanj. Pri vprašanju zaprtega tipa respondent izbere odgovor iz omejenega nabora podanih možnih odgovorov. Ta pristop omogoča raziskovalcem, da zlahka kvantificirajo podatke in jih statistično analizirajo (Roopa in Satya 2012; Jenn, 2006). Pri vprašanjih, ki vključujejo merjenje stališč se uporablja Likertova lestvica (običajno 5-stopenjska ali 7-stopenjska), ki meri intenzivnost stališč anketiranih oseb do različnih vidikov opazovane situacije, teme, izdelka, itd. (Tominc, 2008).

Vprašanja odprtega tipa omogočajo respondentom, da v svojih besedah izrazijo mnenja, občutke ali ideje. Ta vrsta vprašanj je koristna za pridobivanje bolj globokih vpogledov in lahko razkrije nove ideje, ki jih raziskovalec morda ni predvidel. Odgovori na ta vprašanja zahtevajo bolj kompleksno analizo in se običajno obdelujejo s kvalitativnimi metodami, kot sta tematska analiza ali kodiranje. V vprašalniku pa lahko uporabimo tudi kombiniran tip vprašanj, ki združuje vprašanja odprtega in zaprtega tipa (Patten, 2014; Jenn, 2006).

Dobro zasnovan anketni vprašalnik mora biti veljaven, zanesljiv, jasen, zanimiv in jedrnat. Veljaven anketni vprašalnik pomeni, da meri točno to, kar želi izmeriti raziskovalec. Da bi to dosegli, je treba vprašalnik med pilotnim testiranjem pregledati s strani 'strokovnjaka za vsebino' (na primer, če je ciljni respondent diabetični bolnik, potem mora diabetični bolnik komentirati, ali razume vprašalnik). Vse nejasnosti in vprašanja je treba pojasniti, dokler vprašanje ni popolnoma razumljivo. Zanesljiv anketni vprašalnik bi moral ob ponovnem postavljanju istega vprašanja respondentu v kratkem časovnem obdobju pridobiti enak odgovor. To dosežemo z izvedbo testa ponovitve, tj. isti vprašalnik se ponovno predloži respondentu in se preveri doslednost odgovorov. Morebitna neskladja v odgovorih bi lahko bila posledica pomanjkanja jasnosti vprašanj, kar je treba pregledati in preoblikovati. Zanimiv anketni vprašalnik ima večjo verjetnost, da ga bo respondent izpolnil, zato prinaša boljšo stopnjo odzivov. To zahteva, da raziskovalec razmisli o postavljanju vprašanj, ki so relevantna za respondenta in so postavljena v logičnem zaporedju. Jedrnat anketni vprašalnik postavlja vprašanja, ki ciljajo zgolj na odgovore, povezane z raziskovalnimi cilji. Vsa vprašanja, ki presegajo obseg raziskave, je treba izključiti. To se običajno zgodi, ko raziskovalec ni temeljito premislil o raziskovalnih ciljih. Tak pristop tvega, da bo postavil preveč vprašanj in da bo vprašalnik obsegal veliko strani (Jenn, 2006).

Prednosti uporabe anketnega vprašalnika (Tominc, 2008; Patten, 2014; Jenn, 2006):

1. *Anketni vprašalniki so relativno poceni*: odvisno od kompleksnosti študije, je lahko uporaba anketnega vprašalnika stroškovno učinkovita v primerjavi z drugimi metodami.
2. *Rezultati se lahko hitro pridobijo in analizirajo*: odvisno od obsega raziskave, lahko raziskovalec odgovore na anketni vprašalnik hitro prejme, kar pomeni, da lahko tudi hitro začnete z analizo odgovorov.
3. *Anketni vprašalniki so enostavni za analizo*: če je oblikovanje vprašalnika pravilno izvedeno, je analiza rezultatov hitra in enostavna. To je še posebej koristno pri obsežnih tržnih raziskavah. Ker vsi respondentni odgovarjajo na ista vprašanja, je enostavno identificirati odgovore.

4. *Anonimnost anketirane osebe*: z vprašalnikom, ki ga izpolni anketirana oseba sama, lahko zagotovimo večjo anonimnost.

Slabosti uporabe anketnega vprašalnika (Tominc, 2008):

1. Uporaba vprašalnika je omejena na populacijo, ki zna brati in pisati. Problem te vrste se pri poslovnih raziskavah redkeje pojavlja.
2. Stopnja odgovorov je običajno nizka. Pri pošiljanju vprašalnikov se moramo zavedati, da veliko ljudi (ali podjetij, če statistično enoto v naši raziskavi predstavlja podjetje) vprašalnikov ne izpolni in ne vrne. To ima lahko velik vpliv na velikost vzorca.
3. Respondenti morda ne bodo odgovorili na vsa vprašanja. Razlogov, zakaj respondent morda ne odgovori na vsako vprašanje, je več, od dolžine vprašalnika do tega, da morda ne razumejo, kaj se sprašuje, ali preprosto ne želijo odgovoriti.

Zanesljivost vprašalnika oziroma zanesljivost posamezne skale, s katero analiziramo stališče do neke teme, objekta, itd., merimo z ustreznim kazalcem zanesljivost, s pomočjo Cronbachovega koeficienta  $\alpha$ . Zanesljivost vprašalnika (oziroma posamezne skale) je (Tominc, 2008) je:

- zgedna, če je koeficient Cronbach alfa večji ali enak 0,80,
- zelo dobra, če je koeficient Cronbach alfa večji ali enak 0,70 in manjši od 0,80,
- zmerna, če je koeficient Cronbach alfa večji ali enak 0,60 in manjši od 0,70,
- komaj sprejemljiva, če je koeficient Cronbach alfa manjši od 0,60.

**Naloga 25:** Opišite vire podatkov in načine zbiranja podatkov.

---

---

---

---

---

---

**Naloga 26:** Pojasnite razliko med primarnimi in sekundarnimi viri podatkov.

---

---

---

---

---

---

**Naloga 27:** Kaj je značilno za anketni vprašalnik? Pojasnite prednosti in pomanjkljivosti anketnega vprašalnika.

---

---

---

---

---

---

---

**Naloga 28:** Naštejte načine razdeljevanja anketnega vprašalnika.

---

---

---

---

---

---







Trditev 5: Za upravljanje urnika za učenje uporabljam orodja za načrtovanje učenja (npr. koledarje, aplikacije, planerje).

Rezultati analize zanesljivosti so v spodnjih preglednicah. Analizirajte in vsebinsko pojasnite dobljene rezultate.

#### Reliability Statistics

Cronbach's Alpha	N of Items
0,785	6

#### Item Statistics

	Mean	Std. Deviation	N
Najraje se učim vsak dan ob isti uri.	4,14	,925	278
Svoje ure učenja razdelim na več časovnih intervalov ali dni.	4,08	,958	278
Čas učenja enakomerno razdelim med vse predmete, ki se jih moram naučiti.	3,27	,992	278
Svoj urnik učenja prilagodim glede na študijsko obremenitev.	3,28	,984	278
Za upravljanje urnika za učenje uporabljam orodja za načrtovanje učenja (npr. koledarje, aplikacije, planerje).	3,50	1,087	278

#### Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Najraje se učim vsak dan ob isti uri.	17,98	12,639	,554	,749
Svoje ure učenja razdelim na več časovnih intervalov ali dni.	18,04	12,428	,561	,746
Čas učenja enakomerno razdelim med vse predmete, ki se jih moram naučiti.	18,85	12,324	,550	,749
Svoj urnik učenja prilagodim glede na študijsko obremenitev.	18,85	12,473	,532	,753
Za upravljanje urnika za učenje uporabljam orodja za načrtovanje učenja (npr. koledarje, aplikacije, planerje).	18,62	12,402	,464	,804







## 2.6 Normalna porazdelitev

Normalna porazdelitev, znana tudi z imenom Gaussova porazdelitev, je ena izmed najpomembnejših in najpogosteje uporabljenih statističnih porazdelitev v naravoslovju, družboslovju, ekonomiji in inženirstvu. Razumevanje oblike in lastnosti normalne porazdelitve je ključno za interpretacijo mnogih naravnih in družbenih pojavov. Znanje o tem, kako podatki sledijo normalni porazdelitvi, omogoča boljše razumevanje in interpretacijo statističnih podatkov, izvajanje statističnih testov in napovedovanje verjetnosti (Thomopoulos, 2017).

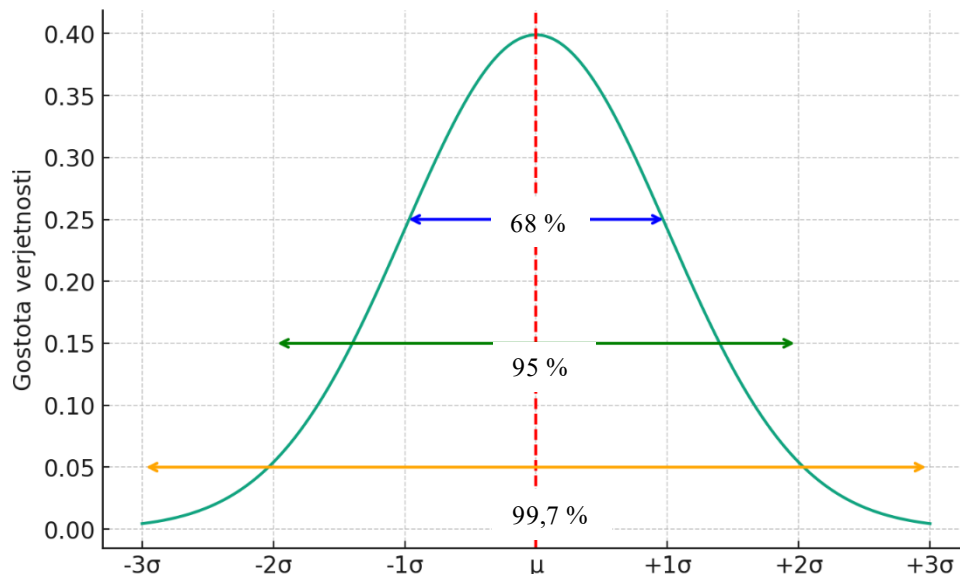
Normalna porazdelitev je definirana z dvema parametroma, aritmetično sredino (označimo jo z  $\bar{y}$ , velikokrat pa se za aritmetično sredino v statistični množici uporablja tudi oznaka  $\mu$ ) in standardnim odklonom ( $\sigma$ ). Normalna porazdelitev je simetrična glede na aritmetično sredino, aritmetična sredina pa je po vrednosti hkrati tudi enaka mediani in modusu (Holmes idr., 2018).

Lastnosti normalne porazdelitve (Thomopoulos, 2017; Holmes idr., 2018; Evans idr., 2010):

- *simetrija*: krivulja je simetrična okoli srednje vrednosti, ki jo označimo  $\bar{y}$ , kar pomeni, da sta repa na obeh koncih krivulje zrcalni sliki drug drugega;
- *srednja vrednost, mediana in modus*: pri normalni porazdelitvi je aritmetična sredina enaka mediani in modusu, ki sovpadajo;
- *asimptotska narava*: krivulja se približuje, a nikoli ne doseže horizontalne osi. Vrednosti daleč od sredine imajo zelo majhno verjetnost.

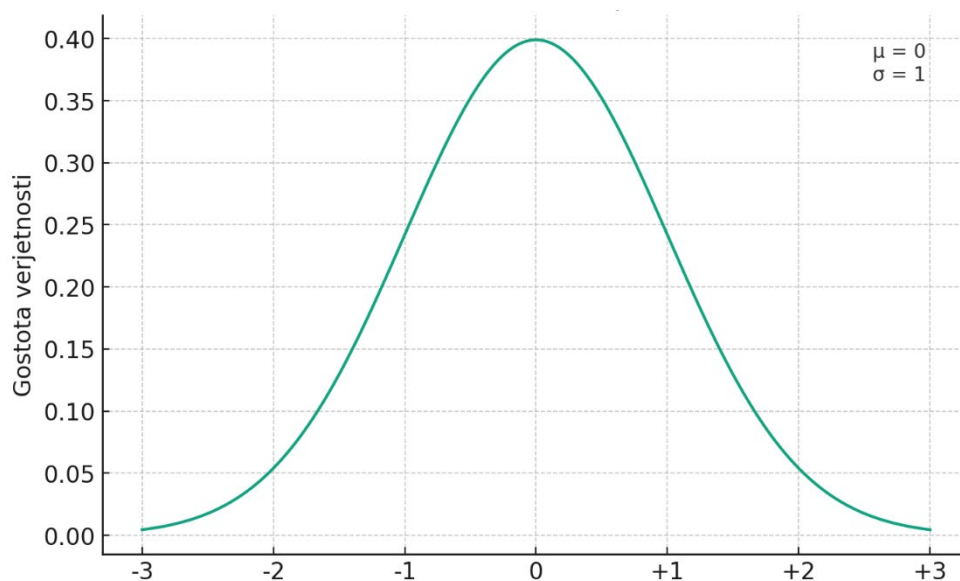
Slika 6 prikazuje normalno porazdelitev z značilno obliko zvona, ki je simetrična okoli aritmetične sredine, označene z  $\bar{y}$  (uporablja se tudi oznaka  $\mu$ ). Središče krivulje je na aritmetični sredini, kjer doseže najvišjo točko gostote verjetnosti. Standardni odklon  $\sigma$  določa širino porazdelitve krivulje. Manjši kot je standardni odklon, bolj je krivulja koncentrirana okoli aritmetične sredine  $\bar{y}$ . Če bi standardni odklon bil večji, bi bila krivulja širša, kar bi nakazovalo večjo variabilnost podatkov. Območja pod krivuljo prikazujejo delež vrednosti v okviru enega ( $\pm 1\sigma$ ), dveh ( $\pm 2\sigma$ ) in treh ( $\pm 3\sigma$ ) standardnih odklonov od aritmetične sredine. Območje znotraj enega standardnega odklona pokriva približno 68 % vseh vrednosti, znotraj dveh standardnih odklonov približno 95 % vseh vrednosti in znotraj treh standardnih odklonov približno 99,7 % vseh vrednosti. Repi krivulje so

asimptotični; približujejo se, vendar ne dosežejo x-osi, kar pomeni, da je verjetnost za zelo velika odstopanja vrednosti spremenljivke od srednje vrednosti izredno majhna.



Slika 6: Prikaz normalne porazdelitve

Standardizirana normalna porazdelitev (slika 7) je normalna porazdelitev standardiziranih vrednosti, imenovanih tudi z-vrednosti. Vsaki vrednosti  $y$  spremenljivke, porazdeljeni po poljubni normalni porazdelitvi, je mogoče izračunati njeno standardizirano z-vrednost (Holmes idr., 2018) z upoštevanjem transformacijske enačbe:  $z_i = \frac{y - \bar{y}}{\sigma}$ . Pri standardizirani normalni porazdelitvi je aritmetična sredina enaka 0 in standardni odklon je enak 1.



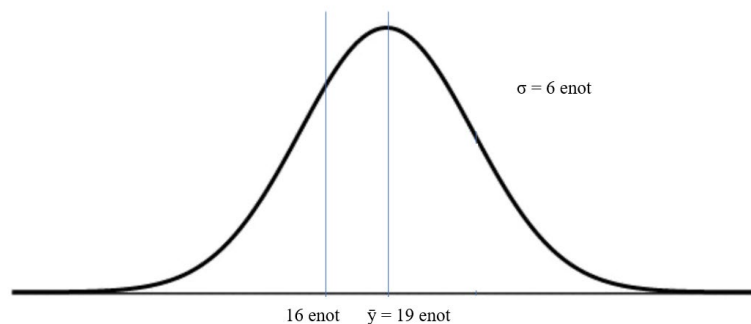
Slika 7: Standardizirana normalna porazdelitev



## Primer rešene naloge

Prodajna cena delnic podjetij določene gospodarske panoge na borzi se porazdeljuje po normalni porazdelitvi. Povprečna prodajna cena delnice je 19 d.e., s standardnim odklonom 6 d.e. Izračunajte verjetnost, da bo naključno izbrana delnica imela prodajno ceno večjo od 16 d.e.

Prikaz normalne porazdelitve:



Standardizirana normalna porazdelitev:



$$z_i = \frac{y - \bar{y}}{\sigma} = \frac{16 - 19}{6} = -0,5$$

$$P(y > 16 \text{ enot}) = 0,5 + H(-0,5) = 0,5 + 0,1915 = 0,6915 = 69,15 \%$$

Vrednost  $H(z_i)$  odčitamo iz tabele *ploščine*  $H(z)$  za standardizirano normalno porazdelitev.

Verjetnost, da bo naključno izbrana delnica imela prodajno ceno večjo od 16 d.e., je 69,15 %.

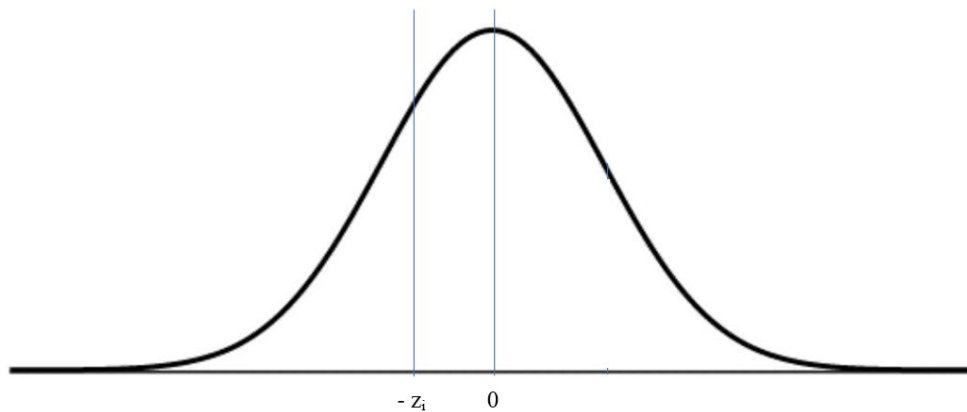


## Primer rešene naloge

Čas, ki ga zaposleni v finančnem oddelku podjetja porabijo za pripravo mesečnega poročila o stroških, se porazdeljuje po normalni porazdelitvi, z aritmetično sredino 40 minut in standardnim odklonom 15 minut.

Izračunajte, koliko minut porabi tistih 20 % zaposlenih, ki so pripravili mesečno poročilo o stroških v najkrajšem času.

Standardizirana normalna porazdelitev:



$$P(y < y_i) = 20 \%$$

$$H(z_i) = 0,3$$

Vrednost  $H(z_i)$  poiščemo v tabeli *ploščine  $H(z)$  za standardizirano normalno porazdelitev*, kjer poiščemo ploščino, ki je po vrednosti najbližje (tj. 0,2995), ter odčitamo pripadajočo standardizirano vrednost v prvem stolpcu ter prvi vrstici tabele, kar pomeni, da je  $z_i = 0,84$ , ki ji dodamo negativni predznak, saj leži iskana vrednost na levo od aritmetične sredine.

$$y_i = \bar{y} + z_i \cdot \sigma$$

$$y_i = 40 - 0,84 \cdot 15 = 27,4 \text{ minut}$$

Zaposleni, ki spadajo med 20 % tistih, ki so pripravili mesečno poročilo o stroških v najkrajšem času, so za pripravo poročila porabili 27,4 minut.



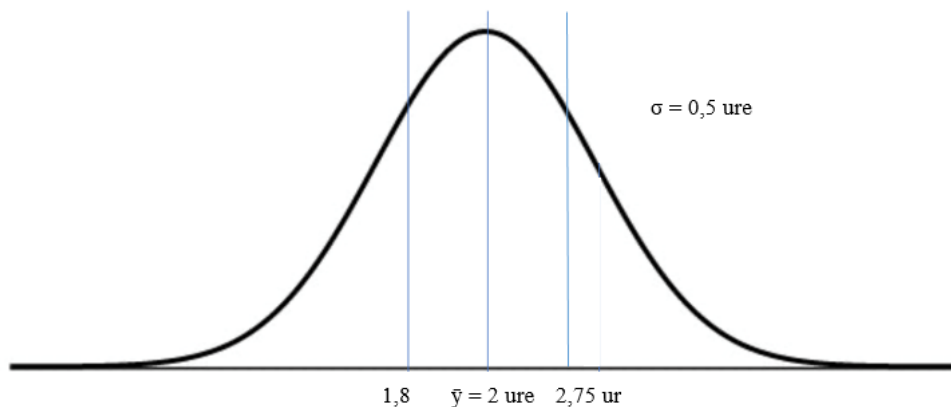
## Primer rešene naloge

V laboratorijih nekega inštituta je čas delovanja določenega laboratorijskega instrumenta za opravljanje testov porazdeljen po normalni porazdelitvi, z aritmetično sredino dve uri in standardnim odklonom 0,5 ure.

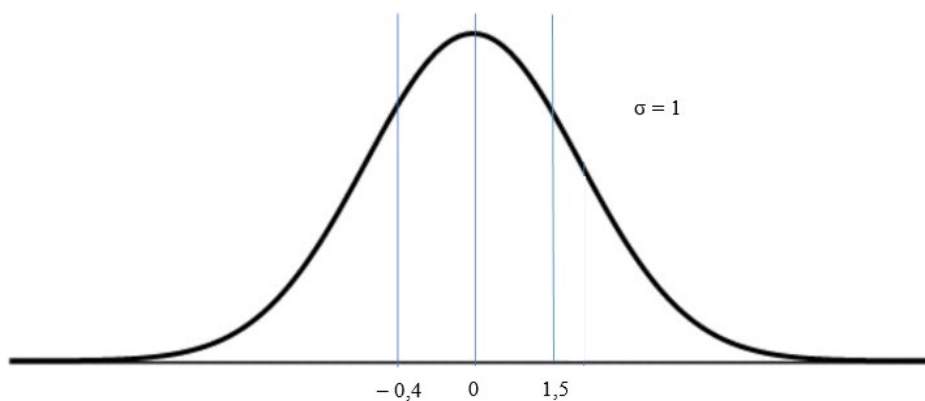
- Izračunajte verjetnost, da je v naključno izbranem laboratoriju instrument v delovanju za opravljanje testov med 1,8 in 2,75 urami na dan.
- Koliko časa je instrument v delovanju za opravljanje testov pri tistih 25 odstotkih laboratorijev, kjer je ta čas najkrajši?

a)

Normalna porazdelitev:



Standardizirana normalna porazdelitev:



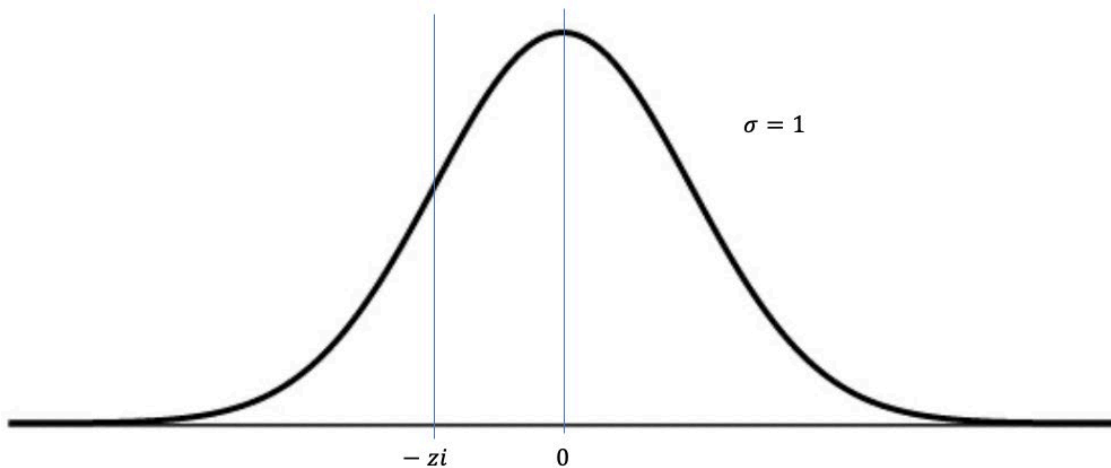
$P(1,8 \text{ ure} < y < 2,75 \text{ ure}) = H(-0,4) + H(1,5) = 0,1554 + 0,4332 = 0,5886 = 58,86 \%$   
 Vrednost  $H(z_i)$  odčitamo iz tabele *ploščine*  $H(z)$  za standardizirano normalno porazdelitev.

Verjetnost, da je v naključno izbranem laboratoriju instrument v delovanju za opravljanje testov med 1,8 in 2,75 urami na dan, je enaka 58,86 %.

b)

V tem primeru je podana verjetnost:

$$P(y < z_i) = 0,25$$



$$H(z_i) = 0,25$$

Iz tabele *ploščine*  $H(z)$  *za standardizirano normalno porazdelitev* odčitamo vrednost  $z_i$ :

$$H(-0,67) = 0,2486 \approx 0,25$$

$$y_i = -0,67 \cdot 0,5 + 2 = 1,7 \text{ ure}$$

V 25 % laboratorijev, kjer je čas uporabe instrumenta v delovanju za opravljanje testov najkrajši, je ta čas do 1,7 ure dnevno.



**Naloga 36**

V gospodinjstvih v neki skupnosti je čas uporabe družinskega avtomobila na dan porazdeljen po normalni porazdelitvi, z aritmetično sredino 3 ure in standardnim odklonom 1 uro.

- a) Izračunajte verjetnost, da je v naključno izbranem gospodinjstvu družinski avtomobil v uporabi med 2 in 5 urami na dan.
- b) V koliko odstotkih gospodinjstev se družinski avtomobil uporablja več kot 6 ur dnevno?
- c) Koliko časa na dan je družinski avtomobil v uporabi pri tistih 35 odstotkih gospodinjstev, kjer je ta čas najdaljši?



### Naloga 37

Letni prihodek samostojnih podjetnikov v določeni industriji se porazdeljuje po normalni porazdelitvi. Povprečni letni prihodek v tej skupini znaša 70.000 evrov, s standardnim odklonom 8.000 evrov. Izračunajte verjetnost, da bo naključno izbrani samostojni podjetnik imel letni prihodek večji od 60.000 evrov.

**Naloga 38**

Letna rast BDP države se porazdeljuje po normalni porazdelitvi. Povprečna letna rast BDP je 3 %, s standardnim odklonom 1,5 %. Izračunajte verjetnost, da bo letna rast BDP te države v določenem letu manjša od 2,5 %.

**Naloga 39**

Trajanje spanja pri odraslih posameznikih se porazdeljuje po normalni porazdelitvi, z aritmetično sredino 7 ur in standardnim odklonom 1,2 ure. Izračunajte, koliko ur spanja spijo tisti posamezniki, ki spadajo med 35 % posameznikov z najkrajšim trajanjem spanja.



## Naloga 40

Raziskovalna skupina za umetno inteligenco meri učinkovitost algoritmov za obdelavo naravnega jezika. Čas, potreben za analizo in odgovor na kompleksno vprašanje, se porazdeljuje po normalni porazdelitvi med vsemi testiranimi algoritmi. Povprečni čas odgovora v zadnjem sklopu testiranj je bil 1,02 sekunde, s standardnim odklonom 0,04 sekunde.

- a) Kolikšna je verjetnost, da bo naključno izbrani algoritem potreboval za odgovor največ 0,98 sekunde?
- b) Kolikšna je verjetnost, da bo naključno izbrani algoritem potreboval za odgovor med 0,97 in 1,05 sekundami?
- c) Določite mejne vrednosti za interval, v katerem se nahaja 50 % algoritmov, če so glede na čas odgovora razporejeni simetrično okoli povprečne vrednosti. Kako se imenuje izračunani interval?

**Naloga 41**

Prodajna cena stanovanj v določeni mestni četrti se porazdeljuje po normalni porazdelitvi z aritmetično sredino 150.000 evrov in standardnim odklonom 40.000 evrov. Na osnovi standardizirane normalne porazdelitve izračunajte verjetnost, da je prodajna cena naključno izbranega stanovanja med 60.000 in 120.000 evrov.



## Naloga 42

Podjetje X razvija nov sistem umetne inteligence za napovedovanje gibanja cen na delniških trgih. Uspešnost sistema pri napovedovanju naraščanja cen delnic (pravilni odstotek naraščanja cene delnic v določenem časovnem obdobju) je normalno porazdeljena. Zanima jih, kakšna je verjetnost, da bo njihov sistem napovedal naraščanje cen z uspešnostjo, ki je večja od 6 %. Povprečna uspešnost sistema je 5 % (povprečen pravilni odstotek naraščanja cene delnic), s standardnim odklonom 2 %.

**Naloga 43**

Čas, potreben za izvedbo ene delavnice v podjetju Y, se porazdeljuje po normalni porazdelitvi, z aritmetično sredino 90 minut in standardnim odklonom 35 minut.

- a) Izračunajte, koliko minut traja izvedba delavnice za tiste organizatorje, ki sodijo med 40 % z najdaljšim trajanjem delavnic?
- b) Izračunajte, koliko minut traja izvedba delavnice za tiste organizatorje, ki sodijo med 30 % z najkrajšim trajanjem delavnic?
- c) Izračunajte, koliko odstotkov organizatorjev je za izvedbo delavnice porabilo med 50 in 105 minut?
- d) Izračunajte, koliko odstotkov organizatorjev je za izvedbo delavnice porabilo med 100 in 130 minut?

## 2.7 Osnove vzorčenja in osnove preizkušanja domnev

Na splošno obstajata dve skupini tehnik vzorčenja: *tehnike verjetnostnega vzorčenja* in *tehnike neverjetnostnega vzorčenja*. Tehnike verjetnostnega vzorčenja se opirajo na teorijo verjetnosti in zagotavljajo, da ima vsak element populacije znano in neodvisno možnost, da je izbran v vzorec. Vse tehnike verjetnostnega vzorčenja ustvarjajo vzorce, ki so reprezentativni za populacijo in s tem omogočajo verodostojno statistično sklepanje. Neverjetnostne tehnike vzorčenja se nanašajo na subjektivno presojo raziskovalca pri izbiri enot iz populacije za vključitev v vzorec. Ker tehnike neverjetnostnega vzorčenja ne zagotavljajo reprezentativnega vzorca populacije, je treba pri interpretaciji in sklepanju o populaciji s podatki, zbranimi s temi metodami, biti previden. Take metode so pogosto primerne za eksploratorne raziskave ali ko so potrebne poglobljene študije o določenih skupinah, kjer ni mogoče rezultatov iz vzorca posplošiti na celotno populacijo (Valliant idr., 2018; Chaudhuri, 2014; Wolf idr., 2016).

Med tehnike verjetnostnega vzorčenja sodi (Heeringa idr., 2010; Johnnie, 2012; Valliant idr., 2018; Chaudhuri, 2014; Wolf idr., 2016):

1. Naključno vzorčenje, ki zagotavlja, da ima vsak element populacije enako verjetnost, da je izbran v vzorec.
2. Stratificirano vzorčenje, za katerega je značilno, da populacijo razdelimo v homogene podskupine ali stratum. Stratificirano vzorčenje zajema proporcionalno stratifikacijo in disproporcionalno stratifikacijo. Pri proporcionalni stratifikaciji je velikost vzorca iz stratuma sorazmerna ustrezni velikosti populacije, kar pomeni, da v vseh stratumih izberemo enak vzorčni delež. Pri disproporcionalni stratifikaciji pa imajo stratumi različne vzorčne deleže, zato v takem primeru struktura stratumov v vzorcu ne odseva strukture stratumov v populaciji. Disproporcionalno stratifikacijo uporabljamo takrat, kadar želimo doseči dovolj natančne ocene za vsak stratum posebej.
3. Sistematično vzorčenje, za katerega je značilno, da izberemo v vzorec vsak k-ti element oziroma statistično enoto iz populacije, pri čemer smo predhodno izbrali določen slučajni začetek.

Med tehnike neverjetnostnega vzorčenja sodi (Heeringa idr., 2010; Johnnie, 2012; Valliant idr., 2018; Chaudhuri, 2014; Wolf idr., 2016):

1. Kvotno vzorčenje, za katerega je značilno, da raziskovalci zberejo določeno število oseb ali enot, ki ustrezajo določenim značilnostim, iz populacije.



2. Priložnostno vzorčenje, kjer raziskovalec izbere elemente, ki so mu lahko dostopni. Priložnostni vzorec je tako sestavljen iz oseb, dogodkov ali predmetov, ki so raziskovalcu najbližje na voljo in najlažje dosegljivi za raziskavo.
3. Vzorčenje po presoji, ki je metoda neverjetnostnega vzorčenja, pri kateri raziskovalec izbere enote za vzorec na podlagi lastne presoje in znanja o populaciji. Pri tem izbira enot temelji na značilnostih, za katere raziskovalec verjame, da so pomembne za specifično raziskovalno vprašanje. Raziskovalec se zanaša na svoje strokovno znanje in izkušnje, da identificira in izbere enote, ki bodo najbolj prispevale k razumevanju problema.
4. Vzorčenje po principu kotaleče snežne kepe, se začne z identifikacijo nekaj potencialnih respondentov, ki ustrezajo kriterijem raziskave. Ti začetni udeleženci nato pripomorejo k identificiranju drugih udeležencev, saj 'kotalijo snežno kepo' tako, da priporočajo naslednje udeležence, ki prav tako ustrezajo raziskovalnim kriterijem. Postopek se ponavlja, dokler se ne zbere dovolj velik vzorec ali dokler novi udeleženci ne morejo več prispevati novega ali relevantnega informacij.

Na osnovi vzorčnih podatkov verjetnostnega vzorca lahko izvedemo postopek statističnega sklepanja, ki omogoča, da rezultate iz vzorca posplošimo na celotno statistično populacijo. Ta pomemben del statistike je znan kot inferenčna statistika. Vzorčni podatki služijo kot osnova za izračun ocene statističnega parametra v statistični populaciji. Vendar se zavedamo, da ta vzorčna ali točkovna ocena najverjetneje ni natančna vrednost parametra v populaciji, temveč le njena ocena. Zato pri izračunu vzorčne ocene statističnega parametra izvedemo tudi izračun intervalne ocene, imenovane *interval zaupanja*. Ta interval zaupanja je določen z določeno verjetnostjo, ki ji rečemo stopnja zaupanja. Na primer, če želimo izračunati interval zaupanja s stopnjo zaupanja 95 %, to pomeni, da je verjetnost, da se resnična vrednost parametra nahaja v tem intervalu, 95 %. Ta koncept intervalov zaupanja je ključen za ocenjevanje zanesljivosti in natančnosti naših ocen parametra na podlagi vzorčnih podatkov (Heumann idr., 2016).

Poleg intervalov zaupanja za oceno vrednosti statističnega parametra v statistični množici se v tem vsebinskem sklopu ukvarjamo tudi s *preverjanjem hipoteze* o vrednosti določenega statističnega parametra. Testiranje hipoteze vključuje zbiranje podatkov iz slučajnega vzorca in vrednotenje zbranih podatkov. Na podlagi analize podatkov se odločimo, ali obstaja dovolj informacij in ustrezne okoliščine za zavrnitev ničelne hipoteze ( $H_0$ ) ali ne.

*Ničelna hipoteza*,  $H_0$ , je trditev, da je vrednost aritmetične sredine (ali strukturnega odstotka ali totala) statistične množice enaka neki vnaprej določeni vrednosti.

*Alternativna ali raziskovalna hipoteza, H1*, pa je trditev o vrednosti statističnega parametra v statistični množici, ki je nasprotna trditvi H0, in menimo, da drži takrat, ko H0 lahko zavrnemo. Običajno pri preizkušanju hipoteze H0 uporabljamo kot najvišjo še dopustno petodstotno stopnjo tveganja (maksimalna dopustna stopnja tveganja).



### Primer rešene naloge

V naključnem vzorcu  $n = 300$  pacientov je bila povprečna količina dnevno zaužitih vitaminov 15,4 mg, nepristranska ocena variance pa 30,25 mg<sup>2</sup>.

- Izračunajte 95-% interval zaupanja za povprečno količino zaužitih vitaminov v osnovni statistični množici.
- Z 10-% tveganjem preizkusite domnevo, da je povprečna količina dnevno zaužitih vitaminov enaka 13,8 mg.

a)

Dvostranski interval za povprečno količino zaužitih vitaminov v osnovni statistični množici s 95-% verjetnostjo:

Imamo velik vzorec, ker je  $n = 300$ , ter dvostransko intervalno ocenjevanje aritmetične sredine:

$$\gamma = 95\%, \alpha = 5\%, z = \pm 1,96$$

*Kritične vrednosti za spremenljivko  $z$ :*

Stopnja tveganja $\alpha$	Dvostransko ocenjevanje
10 %	$\pm 1,645$
5 %	$\pm 1,96$
1 %	$\pm 2,58$

Izračunamo standardno napako ocene aritmetične sredine:

$$SE_{\bar{y}} = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$$

$$\sigma = \sqrt{30,25} = 5,5 \text{ mg}$$

$$se_{\bar{y}} = 5,5 / \sqrt{300} = 0,318 \text{ mg}$$

$$P(\bar{Y} - z \cdot se_{\bar{y}} < \bar{y} < \bar{Y} + z \cdot se_{\bar{y}}) = \gamma$$

$$P(15,4 - 1,96 \cdot 0,318 < \bar{y} < 15,4 + 1,96 \cdot 0,318) = 95 \%$$

$$P(14,777 < \bar{y} < 16,023) = 95 \%$$

S 95-odstotno verjetnostjo ocenjujemo, da je povprečna količina zaužitih vitaminov v osnovni statistični množici med 14,777 in 16,023 mg.

b)

$$H_0: \bar{y}_D = 13,8 \text{ mg}$$

$$H_1: \bar{y}_D \neq 13,8 \text{ mg}$$

$$\bar{y} = 15,4 \text{ mg}$$

$$\sigma = 5,5 \text{ mg}$$

$$se_{\bar{y}} = 0,318 \text{ mg}$$

$$\alpha = 10 \% \text{ (oziroma } 0,10), \text{ kar pomeni, da je kritična vrednost spremenljivke } z = \pm 1,645$$

Izračunamo  $z$  po enačbi (izračun testne vrednosti pri preizkušanju domneve o aritmetični sredini –  $z$ -test):

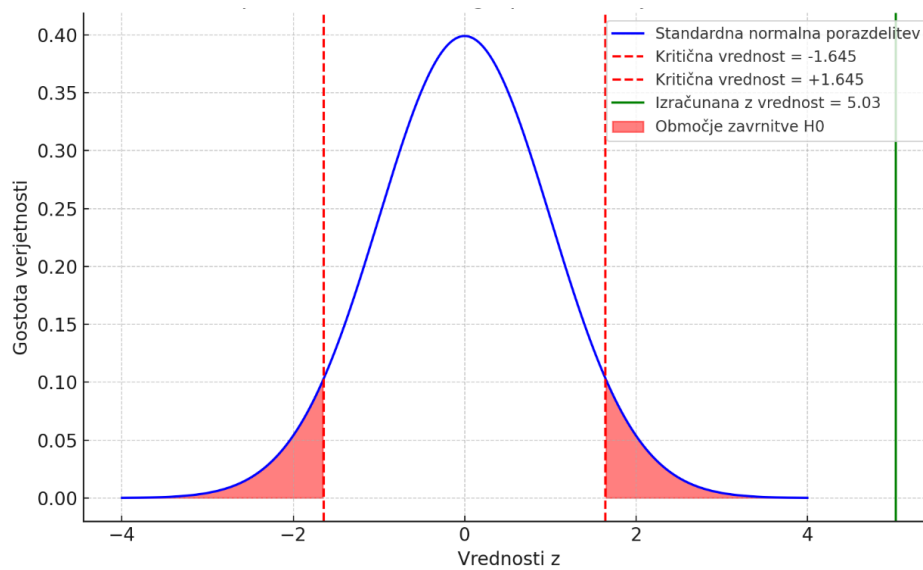
$$z = \frac{\bar{Y} - \bar{y}_D}{SE_{\bar{y}}}$$

$$z = \frac{15,4 - 13,8}{0,318} = 5,03 \quad (\text{dobljena standardizirana vrednost vzorčne vrednosti ne pade v interval } \pm 1,645, \text{ pač pa izven intervala})$$

Na osnovi rezultata vidimo, da povprečna količina dnevno zaužitih vitaminov ni enaka 13,8 mg, tako da zavrnamo domnevo  $H_0$  in sprejmemo raziskovalno domnevo  $H_1$  z manj kot desetodstotnim tveganjem oziroma verjetnostjo, da smo se pri tem našem zaključku zmotili.

Na grafu sta z rdečima črkanima črtama označeni kritični vrednosti  $z = \pm 1,645$ . Te vrednosti določajo meje območja zavrnitve ničelne hipoteze  $H_0: \bar{y}_D = 13,8 \text{ mg}$  pri tveganju  $\alpha = 10 \%$ , razdeljenem na dve strani porazdelitve, kar pomeni 5 % na vsakem koncu.

Grafični prikaz dvostranskega preizkušanja domneve (z-test):



Izračunana z vrednost  $z_i = 5,03$ , prikazana z zeleno črto, se nahaja precej desno od pozitivne kritične vrednosti, kar pomeni, da je vrednost izven območja sprejemanja ničelne hipoteze. To nakazuje, da pri stopnji tveganja 10 % zavrnilo ničelno hipotezo. Na osnovi rezultata sprejmemo hipotezo H1, kar pomeni da je povprečna količina dnevno zaužitih vitaminov statistično značilno različna od 13,8 mg ( $H_1: \bar{y}_D \neq 13,8 \text{ mg}$ ). Območje zavrnitve hipoteze H0 je poudarjeno z rdečo barvo in se nahaja na obeh koncih porazdelitve.



### Primer rešene naloge

V sedmih naključno izbranih šolah v določeni občini je med učenci naslednji odstotek tistih, ki prejemajo štipendijo: 15 %, 22 %, 18 %, 25 %, 20 %, 30 % in 10 %.

- Izračunajte in pojasnite interval zaupanja za povprečen odstotek učencev, ki prejemajo štipendijo, med učenci v statistični množici. Oceno napravite s petodstotnim tveganjem.
- Pri stopnji tveganja 5 % preizkusite domnevo, da je povprečen odstotek učencev, ki prejemajo štipendijo, več kot 23 %.

Imamo mali vzorec, ker je  $n = 7$ , ter dvostransko intervalno ocenjevanje aritmetične sredine:

$$\alpha = 5 \%$$

$$\bar{Y} = \frac{1}{7} \cdot (15 + 22 + \dots + 10) = 20 \%$$

Povprečen odstotek učencev, ki prejemajo štipendijo, znaša 20 %.

Uporabimo enačbo za izračun vzorčne variance iz posamičnih vrednosti:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{Y})^2$$

$$s^2 = \frac{1}{6} \cdot [(15 - 20)^2 + (22 - 20)^2 + \dots + (10 - 20)^2] = \frac{1}{6} \cdot 258 = 43 (\%)^2$$

$$s = 6,56 \%$$

$$se_{\bar{y}} = 6,56 / \sqrt{7} = 2,48 \%$$

$$t_{n-1; \frac{\alpha}{2}} = t_{6; 0,025} = 2,447 \text{ (gledamo tabelo kritična vrednost za } t \text{ porazdelitev)}$$

$$\frac{\alpha}{2} = \frac{0,05}{2} = 0,025$$

$$P(20 - 2,447 \cdot 2,48 < \bar{y} < 20 + 2,447 \cdot 2,48) = 95 \%$$

$$P(13,93 < \bar{y} < 26,07) = 95 \%$$

S 95-% verjetnostjo ocenjujemo, da povprečen odstotek učencev, ki prejemajo štipendijo, med učenci v vseh šolah znaša med 13,93 % in 26,07 %.

b)

Enostransko preizkušanje domneve:

$$\alpha = 5 \% \rightarrow t_{n-1; \alpha} = t_{6; 0,05} = + 1,943 \text{ (kritična vrednost spremenljivke } t)$$

$$H_0: \bar{y}_D \geq 23 \%$$

$$H_1: \bar{y}_D < 23 \%$$

$$\bar{Y} = 20 \% \text{ (izračun pri nalogi a)}$$

$$se_{\bar{y}} = 2,48 \% \text{ (izračun pri nalogi a)}$$

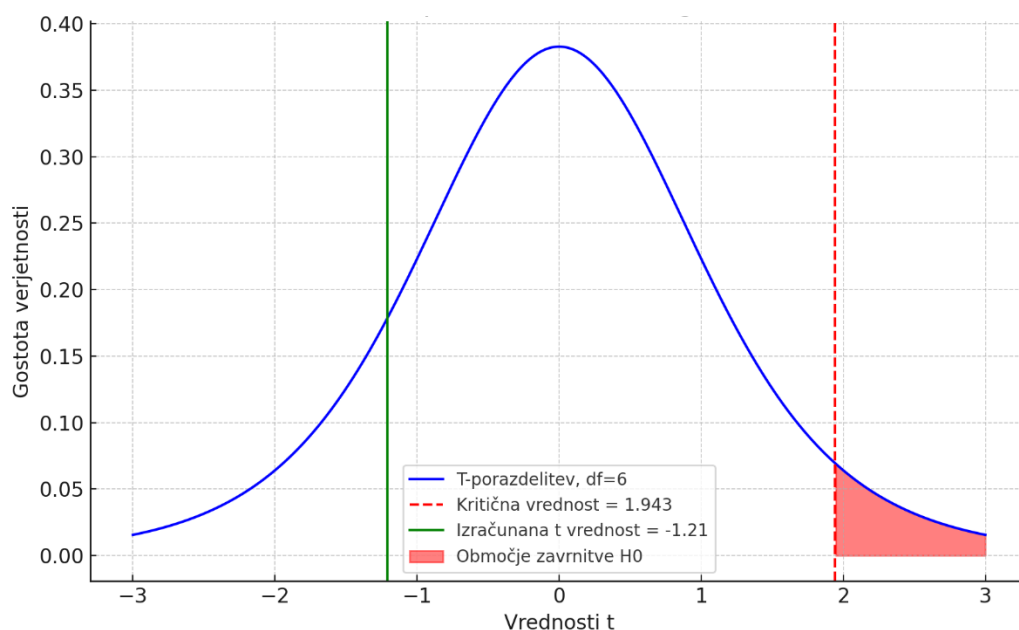
Izračunamo  $t$  po enačbi (testna vrednost pri preizkušanju domneve o aritmetični sredini –  $t$ -test):

$$t = \frac{\bar{Y} - \bar{y}_D}{SE_{\bar{y}}}$$

$$t = \frac{20 - 23}{2,48} = -1,21$$

Sprejmemo domnevo  $H_0$ .

Grafični prikaz enostranskega preizkušanje domneve ( $t$ -test):



Modra črta predstavlja  $t$ -porazdelitev s šestimi stopnjami prostosti. Rdeča črtkana črta označuje kritično vrednost  $t_{6;0,05} = +1,943$ , ki določa mejo za območje zavrnitve  $H_0$ . Zelena črta predstavlja izračunano  $t$  vrednost  $-1,21$ , ki se nahaja levo od kritične vrednosti, kar pomeni znotraj območja sprejemanja. Rdeče obarvano območje na desni predstavlja območje zavrnitve hipoteze  $H_0$  pri stopnji tveganja 5 %. Ker je izračunana  $t$  vrednost  $-1,21$  levo od kritične vrednosti, to pomeni, da na podlagi podatkov ne moremo zavrniti ničelne hipoteze  $H_0: \bar{y}_D \geq 23 \%$ .

**Naloga 44:** Naštejte in opišite tehnike čistega verjetnostnega vzorčenja, ki temeljijo na enostavnem slučajnem vzorčenju. Opišite možne načine za izbiro enot v vzorec pri enostavnem slučajnem vzorčenju.

---

---

---

---

---

---

**Naloga 45:** Opišite značilnosti neverjetnostnega vzorčenja in tehnike neverjetnostnega vzorčenja.

---

---

---

---

---

---

**Naloga 46:** Na osnovi izvlečka iz tabele slučajnih števil izberite iz neke populacije  $N = 400$  enot v enostavni slučajni vzorec  $n = 6$  enot.

---

---

---

---

---



## Naloga 47

Vzemimo, da opazovano populacijo predstavljajo štirje študenti, označimo jih z A, B, C in D, ki imajo takšno število opravljenih izpitov na današnji dan: 3, 5, 4 in 2. V vzorec izberimo  $n = 2$  študenta.

- Izračunajte število vseh možnih vzorcev.
- V primeru vzorcev brez ponavljanja prikažite v tabeli vse možne vzorce in izračunajte in pojasnite vzorčne aritmetične sredine v vseh možnih vzorcih.
- Izračunajte aritmetično sredino iz vseh vzorčnih ocen aritmetične sredine. Izračunajte aritmetično sredino iz podatkov v statistični množici. Kakšen rezultat pričakujete?
- Narišite graf porazdelitve vzorčnih ocen aritmetične sredine iz vseh možnih vzorcev.
- Izračunajte vzorčne variance in nepristranske ocene vzorčnih varianc. Izračunajte tudi varianco iz statistične množice. Kakšen rezultat pričakujete?







## Naloga 48

Z uporabo sistematičnega vzorčenja izberite iz statistične množice  $N = 200$  zaposlenih oseb v vzorec:

- a)  $n = 5$  zaposlenih
- b)  $n = 10$  zaposlenih
- c)  $n = 20$  zaposlenih

Izračunajte in pojasnite za posamezni primer vzorčni delež ter določite prvega in vse nadaljnje zaposlene v vzorec.

**Naloga 49**

Statistična množica obsega  $N = 1500$  zaposlenih, ki jih opazujemo po osebnem dohodku:

Skupina A: 700 zaposlenih, ki dosegajo osebni dohodek do pod 800 €

Skupina B: 500 zaposlenih, ki dosegajo osebni dohodek od 800 do pod 1500 €

Skupina C: 300 zaposlenih, ki dosegajo osebni dohodek 1500 € ali več.

Skupno število statističnih enot v vzorcu naj bo  $n = 200$ . Pri izbiri enot v vzorec uporabite:

- a) metodo stratificiranega – proporcionalnega vzorčenja,
- b) metodo stratificiranega – disproporcionalnega vzorčenja.



## Naloga 50

V slučajnem vzorcu je bilo zajetih  $n = 300$  tekačev, ki so se udeležili maratona. Povprečni čas, ki so ga tekači porabili, da so pretekli celotno razdaljo maratona, je bil 180 minut, z nepristransko oceno variance  $74,13 \text{ minut}^2$ .

- Izračunajte 95-% interval zaupanja za povprečni čas pretečenega maratona v osnovni statistični množici.
- Izračunajte 99-% interval zaupanja za aritmetično sredino statistične množice – enostranski pristop, in sicer s spodnjo in z zgornjo mejo.
- Izračunajte 90-% interval zaupanja za povprečni čas pretečenega maratona v osnovni statistični množici, če smo v slučajni vzorec zajeli  $n = 25$  tekačev.



## Naloga 51

V šestih naključno izbranih srednjih šolah so izmerili povprečno število knjig, ki jih dijaki izposodijo iz šolske knjižnice v enem mesecu: 23, 47, 5, 45, 65 in 12. Izračunajte in pojasnite interval zaupanja za povprečno število knjig, izposojenih na dijaka med srednjimi šolami v statistični množici. Oceno napravite s petodstotnim tveganjem.



## Naloga 52

V mestni knjižnici, ki ima registriranih 11.125 članov, so želeli ugotoviti, kakšen odstotek (%) članov si aktivno izposoja knjige. Iz slučajnega vzorca 700 članov so ugotovili, da si 235 članov ni izposodilo nobene knjige v zadnjem mesecu. Upoštevajte stopnjo tveganja 0,10.



## Naloga 53

Naključno smo izvedli anketo med 45 startup podjetji o njihovem letnem dobičku. Skupni letni dobiček v teh podjetjih je bil 4.500.000 d.e., nepristranska ocena standardnega odklona pa 890 d.e. Z 99-% verjetnostjo ocenite povprečni letni dobiček startup podjetja v statistični množici.

Izračunajte še 95-% interval zaupanja za povprečni letni dobiček startup podjetja v osnovni statistični množici, če smo v slučajni vzorec zajeli  $n = 15$  startup podjetji.



## Naloga 54

Na ravni značilnosti  $\alpha=0,10$  preizkusite domnevo da je povprečna količina recikliranih materialov za 120 gospodinjstev 5000 kg, če smo iz te skupine naključno izbrali 36 gospodinjstev in zanje ugotovili, da je:

$$\sum y_i = 187.200 \quad \text{in} \quad \sum (y_i - \bar{y})^2 = 28.000.000.$$





## Naloga 55

V okviru analize dnevnega trgovanja na borzi bomo raziskali, ali lahko trdimo, da vlagatelji na določeni borzi v povprečju zaslužijo 100 d.e. na dan, če smo za 50 naključno izbranih vlagateljev ugotovili, da v povprečju zaslužijo 102 d.e. na dan in je nepristranska ocena variance dobička  $6,25 \text{ (d.e.)}^2$ . Upoštevajmo stopnjo tveganja 0,05. V naslednjem koraku preizkusite domnevo, da 50 naključno izbranih vlagateljev v povprečju zasluži manj kot 103 d.e. na dan.



## Naloga 56

V raziskavi se želi oceniti povprečno število prebranih knjig na leto med študenti Ekonomsko-poslovne fakultete v Mariboru. V enostavnem naključnem vzorcu so zajeli  $n = 25$  študentov. Študente so analizirali glede na število prebranih knjig v enem letu. Aritmetična sredina znaša 30 knjig, standardni odklon pa je 5 knjig.

- a) Pri 10-% stopnji tveganja preizkusite domnevo, da je povprečno število prebranih knjig na študenta 28 ali več.
- b) Pri 1-% stopnji tveganja preizkusite domnevo, da je povprečno število prebranih knjig na študenta enako 22 knjig.
- c) Pri 5-% stopnji tveganja preizkusite domnevo, da je povprečno število prebranih knjig na študenta manj kot 20 knjig.

**Naloga 57**

Skupina inženirjev je delala na izboljšanju algoritmov za samovozeče avtomobile, s posebnim poudarkom na zmanjševanju reakcijskega časa samovozečih vozil v kritičnih situacijah. Za preizkus so izbrali sedem prototipov samovozečih avtomobilov. Merili so, koliko časa (v milisekundah) je potrebno vsakemu vozilu, da pravilno reagira na nenadno oviro na cesti. Rezultati testov so bili: 25 ms, 30 ms, 21 ms, 60 ms, 17 ms, 14 ms, 24 ms.

- a) Določite 80-% interval zaupanja za povprečni reakcijski čas samovozečih avtomobilov pri zaznavanju nenadnih ovir.
- b) Pri stopnji tveganja 5 % preizkusite domnevo, da je povprečni reakcijski čas samovozečih avtomobilov v kritičnih situacijah največ 30 ms.



## Naloga 58

V okviru iniciative za zmanjšanje količine odpadne hrane so ekološki raziskovalci opravili študijo na vzorcu petih restavracij, da bi ocenili vpliv uvajanja trajnostnih praks. Zbrali so podatke o količini hrane v kilogramih, ki so jo restavracije uspele ohraniti in preprečiti, da bi postala odpadek, v obdobju treh mesecev po uvedbi trajnostnih praks. Zabeleženi četrtni prihranki hrane so bili: 186 kg, 216 kg, 221 kg, 266 kg, 227 kg.

- a) Določite 95-% interval zaupanja za povprečni četrtni prihranek hrane po uvedbi trajnostnih praks v sodelujočih restavracijah.
- b) Izračunajte 90-% interval zaupanja za aritmetično sredino statistične množice – enostranski pristop, in sicer s spodnjo in z zgornjo mejo.
- c) Pri stopnji tveganja 10 % preizkusite domnevo, da povprečni četrtni prihranek hrane presega 220 kg.

**Naloga 59**

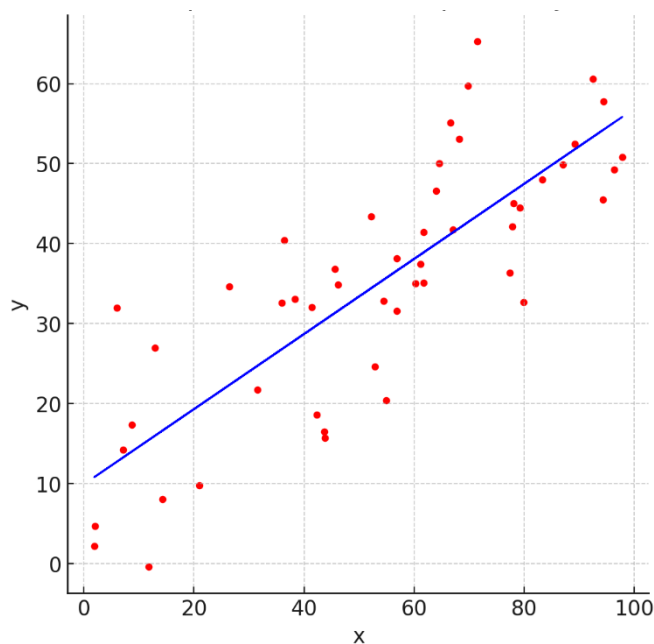
V banki X je bilo v slučajni vzorec izbranih 340 imetnikov vrednostnih papirjev, ki imajo v vrednostne papirje vložene naslednje zneske:

Znesek v 10 <sup>2</sup> EUR	Število oseb
Od 10 do pod 20	45
od 20 do pod 40	132
od 40 do pod 80	92
od 80 do pod 120	41
od 120 do pod 200	22
od 200 do pod 500	8
<b>Skupaj</b>	<b>340</b>

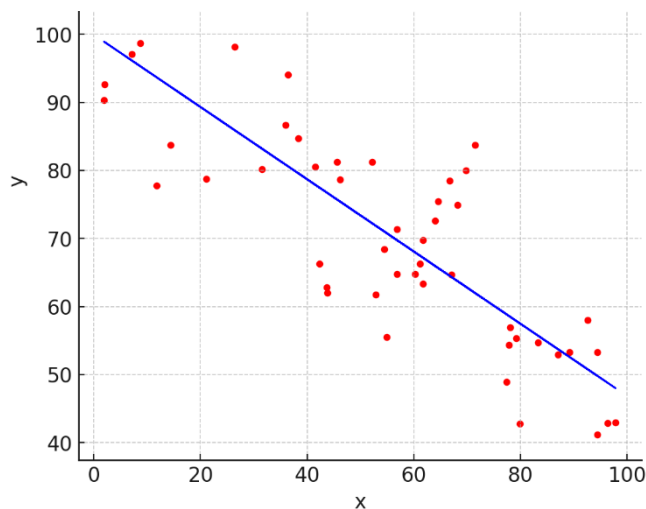
- S 95-% verjetnostjo določite interval zaupanja za povprečni znesek, ki so ga v vrednostne papirje vložili imetniki vrednostnih papirjev pri tej banki.
- Na ravni značilnosti  $\alpha = 0,10$  preverite domnevo, da je povprečen znesek, ki so ga lastniki vložili v vrednostne papirje, 7.500 EUR.

## 2.8 Regresijska analiza

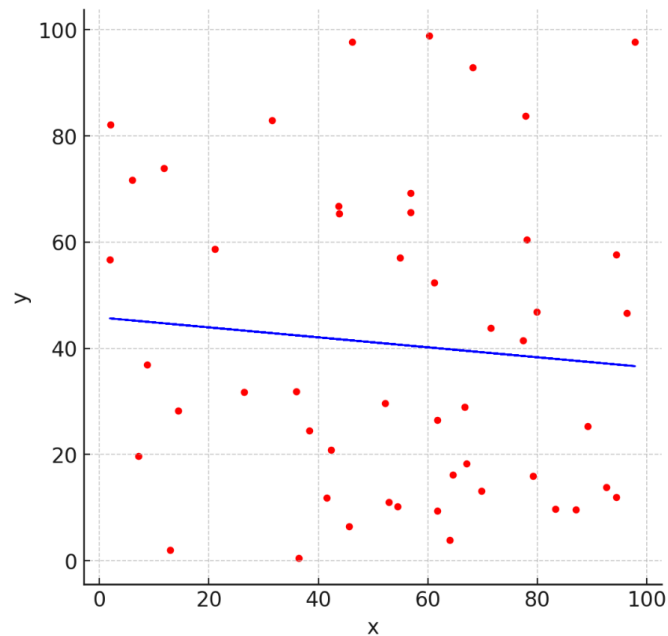
Z enostavno linearno regresijo analiziramo odvisnost med odvisno ( $y$ ) in eno neodvisno (ali pojasnjevalno) spremenljivko ( $x$ ). Grafični prikaz, ki ga uporabljamo pri enostavni regresijski analizi, se imenuje *razsevni grafikon*. Razsevni grafikon je grafični prikaz povezanosti med dvema spremenljivkama. Slika 8 prikazuje primer pozitivne povezanosti med odvisno in neodvisno spremenljivko z vrisano regresijsko premico. Slika 9 prikazuje primer negativne povezanosti med odvisno in neodvisno spremenljivko z vrisano regresijsko premico in slika 10 prikazuje primer, ko ni povezanosti med spremenljivkama.



Slika 8: Pozitivna povezanost med odvisno in neodvisno spremenljivko



Slika 9: Negativna povezanost med odvisno in neodvisno spremenljivko



Slika 10: Ni povezanosti med odvisno in neodvisno spremenljivko

### *Kazalci enostavne linearne regresije*

Regresijski model predvideva, da je vrednost odvisne spremenljivke odvisna od vrednosti pojasnjevalne spremenljivke ter od drugih spremenljivk in slučajnih vplivov, ki jih nismo eksplicitno vključili v model:  $y = f(x) + e$ . Pri tem  $e$  označuje t.i. *preostanek modela* (Montgomery idr., 2021).

Vrednost  $f(x)$  lahko opredelimo z linearno funkcijo in v tem primeru govorimo o linearni regresijski funkciji:  $\hat{y}_i = a + b \cdot x$  ali  $\hat{y}_i = b_0 + b_1 \cdot x_i$  (Montgomery idr., 2021).

Kovarianca ( $C_{xy}$ ) pove, ali sta spremenljivki povezani ter kakšna je smer njune povezanosti. Kadar je kovarianca različna od 0, pomeni, da sta spremenljivki medsebojno povezani. V primeru, ko je kovarianca večja od 0, prevladuje pozitivna smer povezanosti med spremenljivkama, in kadar je kovarianca manjša od 0, prevladuje negativna smer povezanosti (Aickin, 2010).

Ocenjeni vrednosti obeh regresijskih koeficientov: regresijska konstanta ( $a$ ) pove povprečno vrednost odvisne spremenljivke ( $y$ ), ko je neodvisna spremenljivka  $x$  enaka 0, regresijski koeficient pri neodvisni spremenljivki ( $b$ ) pa izraža, za koliko enot se v povprečju spremeni vrednost odvisne spremenljivke, če se neodvisna spremenljivka spremeni za eno enoto (Holmes, 2018).

Determinacijski koeficient ( $r^2_{xy}$ ) pove, kolikšen % celotne variance spremenljivke  $y$  (odvisna spremenljivka) je pojasnjen z regresijsko funkcijo oz. s spremenljivko  $x$  (neodvisna spremenljivka). Opredeljuje jakost linearne povezanosti med spremenljivkama. Vrednost determinacijskega koeficienta se giblje med 0 in 1 ( $0 \leq r^2_{xy} \leq 1$ ) (Holmes, 2018).

Korelacijski koeficient ( $r_{xy}$ ) opredeljuje jakost in smer linearne povezanosti med odvisno in neodvisno spremenljivko. Vrednost korelacijskega koeficienta se giblje med  $-1$  in  $1$  ( $-1 \leq r_{xy} \leq 1$ ) (Holmes, 2018, Montgomery idr., 2021).

Standardna napaka ocene odvisne spremenljivke ( $\sigma_y$ ) pokaže, ali na variabilnost spremenljivke  $y$ , razen spremenljivke  $x$ , vplivajo še druge spremenljivke in slučajni vplivi (Seber in Lee, 2003).

Točkovna ocena vrednosti spremenljivke  $y$  pri izbrani vrednosti spremenljivke  $x = x_0$  je pridobljena tako, da vrednost spremenljivke  $x_0$  vstavimo v regresijsko enačbo. Pri intervalni oceni vrednosti spremenljivke  $y$  pri izbrani vrednosti spremenljivke  $x$  pa upoštevamo, da na odvisno spremenljivko vplivajo še druge spremenljivke in slučajni vplivi (Tabachnick in Fidel, 2013). Intervalna ocena pomeni, da z določeno stopnjo verjetnosti ocenimo, kakšno vrednost spremenljivke  $y$  lahko v povprečju pričakujemo pri izbrani vrednosti spremenljivke  $x = x_0$ , če upoštevamo tudi standardno napako ocene odvisne spremenljivke (Tominc in Kramberger, 2007).



## Primer rešene naloge

V tabeli so podatki o tedenskem številu potrjenih primerov s SARS-CoV-2 glede na dve različni lokaciji prenosa okužbe: javni prevoz in restavracije. Zanima nas, ali okuženost oseb v restavracijah vpliva na povečanje okuženosti oseb v javnem prevozu.

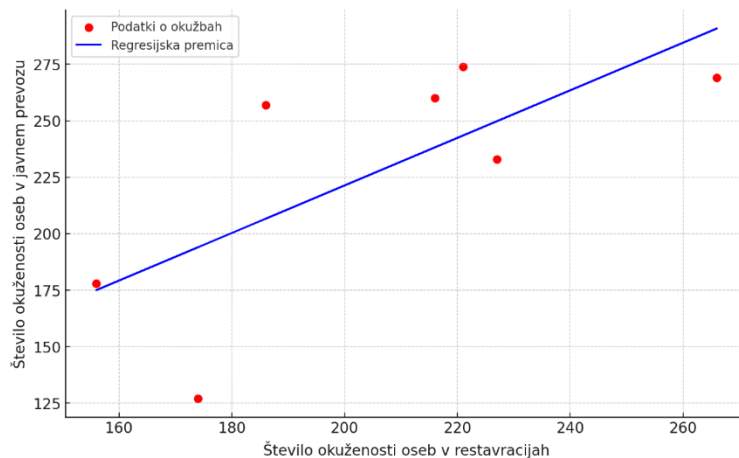
Podatki o tedenskem številu potrjenih primerov s SARS-CoV-2 v javnem prevozu in restavracijah:

Teden	Število okuženosti oseb v javnem prevozu (y)	Število okuženosti oseb v restavracijah (x)
1.	257	186
2.	260	216
3.	274	221
4.	269	266
5.	233	227
6.	178	156
7.	127	174



- Odvisno in neodvisno spremenljivko prikažite v razsevnem grafikonu.
- Ocenite osnovne značilnosti povezanosti med spremenljivkama ter izračunajte in pojasnite vse kazalce linearne korelacije in regresije.
- Izpišite enačbo regresijske premice.
- Ob upoštevanju linearne korelacijske odvisnosti ocenite z verjetnostjo 95 % število okuženih oseb v javnem prevozu pri  $x = 195$ , intervalna ocena (upoštevajte, da je popravek  $h_1$  enak 0).

a)



Na razsevnem grafikonu je prikazan odnos med številom okuženih oseb v restavracijah in številom okuženih oseb v javnem prevozu. Kot kaže regresijska premica, obstaja korelacija med številom okužb v restavracijah in številom okužb v javnem prevozu, kar pomeni, da se s številom okuženosti oseb v restavracijah (neodvisna spremenljivka) povečuje tudi število okuženosti oseb v javnem prevozu (odvisna spremenljivka).

b)

$$n = 7$$

$$\sum x_i = 186 + 216 + 221 + 266 + 227 + 156 + 175 = 1.446$$

$$\sum x_i^2 = 186^2 + 216^2 + 221^2 + 266^2 + 227^2 + 156^2 + 175^2 = 306.990$$

$$\sum y_i = 257 + 260 + 274 + 269 + 233 + 178 + 127 = 1.598$$

$$\sum y_i^2 = 257^2 + 260^2 + 274^2 + 269^2 + 233^2 + 178^2 + 127^2 = 383.188$$

$$\sum x_i \cdot y_i = (186 \cdot 257) + (216 \cdot 260) + (221 \cdot 274) + \dots + (174 \cdot 127) = 338.827$$

$$\bar{x} = \frac{1446}{7} = 206,571$$

$$\bar{y} = \frac{1598}{7} = 228,286$$

Oba regresijska koeficienta izračunamo po enačbi:

$$b_1 = \frac{\frac{1}{n-1} \sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\frac{1}{n-1} \sum x_i^2 - \bar{x}^2} = \frac{c_{xy}}{s_x^2} = \frac{(\sum x_i \cdot y_i) - n\bar{x}\bar{y}}{(n-1) \cdot s_x^2} = \frac{(\sum x_i \cdot y_i) - n\bar{x}\bar{y}}{(\sum x_i^2) - n\bar{x}^2}$$

$$b_0 = \bar{y} - b_1 \cdot \bar{x}$$

$$b_1 = \frac{(338827) - 7 \cdot 206,571 \cdot 228,286}{(306990) - 7 \cdot 206,571^2} = 1,052$$

$$b_0 = 228,286 - 1,052 \cdot 206,571 = 10,973$$

Korelacijski koeficient izračunamo po enačbi:

$$r_{xy} = \frac{(\sum x_i \cdot y_i) - n\bar{x}\bar{y}}{(n-1) \cdot s_x \cdot s_y}$$

Izračun  $s_x$  in  $s_y$ :

$$s_x^2 = \frac{1}{n-1} \cdot \sum (x_i - \bar{x})^2$$

$$s_x^2 = \frac{1}{6} \cdot [(186 - 206,571)^2 + (216 - 206,571)^2 + \dots + (174 - 206,571)^2] = 1.381,286$$

$$s_x = \sqrt{1381,286} = 37,166$$

$$s_y^2 = \frac{1}{n-1} \cdot \sum (y_i - \bar{y})^2$$

$$s_y^2 = \frac{1}{6} \cdot [(257 - 228,286)^2 + (260 - 228,286)^2 + \dots + (127 - 228,286)^2] = 3.064,571$$

$$s_y = \sqrt{3064,571} = 55,359$$

$$(\sum x_i \cdot y_i) - n \cdot \bar{x} \cdot \bar{y} = 338.827 - 7 \cdot 206,571 \cdot 228,286 = 8.726,129$$

$$r_{xy} = \frac{8726,129}{6 \cdot 37,166 \cdot 55,359} = 0,707$$

Na osnovi rezultata ( $r_{xy} = 0,707$ ) vidimo, da obstaja srednje močna povezanost med odvisno (okuženost oseb v javnem prevozu) in neodvisno spremenljivko (okuženost oseb v restavracijah). Smer povezanosti je pozitivna.

Determinacijski koeficient:

$$r_{xy}^2 = 0,707^2 = 49,98 \%$$

Delež pojasnjene variance v skupni varianci za odvisno spremenljivko znaša 49,98 %. Standardno napako ocene odvisne spremenljivke izračunamo po enačbi:

$$s_{y.x} = \sqrt{\frac{(\sum y_i^2) - b_0(\sum y_i) - b_1(\sum x_i \cdot y_i)}{n - 2}}$$

$$s_{y.x} = \sqrt{\frac{383188 - (10,973 \cdot 1598) - (1,052 \cdot 338827)}{7 - 2}} = 42,91$$

Standardna napaka ocene odvisne spremenljivke je različna od 0, kar pomeni, da na okuženost oseb v javnem prevozu (odvisna spremenljivka) poleg okuženosti oseb v restavracijah (neodvisna spremenljivka) vplivajo še druge spremenljivke in slučajni vplivi.

c)

Izračunana regresijska koeficienta vstavimo v enačbo regresijske premice:

$$\hat{y}_i = b_0 + b_1 x_i$$

$$\hat{y} = 10,973 + 1,052 \cdot x_i$$

Izračunana regresijska koeficienta vstavimo v enačbo regresijske premice:

Pomen regresijskega koeficienta  $b_0$ : Pri okuženosti oseb v restavracijah  $x = 0$  lahko v povprečju pričakujemo, da bo okuženost oseb v javnem prevozu 10,973 oz. 11 oseb.

Pomen regresijskega koeficienta  $b_1$ : Če se okuženost oseb v restavracijah ( $x$ ) poveča za eno enoto oz. osebo, se okuženost oseb v javnem prevozu ( $y$ ) v povprečju poveča za 1,052 oz. eno osebo.

d)

Intervalno oceno izračunamo po enačbi:

$$\hat{y}_{x_0} \pm t_{n-2; \alpha/2} \cdot s_{y.x} \cdot \sqrt{1 + h_1}$$

$$\hat{y}_{x=195} = 10,973 + 1,052 \cdot 195$$

$$\hat{y}_{x=195} = 216,113$$

$$s_{y.x} = 42,91 \text{ (izračunano pri nalogi a)}$$

Upoštevamo, da je pri  $\gamma = 95 \%$ ,  $\alpha = 5 \%$ .

Izračun:  $t_{n-2, \gamma/2} = t_{5, 0,025} = 2,571$  (gledamo tabelo *kritične vrednosti za t porazdelitev*)

Upoštevamo, da je popravek  $h_1$  enak 0.

Izračunamo intervalno oceno:

$$P(216,113 - 2,571 \cdot 42,91 \cdot 1 < y_{x=195} < 216,113 + 2,571 \cdot 42,91 \cdot 1) = 95 \%$$

$$P(105,791 < y_{x=195} < 326,435) = 95 \%$$

Pri številu okuženih oseb v restavracijah pri  $x = 195$  bo število okuženih oseb v javnem prevozu med 105,791 in 326,435 oseb, kar trdimo s 95 % verjetnostjo.



## Naloga 60

Podatki o šestih državah, za katere so zabeležene investicije v prehod na pametno gospodarstvo (v  $10^6$  €) in ustvarjen BDP (v  $10^6$  €) v določenem letu, so naslednji:

Država	Investicije v pametno gospodarstvo (x)	Ustvarjen BDP (y)
A	115	328
B	130	330
C	140	390
D	149	361
E	160	421
F	171	400

- S prikazom dvojic vrednosti opazovanih spremenljivk v razsevnem grafikonu določite obliko, smer in jakost povezanosti med spremenljivkama.
- Z metodo najmanjših kvadratov izračunajte parametre regresijske premice, izračunano regresijsko premico vrišite v razsevni grafikon.
- Ocenite ustvarjen BDP za državo, ki bi investirala  $x = 180$  (v  $10^6$  €) v prehod na pametno gospodarstvo, točkovna ocena.
- Izračunajte parameter, na osnovi katerega določite smer in jakost linearne korelacijske odvisnosti.
- Izračunajte delež pojasnjene variance v skupni varianci za odvisno spremenljivko.
- Izračunajte standardno napako ocene odvisne spremenljivke.
- Ob upoštevanju linearne korelacijske odvisnosti ocenite z verjetnostjo 95 % ustvarjen BDP pri  $x = 180$  (v  $10^6$  €), ob upoštevanju, da je popravek  $h_1$  enak 0.





**Naloga 61**

V podjetju X so želeli analizirati vpliv usposabljanj za zaposlene na rast prihodkov podjetja. Tako so v podjetju ob različnem številu opravljenih usposabljanj za zaposlene ( $x$ ) zabeležili naslednje stopnje rasti prihodkov ( $y$  %):

Število usposabljanj ( $x$ )	2	5	6	8	10
Rast prihodkov $y$ % ( $y$ )	3,5	5,2	5,5	6,4	7,8

- Ocenite z zanesljivostjo 95 % rast prihodkov podjetja pri usposabljanju zaposlenih  $x = 15$ .
- Pojasnite vse kazalce linearne regresije in korelacije, ki ste jih izračunali pod a).

**Naloga 62**

Pojasnite:

- smer in jakost povezanosti med odvisno in neodvisno spremenljivko, če je  $r_{xy} = -0,854$ ;
- determinacijski koeficient, če je  $r_{xy}^2 = 0,729$ ;
- velikost standardne napake ocene ( $\sigma_{ey}$ ), če je  $r_{xy} = -1$ .

---

---

---

---

---

---

---

---

---

---



**Naloga 63**

Predpostavimo, da je čas dostave (v minutah), ki ga podjetje za dostavo hrane potrebuje za dostavo naročila, odvisen od razdalje dostave (v kilometrih). Za pet naročil so podatki v preglednici.

Razdalja dostave (km)	Čas dostave (minute)
10	30
20	60
50	120
35	90
15	45

- Narišite razsevni grafikon in ga pojasnite.
- Izračunajte in pojasnite vse kazalce linearne korelacije in regresije.
- Ocenite čas dostave za naročilo, če je razdalja dostave 25 kilometrov. Upoštevajte  $\alpha = 5\%$  (upoštevajte, da je popravek  $h_1$  enak 0).





## Naloga 64

V preglednici so zbrani podatki o povprečni stopnji vključenosti uporabnikov (v %) za različne oglaševalske kampanje na družbenih omrežjih in povezanimi stroški teh kampanj (v d.e.). Predpostavljamo, da povprečna stopnja vključenosti uporabnikov (v %) vpliva na stroške oglaševalske kampanje (v d.e.).

Povprečna stopnja vključenosti uporabnikov (%)	Stroški oglaševalske kampanje (v d.e.)
78,70	80,99
68,70	82,48
63,90	76,50
71,70	101,76
70,80	100,25
80,50	169,19
66,90	78,74

Izračunajte koeficiente linearne regresijske funkcije in ocenite pričakovane stroške oglaševalske kampanje pri povprečni stopnji vključenosti uporabnikov 90 % (upoštevajte, da je popravek  $h_1$  enak 0). Intervalno oceno napravite s 95-odstotno verjetnostjo.



**Naloga 65**

V preglednici so podatki, ki prikazujejo povprečno oceno zadovoljstva strank (na lestvici od 1 do 100) z različnimi telekomunikacijskimi paketi in povprečno mesečno porabo podatkov (v gigabajtih – GB) teh strank. Ocenite osnovne značilnosti povezanosti med spremenljivkama ter izračunajte in pojasnite vse kazalce linearne korelacije in regresije.

Povprečna mesečna poraba podatkov (v GB) (y)	Povprečna ocena zadovoljstva strank (x)
105,1	100
96	91
93,3	90
101	87
85,7	69
84,1	63
72,3	61
69,7	60



## Naloga 66

Rezultati analize linearne korelacijske odvisnosti povprečne mesečne prodaje izdelka A ( $y$ ) od dnevnega števila oglaševanj izdelka A preko spleta ( $x$ ) s programom SPSS so v spodnji tabeli:

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0,840(a)	0,706	0,608	11,334

a Predictors: (Constant), dnevno število oglaševanj izdelka A preko spleta

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta	B	Std. Error
1	(Constant)	-7,647	36,282		-0,226	0,536
	Dnevno število oglaševanj izdelka A preko spleta	0,634	0,173	0,840	2,684	0,045

a Dependent Variable: povprečna prodaja izdelka A

- Izpišite in vsebinsko pojasnite vse kazalce linearne korelacije in regresije.
- Ocenite povprečno letno prodajo izdelka A ( $y$ ) pri dnevnem številu oglaševanj izdelka A preko spleta pri  $x = 35$ .

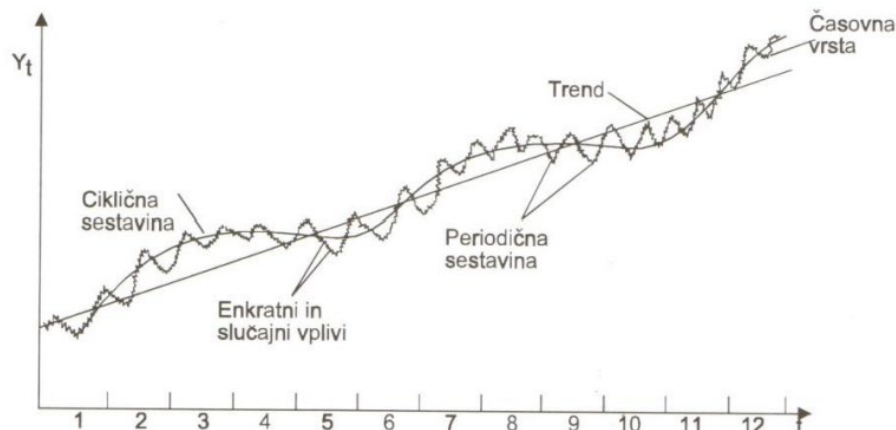


## 2.9 Napovedovanje vrednosti v časovni vrsti z uporabo trenda in sezonske komponente

Prvi korak pri analizi katerekoli časovne vrste je običajno grafična predstavitev podatkov – o grafičnem prikazu časovne statistične vrste smo govorili v drugem poglavju. Iz grafičnega prikaza lahko ocenimo, katere komponente ima časovna vrsta: *osnovno smer razvoja – trend, sezonsko ali periodično komponento* ter *slučajne vplive* (Brockwell in Davis, 2016). Tukaj bomo obravnavali komponento trenda, ki jo analitično opisujemo s *trendno funkcijo* (omejili se bomo predvsem na linearno funkcijo), ter sezonsko komponento, ki jo opisujemo s *sezonskimi ali periodičnimi indeksi*.

Trend prikazuje osnovno smer razvoja pojava in je najpomembnejša sestavina časovne vrste. V večini primerov je namreč mogoče ugotoviti, da se vrednosti spremenljivke gibljejo približno po neki zakonitosti, ki jo odraža linearna ali nelinearna funkcija trenda (Tominc, 2016). Trend lahko definiramo kot dolgoročno gibanje povprečja časovne vrste oz. smer razvoja časovne vrste. Sezonska komponenta vključuje vplive sezone in koledarja. Sem spadajo naravni in socialni vplivi, kot so spreminjanje letnih časov, število delovnih dni in vpliv praznikov. Sezonska komponenta ima periodo eno leto. Iz leta v leto se lahko počasi spreminja, njeni vplivi pa se znotraj enega leta približno izničijo (Golmajer, 2013).

Periodična sestavina časovnih vrst se odraža v spreminjanju vrednosti spremenljivke v krajših enakih časovnih podintervalih osnovne časovne enote. Ciklična sestavina časovnih vrst se kaže v spreminjanju vrednosti spremenljivke v enakih časovnih intervalih, običajno daljših od enega leta. Enkratni in slučajni vplivi pa vplivajo običajno na vrednost časovne spremenljivke v določeni časovni enoti (Tominc, 2016).



Slika 11: Prikaz sestavin v časovnih vrstah

Vir: Tominc, 2016



Slika 11 prikazuje časovno vrsto s trendom (linearnim), periodično in ciklično sestavino ter z enkratnimi vplivi.



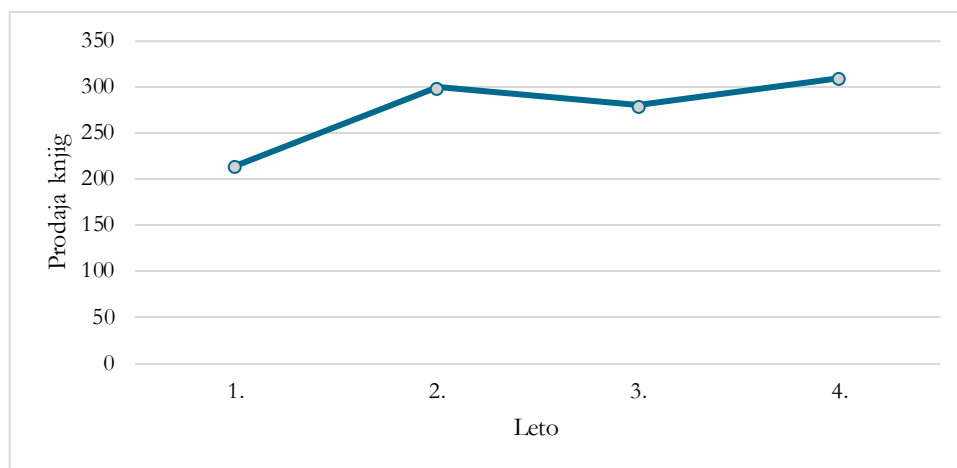
### Primer rešene naloge

V preglednici so podatki o prodaji knjig tujim študentom v knjigarni X v štirih letih:

Leto	2020	2021	2022	2023
Prodaja knjige	215	300	280	310

Časovno vrsto prikažite grafično. Izračunajte in pojasnite linearno funkcijo trenda ter ocenite število prodanih knjig tujim študentom v knjigarni X v letu 2024.

Grafični prikaz:



*Linearna funkcija trenda:*

*Sistem normalnih enačb:*

$$Ta + \left( \sum_{t=1}^T t \right) b = \sum_{t=1}^T Y_t$$

$$\left( \sum_{t=1}^T t \right) a + \left( \sum_{t=1}^T t^2 \right) b = \sum_{t=1}^T t Y_t,$$

Leto	t	$Y_t$	$t \cdot Y_t$	$t^2$
2020	1	215	215	1
2021	2	300	600	4
2022	3	280	840	9
2023	4	310	1240	16
Skupaj	10	1105	2895	30

$$\begin{array}{r} 4a + 10b = 1105 \quad / : 4, \cdot (-10) \\ 10a + 30b = 2895 \\ \hline 0 + 5b = 132,5 \end{array}$$

$\begin{array}{l} 1 \cdot (-10) + 10 = 0 \\ 2,5 \cdot (-10) + 30 = 5 \\ 276,25 \cdot (-10) + 2895 = 132,5 \end{array}$
--

$$b = \frac{132,5}{5} = 26,5$$

$$4a + 10 \cdot 26,5 = 1105$$

$$4a + 265 = 1105$$

$$4a = 840$$

$$a = \frac{840}{4} = 210$$

Funkcija trenda:  $\hat{Y} = 210 + 26,5 \cdot t$

Napoved za število prodanih knjig tujim študentom za 5. leto:  $\hat{Y}_t = 210 + 26,5 \cdot 5 = 342,5$



## Naloga 67

Na opazovanem področju smo zabeležili naslednje število nočitev tujih gostov:

Leto	I–IV	V–VIII	IX–XII	Skupaj
1	1.200	2.500	1.000	4.700
2	1.100	2.600	1.200	4.900
3	1.000	2.800	1.100	4.900
4	1.300	2.500	1.200	5.000
5	1.200	2.700	1.300	5.200

- Letne podatke, ki prikazujejo skupno letno število nočitev, narišite v linijskem grafikonu, s prostoročno metodo določite osnovno smer razvoja pojava; z analitično metodo določite parametre funkcije; ocenite število nočitev v prvem prihodnjem letu.
- Izračunajte sezonske indekse, upoštevajte oceno za število nočitev v prihodnjem letu na osnovi funkcije trenda (rezultat naloge a) in izračunane sezonske indekse ter predvidite število nočitev po sezonah v prihodnjem letu.

**Naloga 68**

V podjetju, ki se ukvarja s popravilom strojev, načrtujejo stroške v zvezi z njihovim delom. Na osnovi četrtnih podatkov za štiri zaporedna leta je podjetje izračunalo funkcijo trenda za stroške ( $t = \text{leto}$ ), in sicer:

$$\hat{Y} = 58,5 + 2,1t$$

- Ocenite stroške za prvo prihodnje leto.
- Sezonski indeks za zadnje četrtnje znaša  $SI_4 = 225,5 \%$ . Kaj ta indeks vsebinsko pomeni?
- Ocenite stroške popravil v zadnjem četrtnju prvega prihodnjega leta.

**Naloga 69**

Prodaja jagod je sezonskega značaja. V tabeli so podatki o prodaji jagod (v količinskih enotah) v trgovski verigi supermarketov, za štiri zaporedna leta, po 4-mesečjih (tri periode ali sezone znotraj leta):

Leto	Prvo 4-mesečje	Drugo 4-mesečje	Tretje 4-mesečje
1.	15	45	20
2.	20	55	15
3.	25	60	40

- Prikažite časovno vrsto po letih grafično.
- Izračunajte in pojasnite funkcijo trenda na osnovi letnih podatkov.
- Izračunajte in pojasnite sezonske indekse.
- Napovejte prodajo za 5. zaporedno leto, po 4-mesečjih.

**Naloga 70**

Podjetje, specializirano za izdelavo kozmetičnih izdelkov, želi analizirati in načrtovati svoje prihodnje prihodke. Na osnovi četrletnih prihodkov za štiri zaporedna leta je podjetje izračunalo funkcijo trenda za prihodke ( $t = \text{leto}$ ), in sicer:

$$\hat{Y} = 120 + 4,5t$$

- a) Ocenite prihodke za prvo prihodnje.
- b) Ocenite prihodke podjetja v zadnjem četrletju prvega prihodnjega leta. Sezonski indeks za zadnje četrletje znaša  $SI_4 = 195,4 \%$ .

**Naloga 71**

V preglednici so podatki o prodaji električnih avtomobilov podjetja X v regiji A v zadnjih petih letih:

Leto	2019	2020	2021	2022	2023
Prodaja električnih avtomobilov (v enotah)	80	120	200	300	450

- Kako imenujemo statistično vrsto v preglednici? Statistično vrsto prikažite grafično.
- Izračunajte in pojasnite linearno funkcijo trenda ter ocenite število prodanih električnih avtomobilov podjetja X v letu 2027.





## 3 Naloge za laboratorijske vaje

### 3.1 Deskriptivna statistika in vzorčni pristop

V prejšnjih poglavjih smo raziskali temelje deskriptivne statistike in vzorčenja ter predstavili ključne koncepte in teoretične osnove. V 1.3 podpoglavju smo se osredotočili na razumevanje in interpretacijo deskriptivne statistike, medtem ko smo v 1.5 podpoglavju razpravljali o metodah in tehnikah vzorčenja ter vzorčnem pristopu, ki so ključni za pridobivanje reprezentativnih podatkovnih nizov.

To poglavje gradi na tej teoretični podlagi in vas vodi skozi praktično uporabo programa SPSS, enega izmed najbolj razširjenih statističnih orodij. Osredotočili se bomo na to, kako lahko program SPSS uporabimo za izvajanje deskriptivnih statističnih analiz in vzorčenje v realnih raziskovalnih scenarijih. Skozi praktične primere in vaje bomo analizirali, kako transformirati podatke v uporabne informacije, ki so nujne za odločanje in znanstveno sporočanje.



## Naloga 1

Raziskovalci so želeli analizirati prodajo domačih proizvodov v letu 2023 v desetih izbranih trgovinah po Sloveniji. V preglednici so podani podatki o prodaji domačih proizvodov v letu 2023 za vsako izmed desetih trgovin (v % od celotne prodaje):

Trgovina	1	2	3	4	5	6	7	8	9	10
Prodaja domačih proizvodov (%)	15,2	18,5	20,1	22,3	23,7	19,8	17,4	16,6	21,9	20,1

- Opreделите statistično enoto, statistično spremenljivko in moč vzorca.
- Podatke o prodaji domačih proizvodov v letu 2023 za deset izbranih trgovin vnesite v program SPSS. Izračunajte in pojasnite rezultate opisne statistike za prodajo domačih proizvodov v letu 2023 za teh deset izbranih trgovin.
- Na osnovi mere asimetričnosti in mere sploščenosti pojasnite ali so vrednosti navedene spremenljivke normalno porazdeljene.
- Za spremenljivko prodaja domačih proizvodov v letu 2023 za deset izbranih trgovin prikažite frekvenčno preglednico.
- Narišite še frekvenčni histogram z normalno krivuljo.

Postopek vnosa podatkov: podatke o prodaji domačih proizvodov v letu 2023 za deset izbranih trgovin vnesemo v SPSS tako, da spodaj v levem kotu kliknemo desno okence *Variable View*. V okence *Name* vpišemo ime spremenljivke *Prodaja*, v okencu *Type* določimo vrsto spremenljivke (numeric), v okencu *Decimals* določimo število decimalnih mest, v okencu *Label* zapišemo celotno ime spremenljivke (*Prodaja domačih proizvodov v letu 2023 za deset izbranih trgovin*). Nato kliknemo spodnje levo okence *Data View* in zapišemo vrednosti spremenljivk v stolpec.

Izračun rezultatov opisne statistike: kliknemo *Analyze* in nato *Descriptive Statistics* ter *Descriptives (ali Frequencies)*. Odpre se pogovorno okno, v katerem kliknemo spremenljivko *Prodaja* in jo prenesemo v desno okence *Variable(s)*. Kliknemo okence *Options* in izberemo statistike, ki jih želimo za prikaz rezultata.

Frekvenčni histogram: kliknemo *Frequencies* in v desno okence *Variable(s)* prenesemo spremenljivko *Prodaja*. Kliknemo okence *Charts*, ter kliknemo na *Histograms*.

Odgovori in izpisi rezultatov:

a)

Statistična enota: 1 trgovina

Statistična spremenljivka: prodaja domačih proizvodov (numerična, zvezna spremenljivka).

b in c)

Tabela 1: Opisna statistika za spremenljivko prodaja domačih proizvodov v letu 2023 za deset izbranih trgovin

N	Valid	10
	Missing	0
	Mean	19,560
	Std. Error of Mean	,8454
	Median	19,950
	Mode	20,1
	Std. Deviation	2,6734
	Variance	7,147
	Skewness	-,112
	Std. Error of Skewness	,687
	Kurtosis	-,744
	Std. Error of Kurtosis	1,334
	Range	8,5
	Minimum	15,2
	Maximum	23,7

V tabeli 1 so prikazani rezultati opisne statistike za spremenljivko prodaja domačih proizvodov v letu 2023 za deset izbranih trgovin. Vidimo, da je bilo v vzorec vključenih 10 trgovin ( $n$ ) in ni nobene manjkajoče vrednosti (angl. Missing = 0).

Povprečna vrednost (angl. Mean) prodaje domačih proizvodov v letu 2023 za deset trgovin znaša 19,560 %.

Standardna napaka ocene aritmetične sredine (angl. Standard error of mean) znaša 0,8454 % in pomeni, da manjša kot je njena vrednost, manjša je variabilnost med vzorčnimi povprečnimi vrednostmi in boljši predstavnik statistične množice je vzorec.

Mediana (angl. Median) znaša 19,950 %, kar pomeni, da ima polovica (50 %) trgovin prodajo domačih proizvodov manjšo ali enako od 19,950 %, ter polovica (50 %) trgovin več kot 19,950 % prodajo domačih proizvodov.

Modus (angl. Mode) je 20,1 % in predstavlja tisto vrednost spremenljivke *prodaja domačih proizvodov* za deset izbranih trgovin, ki se najpogosteje pojavlja.

Standardni odklon je kvadratni koren iz variance in v našem primeru znaša 2,6734 %, ki opisuje razpršenost vrednosti spremenljivke okoli vrednosti aritmetične sredine. Varianca znaša 7,147 (%)<sup>2</sup>.

Koeficient asimetrije (angl. Skewness) znaša -0,112, kar pomeni, da je porazdelitev asimetrična v levo (negativna asimetrična porazdelitev). Koeficient sploščenosti (angl. Kurtosis) znaša -0,744, kar kaže na sploščeno porazdelitev (negativna vrednost).

Variacijski razmik (angl. Range) znaša 8,5 odstotnih točk in je enak razliki med največjo (angl. Maximum = 23,7 %) in najmanjšo (angl. Minimum = 15,2 %) vrednostjo spremenljivke.

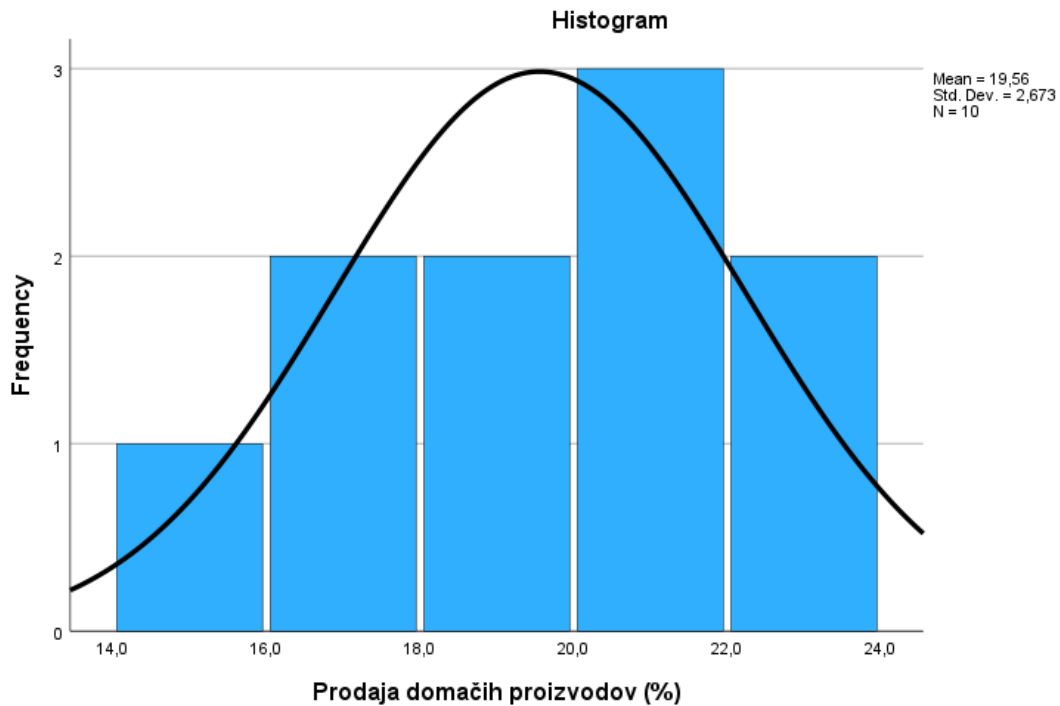
d)

**Tabela 2: Frekvenčna preglednica – prodaja domačih proizvodov v letu 2023 za deset trgovin**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	15,2	1	10,0	10,0	10,0
	16,6	1	10,0	10,0	20,0
	17,4	1	10,0	10,0	30,0
	18,5	1	10,0	10,0	40,0
	19,8	1	10,0	10,0	50,0
	20,1	2	20,0	20,0	70,0
	21,9	1	10,0	10,0	80,0
	22,3	1	10,0	10,0	90,0
	23,7	1	10,0	10,0	100,0
	Total	10	100,0	100,0	

Tabela 2 kaže, da je imela ena trgovina od desetih trgovin prodajo domačega proizvoda 15,2 %. Prav tako je imela ena trgovina od desetih trgovin prodajo domačega proizvoda 16,6 % in enako velja za ostale primere trgovin. Dve trgovini od desetih trgovin sta imeli prodajo domačega proizvoda 20,1 %.

e)



**Histogram 1: Frekvenčni histogram s krivuljo prilagojene normalne porazdelitve**



## Naloga 2

Raziskava se osredotoča na število knjig, ki jih posamezniki preberejo v enem letu, z namenom razumeti, kako se bralne navade razlikujejo med študenti in osebami, zaposlenimi za polni delovni čas (1 – študent, 2 – zaposleni). V naključni vzorec je bilo zajetih 14 oseb. Podatki so v preglednici:

Posamezniki	Število prebranih knjig na leto
Študent	11
Zaposleni	5
Študent	9
Študent	12
Študent	8
Zaposleni	4
Študent	14
Zaposleni	6
Zaposleni	3
Zaposleni	2
Študent	7
Študent	10
Zaposleni	1
Zaposleni	7

- a) Podatke o številu prebranih knjig v enem letu za 14 oseb vnesite v program SPSS. Izračunajte in pojasnite rezultate opisne statistike za število knjig, ki jih posamezniki preberejo v enem letu.
- b) Na osnovi rezultatov, ki ste jih dobili v točki a), izračunajte koeficient variabilnosti.
- c) S pomočjo programa SPSS pojasnite:
- prvi in deveti decil,
  - prvi in tretji kvartil.

Potek: kliknemo *Analyze*, nato *Descriptive Statistics* in *Frequencies*. V okence *Dependent List* prenesemo spremenljivko in kliknemo *Statistics* ter v okencu *Percentile Values* obkljukamo *Percentiles* ter vpišemo vrednosti, ki nas zanimajo, na primer 10 (za prvi decil), 90 (za deveti decil), 25 (za prvi kvartil) in 75 (za tretji kvartil). Kliknemo *Continue* in nato še *OK*.

- d) Izračunajte in pojasnite decilni in kvartilni razmik s podatki, ki ste jih dobili v točki c).

### Odgovori in izpisi rezultatov:

a)

Tabela 3: Opisna statistika za spremenljivko *število prebranih knjig na leto za 14 oseb*

N	Valid	14
	Missing	0
	Mean	7,07
	Std. Error of Mean	1,035
	Median	7,00
	Mode	7
	Std. Deviation	3,872
	Variance	14,995
	Skewness	,132
	Std. Error of Skewness	,597
	Kurtosis	-,779
	Std. Error of Kurtosis	1,154
	Range	13
	Minimum	1
	Maximum	14

- b) Koeficient variabilnosti v odstotku:

$$KV\% = \frac{\sigma}{\bar{y}} \times 100$$

$$KV\% = \frac{3,872}{7,07} \cdot 100 = 54,767 \%$$

Koeficient variabilnosti znaša 54,767 %, kar pomeni, da standardni odklon pri tej spremenljivki predstavlja 54,767 % aritmetične sredine.

c)

Tabela 4: Izračun kvartilov in decilov

N	Valid	14
	Missing	0
Percentiles	10	1,50
	25	3,75
	50	7,00
	75	10,25
	90	13,00

Prvi decil ( $D_1 = 10\%$ ) znaša 1,50 knjig, kar pomeni, da je 10 % oseb prebralo na leto 1,50 knjig ali manj.

Deveti decil ( $D_9 = 90\%$ ) znaša 13 knjig, kar pomeni, da je 90 % oseb prebralo na leto 13 knjig ali manj.

Prvi kvartil ( $Q_1 = 25\%$ ) znaša 3,75 knjig, kar pomeni da je 25 % oseb prebralo na leto 3,75 knjig ali manj.

Tretji kvartil ( $Q_3 = 75\%$ ) znaša 10,25 knjig, kar pomeni da je 75 % oseb prebralo na leto 10,25 knjig ali manj.

d)

Kvartilni razmik:  $Q = Q_3 - Q_1 = 10,25 - 3,75 = 6,5$  knjig

Odgovor: 50 % oseb, ki glede na število prebranih knjig na leto ležijo na sredini ranžirne vrste, se razlikuje za največ 6,5 knjig. Ranžirno vrsto predstavljajo vrednosti spremenljivke, urejene po velikosti od najmanjše do največje vrednosti.

Decilni razmik:  $D = D_9 - D_1 = 13 - 1,50 = 11,5$  knjig

Odgovor: 80 % oseb, ki glede na število prebranih knjig na leto ležijo na sredini ranžirne vrste, se razlikuje za največ 11,5 knjig.



### Naloga 3

Naloga obravnava količino recikliranega materiala, zbranega od desetih gospodinjstev v enem tednu. Količina recikliranega materiala (v kg), zbranega od desetih gospodinjstev v enem tednu, je prikazana v preglednici:

Gospodinjstvo	1	2	3	4	5	6	7	8	9	10
Količina v kg	5,2	4,6	6,1	7,4	3,3	2,8	4,9	8,0	5,5	3,7

- Izračunajte 95-odstotni interval zaupanja za povprečno količino recikliranega materiala, zbranega od gospodinjstev v enem tednu.
- Izračunajte 99-odstotni interval zaupanja za povprečno količino recikliranega materiala, zbranega od gospodinjstev v enem tednu.

Potek: kliknemo *Analyze*, nato *Descriptive Statistics* in *Explore*. V okence *Dependent List* prenesemo spremenljivko *Količina recikliranega materiala* in kliknemo na okvirček *Statistics* ter obkljukamo *Descriptives* – pri *Confidence Interval for Mean* napišemo 95 %, ker želimo pojasniti 95-odstotni interval zaupanja za spremenljivko *Količina recikliranega materiala* (v primeru, če želimo pojasniti 99-odstotni interval zaupanja za aritmetično sredino spremenljivke, napišemo 99 %).

### Odgovori in izpisi rezultatov:

a)

Tabela 5: 95-odstotni interval zaupanja za povprečno količino recikliranega materiala, zbranega od gospodinjstev v enem tednu.

		Statistic	Std. Error	
Količina recikliranega materiala v kg	Mean	5,150	,5336	
	95 % Confidence Interval for Mean	Lower Bound	3,943	
		Upper Bound	6,357	
	5 % Trimmed Mean	5,122		
	Median	5,050		
	Variance	2,847		
	Std. Deviation	1,6874		
	Minimum	2,8		
	Maximum	8,0		
	Range	5,2		
	Interquartile Range	2,8		
	Skewness	,373	,687	
	Kurtosis	-,593	1,334	



V tabeli 5 vidimo, da spodnja meja (angl. Lower Bound) intervala zaupanja znaša 3,943 kg in zgornja meja (angl. Upper Bound) znaša 6,357 kg, kar pomeni, da s 95-odstotno verjetnostjo ocenjujemo, da je povprečna količina recikliranega materiala, zbranega od gospodinjestev v enem tednu, med 3,943 kg in 6,357 kg.



## Naloga 4

V preglednici imamo podatke o povprečnem času, ki ga posamezniki ( $n=10$ ) različnih spolov (1 – moški spol, 2 – ženski spol) preživijo na socialnih omrežjih na dan.

Spol	Povprečen čas preživet na socialnih omrežjih na dan (ure)
1	3,5
2	3,8
1	4,5
2	4,8
1	3
2	3,7
1	2,9
2	4
1	2,2
2	3,6

- Podatke za deset oseb o povprečnem času preživetem na socialnih omrežjih na dan vnesite v program SPSS.
- Naredite ločeno analizo za moški in ženski spol za povprečen čas, preživet na socialnih omrežjih na dan. Ugotovite kdo v povprečju preživi več časa na socialnih omrežjih na dan?
- Izračunajte delež standardnega odklona v aritmetični sredini.
- Izračunajte in pojasnite 80-odstotni interval zaupanja za povprečen čas, preživet na socialnih omrežjih na dan.
- Kolikšen je povprečen čas, preživet na socialnih omrežjih na dan, za 50 % oseb, ki so imeli najmanj ur preživetega časa.
- Pojasnite mero variabilnosti, ki upošteva variabilnost za 50 % oseb, ki glede na povprečen čas, preživet na socialnih omrežjih na dan, ležijo na sredini ranžirne vrste.
- Pojasnite mero variabilnosti, ki upošteva variabilnost za 80 % oseb, ki glede na povprečen čas, preživet na socialnih omrežjih na dan, ležijo na sredini ranžirne vrste.
- Kolikšen je povprečen čas preživet na socialnih omrežjih na dan tistih oseb, ki spadajo med 30 % oseb z najmanj preživetega časa na socialnih omrežjih?
- Kolikšen je povprečen čas, preživet na socialnih omrežjih na dan, tistih oseb, ki spadajo med 40 % oseb z največ preživetega časa na socialnih omrežjih?



## 3.2 Normalna porazdelitev

### 3.2.1 Preverba domneve o normalni porazdelitvi obravnavane spremenljivke

Preverba domneve o normalni porazdelitvi obravnavane spremenljivke je predpogoj za številne statistične teste, saj so normalno porazdeljeni podatki osnovna predpostavka pri uporabi parametričnih statističnih testov. Kolmogorov-Smirnov test in Shapiro-Wilk  $W$  test nam omogočita, da analiziramo, ali je podatkom obravnavane spremenljivke dopustno prilagoditi normalno porazdelitev (Campbell idr., 2007; Mishra idr., 2019).

### 3.2.2 Kolmogorov-Smirnov test in Shapiro-Wilk $W$ test

Kolmogorov-Smirnov test in Shapiro-Wilk  $W$  test uporabimo, kadar želimo preveriti trditev, da je obravnavana spremenljivka v statistični množici porazdeljena po normalni porazdelitvi.

Pri normalni porazdelitvi preizkušamo dve domnevi:

$H_0$ : Obravnavani spremenljivki je dopustno prilagoditi normalno porazdelitev.

$H_1$ : Obravnavani spremenljivki ni dopustno prilagoditi normalne porazdelitve.

V primeru, ko je test neznačilen (stopnja značilnosti preizkusa ali stopnja tveganja,  $p > 0,05$ ), tedaj ničelne domneve  $H_0$  ne zavrnemo in lahko sklepamo, da se porazdelitev proučevane spremenljivke v statistični množici ne razlikuje od normalne porazdelitve. Proučevana spremenljivka ni normalno porazdeljena, ko je test statistično značilen ( $p < 0,05$ ) in v tem primeru sprejememo domnevo  $H_1$ .



## Naloga 1

Raziskovalci so želeli preučiti, ali uporaba umetne inteligence pomaga pri boljši kakovosti odločitev pri zaposlovanju in selekciji kandidatov. V vzorec so naključno izbrali 380 velikih podjetij. Lastniki podjetij so na 5-stopenjski Likertovi lestvici izrazili stopnjo strinjanja (kjer pomeni 1 – sploh se ne strinjam in 5 – popolnoma se strinjam) z navedeno trditvijo:

AR2: UI pomaga pri boljši kakovosti odločitev pri zaposlovanju in selekciji kandidatov.

Podatki so v datoteki *Normalna porazdelitev\_umetna inteligenca.sav*.

- Zapišite ničelno domnevo o normalni porazdelitvi obravnavane spremenljivke.
- Preverite, ali je spremenljivka porazdeljena po normalni porazdelitvi in narišite frekvenčni histogram za spremenljivko.

Potek: kliknemo na *Analyze* in nato *Descriptive Statistics* ter *Explore*. V okence *Dependent List* prenesemo vse štiri spremenljivke – komponente motiviranosti zaposlenih. Kliknemo gumb *Plots* in izberemo *Normality plots with tests* ter kliknemo *Continue*.

Izpis rezultata:

Odgovori in izpisi rezultatov:

a)

H0: Obravnavani spremenljivki AR2 je dopustno prilagoditi normalno porazdelitev.

H1: Obravnavani spremenljivki AR2 ni dopustno prilagoditi normalne porazdelitve.

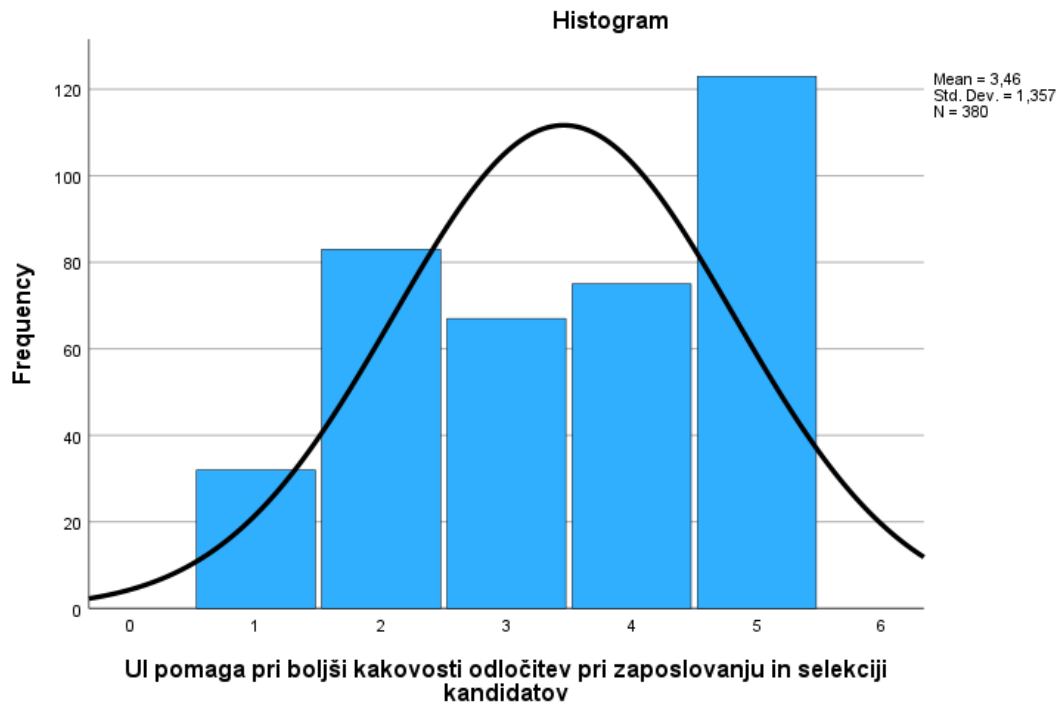
b)

Tabela 6: Kolmogorov-Smirnov test in Shapiro-Wilk W test za spremenljivko AR2

Tests of Normality						
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
AR2: UI pomaga pri boljši kakovosti odločitev pri zaposlovanju in selekciji kandidatov.	,196	380	<,001	,865	380	<,001

a. Lilliefors Significance Correction

Kolmogorov-Smirnov test in Shapiro-Wilk W test (tabela 6) kažeta, da obravnavana spremenljivka ni porazdeljena po normalni porazdelitvi, saj je stopnja tveganja manjša od 0,001 ( $p < 0,05$ ), kar pomeni, da ničelno domnevo zavrnamo in potrdimo raziskovalno domnevo, da spremenljivka ni normalno porazdeljena. Odstopanja od oblike normalne porazdelitve so opazna na sliki frekvenčnega histograma.



**Histogram 2:** Frekvenčni histogram za spremenljivko UI pomaga pri boljši kakovosti odločitev pri zaposlovanju in selekciji kandidatov



## Naloga 2

Želimo ugotoviti, ali se spremenljivka *tehnologija UI zmanjša čas, porabljen za tečaje usposabljanja v podjetju (T1)* in spremenljivka *starost* porazdelujeta po normalni porazdelitvi.

Podatki so v datoteki *Normalna porazdelitev\_umetna inteligenca.sav*.

- Preverite, ali lahko sklepamo, da je spremenljivki *T1* in spremenljivki *starost* dopustno prilagoditi normalno porazdelitev.
- Na osnovi mer centralne tendence, koeficienta asimetrije in koeficienta sploščenosti pojasnite, ali je porazdelitev obeh obravnavanih spremenljivk (empirični podatki) po svojih značilnostih podobna normalni porazdelitvi.

---



---



---



---



## Naloga 3

Regionalna banka je nedavno začela ponujati novo storitev spletne banke in želi oceniti njeno sprejetost med svojimi strankami. Zbrali so podatke o številu novih registracij za spletno banko v prvem tednu po uvedbi, zbranih iz naključnega vzorca 25 poslovalnic po državi. Preverite, ali je spremenljivka v statistični množici porazdeljena po normalni porazdelitvi. Prav tako zapišite ničelno in raziskovalno domnevo.

V preglednici (tabela 7) je izpis rezultata:

**Tabela 7: Kolmogorov-Smirnov test in Shapiro-Wilk W test za spremenljivko število novih registracij za spletno banko**

Tests of Normality						
	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Število novih registracij za spletno banko	0,136	30	0,215	0,958	30	0,374

---



---



---



---



---



---



---



---



---



---



---

### 3.3 Univariatni statistični testi

Parametrični test za ugotavljanje statistično značilnih razlik med povprečnimi vrednostmi spremenljivk v vzorcih uporabljamo, kadar je podatkom za odvisno spremenljivko na intervalni ali razmernostni lestvici dopustno prilagoditi normalno porazdelitev (Corder in Foreman, 2014).

Kadar enote vzorcev pripadajo isti statistični množici, govorimo o odvisnih vzorcih. Kadar enote vzorca pripadajo različnim statističnim množicam, govorimo o neodvisnih vzorcih.

Univariatni parametrični statistični testi, ki jih bomo spoznali, so:

- t-test za testiranje razlik med dvema povprečnima vrednostima za dva neodvisna vzorca (angl. independent samples t-test),
- t-test za testiranje razlik med dvema povprečnima vrednostima za dva odvisna vzorca (angl. paired samples t-test),
- ANOVA (enostranska analiza variance ANOVA) za testiranje razlik med povprečnimi vrednostmi več, med seboj neodvisnih, vzorcev (angl. one-way ANOVA).

Pri preizkušanje domnev med dvema odvisnima ali neodvisnima vzorcema preverjamo naslednjo ničelno hipotezo:

*Ničelna domneva*  $H_0$ :  $H_0: \bar{y}_1 = \bar{y}_2$

*Raziskovalna domneva*  $H_1$ :  $H_1: \bar{y}_1 \neq \bar{y}_2$

$\bar{y}_1$  pomeni vzorčna aritmetična sredina spremenljivke v prvem vzorcu.

$\bar{y}_2$  pomeni vzorčna aritmetična sredina spremenljivke v drugem vzorcu.

Pri enostranski nalizi variance (ANOVA) preizkušamo domnevo o enakosti povprečne vrednosti spremenljivke v več kot dveh neodvisnih vzorcih:

$H_0: \bar{y}_1 = \bar{y}_2 = \bar{y}_3 = \dots = \bar{y}_k$

$H_1$ : vsaj dve povprečji se statistično značilno razlikujeta.

Neparametrični test pa uporabimo, kadar za odvisno numerično spremenljivko podatkom, ki sicer temeljijo na intervalni ali razmernostni lestvici, ni dopustno prilagoditi normalne porazdelitve, ali kadar podatki za odvisno spremenljivko temeljijo na ordinalni lestvici (Corder in Foreman, 2014).

V nadaljevanju bomo uporabili tudi  $\chi^2$ -test za analizo povezanosti dveh spremenljivk, pri čemer je vsaj ena opisna spremenljivka.

### 3.3.1 Parametrični test za odvisna vzorca: t-test za odvisna vzorca



#### Naloga 1

Ekonomsko-poslovna fakulteta UM je razvila nov spletni izobraževalni program za izboljšanje znanja študentov v statistiki. Da bi ocenili učinkovitost programa, so 30 študentov testirali na začetku programa in 30 študentov po končanem izobraževanju (ocena je na lestvici od 0 do 100 točk). Zanima jih, ali obstajajo statistično značilne razlike v znanju študentov pred in po zaključku izobraževalnega programa.

Podatki so v datoteki: *t-test za dva odvisna vzorca\_izobraževalni program.sav*

V ta namen bomo preverili naslednji domnevi:

H0: Povprečna ocena študentov pred izobraževalnim programom je enaka povprečni oceni študentov po izobraževalnem programu ( $H0: \bar{y}_1 = \bar{y}_2$ ).

H1: Povprečna ocena študentov pred izobraževalnim programom se statistično značilno razlikuje od povprečne ocene študentov po izobraževalnem programu ( $H1: \bar{y}_1 \neq \bar{y}_2$ ).

Postopek:

S Kolmogorov-Smirnovim in Shapiro-Wilkovim testom smo najprej ugotovili, da je podatkom dopustno prilagoditi normalno porazdelitev (tabela 8). Zato za preverjanje zapisanih domnev uporabimo parametrični test za odvisna vzorca, tj. t-test za odvisna vzorca.



V izboru *Analyze* izberemo *Compare Means*, nato pa *Paired-Samples T Test*. Označimo spremenljivko *pred\_test* in jo s klikom na gumb s puščico prenesemo v okvir *Paired Variables*, in sicer v celico pod *Variable 1*, spremenljivko *po\_test* pa prenesemo v celico pod *Variable 2*. Kliknemo *OK*, da izvedemo t-test za dva odvisna vzorca.

Odgovori in izpisi rezultatov:

Tabela 8: Kolmogorov-Smirnov test in Shapiro-Wilk W test za oceno študentov pred in po izobraževalnem programom

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Pred_test	,090	30	,200*	,975	30	,687
Po_test	,083	30	,200*	,965	30	,422

\*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Iz tabele 8 je razvidno, da je pri obeh spremenljivkah  $p > 0,05$ , zato ničelne domneve Kolmogorov-Smirnovega testa in Shapiro-Wilkovega testa ne zavrnemo. Spremenljivki *ocena študentov pred izobraževalnim programom* je dopustno prilagoditi normalno porazdelitev. Normalno porazdelitev je dopustno prilagoditi tudi spremenljivki *ocena študentov po izobraževalnem programu*.

Tabela 9: Povprečna ocena študentov pred in po izobraževalnem programom

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Pred_test	49,0600	30	4,50008	,82160
	Po_test	53,4537	30	6,78767	1,23925

Iz tabele 9 je razvidno, da je povprečna ocena študentov pred izobraževalnim programom 49,06 točk in povprečna ocena študentov po izobraževalnem programom 53,45 točk.

Tabela 10: Rezultati t-testa za odvisna vzorca

		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Two-Sided p
					Lower	Upper			
Pair 1	Pred_test Po_test	-4,39367	4,65622	,85011	-6,13233	-2,65501	-5,168	29	<,001

Iz rezultatov v zadnjem stolpcu v tabeli 10 je razvidno, da je izračunana dvostranska stopnja tveganja (angl. Two-Sided p)  $p < 0,001$ . Ker je  $p < 0,05$ , ničelno domnevo  $H_0: \bar{y}_1 = \bar{y}_2$  zavrnamo. Povprečna ocena študentov pred izobraževalnem programu se statistično značilno razlikuje od povprečne ocene študentov po izobraževalnem programu, kar pomeni, da sprejmemo  $H_1: \bar{y}_1 \neq \bar{y}_2$ .



## Naloga 2

Podjetje X je izvedlo novo marketinško kampanjo za promocijo enega od svojih ključnih izdelkov. Da bi ocenili učinkovitost kampanje, so zbrali podatke o prodaji izdelka pred začetkom kampanje in po njenem zaključku. V podjetju želijo ugotoviti, ali obstaja statistično značilna razlika v prodaji izdelka pred in po marketinški kampanji.

Podatki so v datoteki: *t-test za dva odvisna vzorca\_marketinška kampanja.sav*

Podatki vključujejo prodajne številke (v d.e.) 35 izbranih dni pred začetkom kampanje (*Prodaja\_pred*) in 35 izbranih dni po zaključku kampanje (*Prodaja\_po*).

- Preverite, ali je podatkom za prodajo izdelka pred marketinško kampanjo in podatkom po marketinški kampanji dopustno prilagoditi normalno porazdelitev.
- Utemeljite uporabo t-testa za odvisna vzorca, če želimo preveriti, ali se povprečna prodaja izdelka pred marketinško kampanjo statistično značilno razlikuje od povprečne prodaje izdelka po marketinški kampanji.
- Zapišite ničelno in raziskovalno domnevo.
- Preverite, ali se povprečna prodaja izdelka pred marketinško kampanjo statistično značilno razlikuje od povprečne prodaje izdelka po marketinški kampanji.



## Naloga 3

Podjetje Y je uvedlo novo okolju prijazno iniciativo, ki vključuje zmanjšanje porabe električne energije v vseh svojih pisarnah. Iniciativa vključuje ukrepe, kot so izboljšana izolacija, uporaba energetske učinkovitejše razsvetljave in spodbujanje zaposlenih k manjši uporabi električnih naprav. Podatki vključujejo mesečno porabo električne energije (v kilovatnih urah) za 50 pisarn pred uvedbo iniciative in po njeni uvedbi. V podjetju želijo ugotoviti, ali obstajajo statistično značilne razlike v mesečni porabo električne energije (v kilovatnih urah) pred uvedbo iniciative in po njeni uvedbi.

Podatki so v datoteki: *t-test za dva odvisna vzorca\_poraba električne energije.sav*

- Preverite, ali je podatkom za porabo električne energije pred uvedbo iniciative in po njeni uvedbi dopustno prilagoditi normalno porazdelitev.
- Utemeljite uporabo t-testa za odvisna vzorca, če želimo preveriti, ali se povprečna poraba električne energije pred uvedbo iniciative in po njeni uvedbi statistično značilno razlikuje.
- Zapišite ničelno in raziskovalno domnevo.
- Preverite, ali se povprečna poraba električne energije pred uvedbo iniciative statistično značilno razlikuje od povprečne porabe električne energije po uvedbi iniciative.

### 3.3.2 Parametrični test za neodvisna vzorca: t-test za neodvisna vzorca



#### Naloga 1

Želimo ugotoviti, ali obstajajo statistično značilne razlike med malimi in velikimi podjetji v uspešnosti implementacije umetne inteligence v njihove poslovne procese. Uspešnost implementacije umetne inteligence je merjena s povečanjem učinkovitosti operacij, zmanjšanjem stroškov in izboljšanjem zadovoljstva strank, ocenjeno na lestvici od 0 do 100. V vzorec je naključno izbranih 120 majhnih in velikih podjetij v Sloveniji.

Podatki so v datoteki: *t-test za dva neodvisna vzorca in ANOVA\_implementacija umetne inteligence.sav*

V ta namen bomo preverili naslednji domnevi:

$H_0$ : Povprečna uspešnost implementacije umetne inteligence v malih podjetjih je enaka povprečni uspešnosti implementacije umetne inteligence v velikih podjetjih ( $H_0: \bar{y}_1 = \bar{y}_2$ ).

$H_1$ : Povprečna uspešnost implementacije umetne inteligence v malih podjetjih se statistično značilno razlikuje od povprečne uspešnosti implementacije umetne inteligence v velikih podjetjih ( $H_1: \bar{y}_1 \neq \bar{y}_2$ ).

Potek: s Kolmogorov-Smirnovim in Shapiro-Wilkovim testom smo najprej ugotovili, da je podatkom dopustno prilagoditi normalno porazdelitev (tabela 11). Zato za preverjanje zapisanih domnev uporabimo parametrični test za neodvisna vzorca, tj. t-test za neodvisna vzorca.

V izboru *Analyze* izberemo *Compare Means*, nato pa *Independent-Samples T Test*. Odvisno spremenljivko *Implementacija* prenesemo v desno okno *Test Variable(s)*. Nato spremenljivko *Podjetje* prenesemo v okno *Grouping Variable*. Kliknemo na *Define Groups* in definiramo vrednosti skupin: pri *Group 1* vpišemo 1 (tj. malo podjetje), pri *Group 2* pa vpišemo 2 (tj. veliko podjetje) in kliknemo *Continue*. Kliknemo *OK*, da izvedemo t-test za dva neodvisna vzorca.

Odgovori in izpisi rezultatov:

**Tabela 11: Kolmogorov-Smirnov test in Shapiro-Wilk W test za uspešnost implementacije umetne inteligence**

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Implementacija	,060	120	,200*	,983	120	,132

\*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

**Tabela 12: Povprečna uspešnost implementacije umetne inteligence v malih in velikih podjetjih**

	Podjetje	N	Mean	Std. Deviation	Std. Error Mean
Implementacija	Malo podjetje	58	49,4697	8,72888	1,14616
	Veliko podjetje	62	65,8485	15,31077	1,94447

**Tabela 13: Rezultati t-testa za neodvisna vzorca (naloga 1)**

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Significance		Mean Difference	Std. Error Difference	95 % Confidence Interval of the Difference	
						One-Sided p	Two-Sided p			Lower	Upper
Implementacija	Equal variances assumed	7,409	,007	-7,133	118	<,001	<,001	-16,37889	2,29611	-20,92582	-11,83196
	Equal variances not assumed			-7,257	98,081	<,001	<,001	-16,37889	2,25713	-20,85805	-11,89974

V tabeli 13 z Levenovim testom enakosti varianc najprej preverimo, ali lahko predpostavimo enake variance v obeh neodvisnih vzorcih:

H0: Predpostavimo enake variance.

H1: Enakih varianc ne predpostavimo.

Iz izpisa rezultatov Levenovega testa enakosti varianc vidimo, da je  $p = 0,007$ , torej  $p < 0,05$ , zato ničelno domnevo zavrnamo in sprejmemo  $H_1$ : Enakih varianc ne predpostavimo. V izpisu rezultatov t-testa za primerjavo povprečnih vrednosti neodvisnih vzorcev zato upoštevamo drugo vrstico izpisa. Odčitamo, da je  $p < 0,001$ , torej ničelno domnevo o enakosti povprečnih vrednosti zavrnamo: Povprečna uspešnost implementacije umetne inteligence v malih podjetjih je enaka povprečni uspešnosti implementacije umetne inteligence v velikih podjetjih ( $H_0: \bar{y}_1 = \bar{y}_2$ ) in sprejmemo raziskovalno domnevo: Povprečna uspešnost implementacije umetne inteligence v malih podjetjih se statistično značilno razlikuje od povprečne uspešnosti implementacije umetne inteligence v velikih podjetjih ( $H_1: \bar{y}_1 \neq \bar{y}_2$ ).



## Naloga 2

Za 100 naključno izbranih podjetij v Sloveniji želimo preučiti, ali obstajajo statistično značilne razlike v zadovoljstvu zaposlenih med podjetji iz proizvodnega in storitvenega sektorja. Zadovoljstvo zaposlenih je ocenjeno na lestvici od 0 do 100.

Podatki so v datoteki: *t-test za dva neodvisna vzorca\_naloga2*

V ta namen bomo preverili naslednji domnevi:

$H_0$ : Povprečno zadovoljstvo zaposlenih v proizvodnem sektorju in povprečno zadovoljstvo zaposlenih v storitvenem sektorju se statistično značilno ne razlikujeta ( $H_0: \bar{y}_1 = \bar{y}_2$ ).

$H_1$ : Povprečno zadovoljstvo zaposlenih v proizvodnem sektorju je statistično značilno višje kot povprečno zadovoljstvo zaposlenih v storitvenem sektorju ( $H_1: \bar{y}_1 > \bar{y}_2$ ).

Odgovori in izpisi rezultatov:

**Tabela 14: Kolmogorov-Smirnov test in Shapiro-Wilk W test za zadovoljstvo zaposlenih**

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Zadovoljstvo	,053	100	,200*	,976	100	,063

\*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Tabela 15: Povprečno zadovoljstvo zaposlenih v podjetjih proizvodnega in storitvenega sektorja

	Podjetje	N	Mean	Std. Deviation	Std. Error Mean
Zadovoljstvo	Proizvodni sektor	50	74,50	14,577	2,062
	Storitveni sektor	50	64,50	14,577	2,062

Tabela 16: Rezultati t-testa za neodvisna vzorca (naloga 2)

		Levene's Test for Equality of Variances		t-test for Equality of Means							
Zadovoljstvo		F	Sig.	t	df	Significance		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						One-Sided p	Two-Sided p			Lower	Upper
			Equal variances assumed	,000	1,000	3,430	98	<,001	<,001	10,000	2,915
	Equal variances not assumed			3,430	98,000	<,001	<,001	10,000	2,915	4,214	15,786

V tabeli 16 z Levenovim testom enakosti varianc najprej preverimo, ali lahko predpostavimo enake variance v obeh neodvisnih vzorcih:

H0: Predpostavimo enake variance.

H1: Enakih varianc ne predpostavimo.

Iz izpisa rezultatov Levenovega testa enakosti varianc vidimo, da je  $p = 1,000$ , torej  $p > 0,05$ , zato sprejmemo ničelno domnevo. V izpisu rezultatov t-testa za primerjavo povprečnih vrednosti neodvisnih vzorcev zato upoštevamo prvo vrstico izpisa. Odčitamo, da je  $p < 0,001$ , torej ničelno domnevo o enakosti povprečnih vrednosti zavrnilo in sprejmemo raziskovalno domnevo H1: Povprečno zadovoljstvo zaposlenih v proizvodnem sektorju je statistično značilno višje kot povprečno zadovoljstvo zaposlenih v storitvenem sektorju ( $H1: \bar{y}_1 > \bar{y}_2$ ).



### Naloga 3

Za 120 naključno izbranih majhnih in velikih podjetij v Sloveniji želimo preučiti, ali obstajajo statistično značilne razlike med malimi in velikimi podjetji v zadovoljstvu strank s podjetjem po implementaciji umetne inteligence, na lestvici od 0 do 100.

Podatki so v datoteki: *t-test za dva neodvisna vzorca in ANOVA\_implementacija umetne inteligence.sav*

- a) Ugotovite, ali je podatkom o zadovoljstvu strank dopustno prirediti normalno porazdelitev.
- b) Utemeljite uporabo t-testa za neodvisna vzorca, če želimo preveriti, ali se povprečno zadovoljstvo strank v malih podjetjih statistično značilno razlikuje od povprečnega zadovoljstva strank v velikih podjetjih. Zapišite ničelno in raziskovalno domnevo.
- c) Ugotovite, ali se povprečno zadovoljstvo strank v malih podjetjih statistično značilno razlikuje od povprečnega zadovoljstva strank v velikih podjetjih.



## Naloga 4

Cilj raziskave je ugotoviti, ali obstajajo statistično značilne razlike v uspešnosti uporabe umetne inteligence med podjetji, ki jih vodijo moški, in podjetji, ki jih vodijo ženske. Uspešnost uporabe umetne inteligence smo merili s povečanjem produktivnosti po implementaciji sistemov umetne inteligence, izraženo v odstotkih. Vsako podjetje je ocenilo povečanje svoje produktivnosti po implementaciji umetne inteligence, na lestvici od 0 do 100%. Vzorec zajema 120 naključno izbranih majhnih in velikih podjetij v Sloveniji.

Podatki so v datoteki: *t-test za dva neodvisna vzorca in ANOVA\_implementation\_umetne\_intelligence.sav*

- a) Ugotovite, ali je podatkom o produktivnosti po implementaciji umetne inteligence dopustno prirediti normalno porazdelitev.
- b) Ugotovite, ali se povprečna produktivnosti po implementaciji sistemov umetne inteligence statistično značilno razlikuje med podjetji, ki jih vodijo moški, in tistimi, ki jih vodijo ženske. Zapišite ničelno in raziskovalno domnevo.

### 3.3.3 Parametrični test za več kot 2 neodvisna vzorca: ANOVA



## Naloga 1

Cilj raziskave je ugotoviti, ali obstajajo statistično značilne razlike v uspešnosti implementacije umetne inteligence (UI) med podjetji, ki jih vodijo mladi, srednje stari in starejši lastniki podjetij. Lastniki podjetij so razvrščeni v tri skupine: mladi (do 35 let), srednje starosti (36 do 55 let) in starejši (nad 55 let). Uspešnost implementacije UI je merjena z izboljšanjem poslovnih procesov, ocenjenim na lestvici od 0 do 100. Vzorec

zajema 120 naključno izbranih majhnih in velikih podjetij v Sloveniji. Podatki so v datoteki: *t-test za dva neodvisna vzorca in ANOVA\_implementacija umetne inteligence.sav*

V ta namen smo oblikovali domnevi:

H0: V povprečni uspešnosti implementacije umetne inteligence ni statistično značilnih razlik med podjetji, ki jih vodijo mladi, srednje stari in starejši lastniki podjetij ( $H_0: \bar{y}_1 = \bar{y}_2 = \bar{y}_3$ ).

H1: Vsaj dve skupini lastnikov podjetij se statistično značilno razlikujeta v povprečni uspešnosti implementacije umetne inteligence.

Postopek:

Najprej smo preverili, da je spremenljivki *uspešnost implementacije umetne inteligence* dopustno prilagoditi normalno porazdelitev.

Nato za izvedbo ANOVA v izboru *Analyze* izberemo *Compare Means, One-way ANOVA* in odpre se okno *One-way ANOVA*. Spremenljivko *uspešnost imeplementacije* prenesemo v okno pri *Dependent List*. Spremenljivko *starost* prenesemo v okno pri *Factor*. Kliknemo na *Options*, nato na *Descriptive* in potem *Continue*. Po kliku na *OK* dobimo izpis rezultatov.

Odgovori in izpisi rezultatov:

Tabela 17: Rezultati analize variance

ANOVA					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7336,333	2	3668,166	22,185	<,001
Within Groups	19345,398	117	165,345		
Total	26681,731	119			

Iz izpisa rezultatov analize variance v tabeli 17 je razvidno, da je stopnja tveganja  $p < 0,05$ , zato ničelno domnevo zavrnamo in sprejmemo raziskovalno domnevo H1: Vsaj dve skupini lastnikov podjetij se statistično značilno razlikujeta v povprečni uspešnosti implementacije umetne inteligence.

Če želimo ugotoviti, katere skupine se med seboj značilno razlikujejo, moramo opraviti dodatno analizo, ki se imenuje post hoc. Z njo primerjamo povprečja med tremi skupinami (povprečje vsake skupine primerjamo s povprečjem preostalih dveh skupin). To analizo



izvedemo tako, da kliknemo *Post hoc* gumb v pogovornem oknu *One-way ANOVA*. Med vsemi možnostmi se odločimo za Tukeyevo metodo tako, da izberemo *Tukey* in dobimo izpis v tabeli 18.

Tabela 18: Rezultati post-hoc testa pri analizi variance

Multiple Comparisons						
Tukey HSD						
(I) Starost	(J) Starost	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
do 35 let	od 36 do 55 let	-7,05352	3,27366	,083	-14,8249	,7179
	nad 55 let	-17,83868*	2,73747	< ,001	-24,3372	-11,3402
od 36 do 55 let	do 35 let	7,05352	3,27366	,083	-,7179	14,8249
	nad 55 let	-10,78517*	3,00411	,001	-17,9167	-3,6537
nad 55 let	do 35 let	17,83868*	2,73747	< ,001	11,3402	24,3372
	od 36 do 55 let	10,78517*	3,00411	,001	3,6537	17,9167

\*. The mean difference is significant at the 0.05 level.

Iz tabele 18 je razvidno, da se povprečna uspešnost implementacije umetne inteligence razlikuje po starostnih skupinah do 35 let in nad 55 let ter starostnih skupinah od 36 do 55 let in nad 55 let ( $p < 0,05$ ). Med starostnima skupinama do 35 let in od 36 do 55 let ni statistično značilnih razlik, ker je stopnja tveganja 0,083 ( $p > 0,05$ ).



## Naloga 2

V nadaljevanju še nas zanima, ali obstajajo statistično značilne razlike v uspešnosti implementacije umetne inteligence med podjetji v treh različnih državah: država A, država B in država C. Uspešnost implementacije umetne inteligence je merjena s povečanjem učinkovitosti operacij, zmanjšanjem stroškov in izboljšanjem zadovoljstva strank, ocenjeno na lestvici od 0 do 100.

Podatki so v datoteki: *t-test za dva neodvisna vzorca in ANOVA\_implementation\_umetne\_intelligence.sav*

Zapišite ustrezni domnevi za opisani problem in preverite domnevi. V primeru, če na osnovi stopnje značilnosti ugotovite, da med državami obstajajo statistično značilne razlike, pojasnite, med katerimi skupinami prihaja do statistično značilnih razlik.

### 3.3.4 Neparometrični test: $\chi^2$ -test za analizo povezanosti dveh nominalnih spremenljivk



#### Naloga 1

Ugotoviti želimo, ali sta spol (1 – moški, 2 – ženski) in izbira plačilne metode pri nakupovanju (1 – gotovina, 2 – kreditna kartica, 3 – spletni plačilni sistem) povezani. Naključno izbranih 300 potrošnikov je v vprašalniku poleg spola označilo, kako običajno plačujejo za svoje nakupe.

Podatki so v datoteki: *Hi-kvadrat\_naloga1.sav*.

Zapisali smo domnevi:

H0: Spremenljivki *spol* in *izbira plačilne metode* pri nakupovanju nista povezani.

H1: Spremenljivki *spol* in *izbira plačilne metode* pri nakupovanju sta povezani.

Postopek:

Izberemo *Analyze, Descriptive Statistics, Crosstabs*. Iz okna *Variables* prenesemo spremenljivko *Spol* v *Rows*, spremenljivko *Izbira plačilne metode pri nakupovanju* pa v *Columns* v okno na desni strani. Pri *Statistics* izberemo *Chi-square* in nadaljujemo s *Continue*.

Odgovori in izpisi rezultatov:

**Tabela 19: Rezultati  $\chi^2$ -testa za preverjanje povezanosti spremenljivk *spol* in *izbira plačilne metode pri nakupovanju***

#### Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	3,058 <sup>a</sup>	2	,217
Likelihood Ratio	3,066	2	,216
Linear-by-Linear Association	,219	1	,640
N of Valid Cases	300		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 35,00.

Iz tabele 19 je v vrstici Pearson Chi-Square razvidno, da je  $p = 0,217$  ( $p > 0,05$ ), zato sprejmemo ničelno domnevo H0: Spremenljivki *spol* in *izbira plačilne metode* pri nakupovanju nista povezani.



## Naloga 2

Ugotoviti želimo, ali sta področje zaposlitve (1 – IT, 2 – zdravstvo, 3 – izobraževanje, 4 – umetnost) in preferenca delovnega okolja (1 – pisarna, 2 – delo od doma, 3 – terensko delo) povezani. Naključno izbranih 434 zaposlenih je v vprašalniku poleg svojega področja zaposlitve označilo tudi svojo preferenco delovnega okolja.

Podatki so v datoteki: *Hi-kvadrat\_naloga2.sav*.

Tabela 20: Rezultati  $\chi^2$ -testa za preverjanje povezanosti spremenljivk *področje zaposlitve* in *preferenca delovnega okolja*

### Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	94,806 <sup>a</sup>	6	<,001
Likelihood Ratio	82,893	6	<,001
Linear-by-Linear Association	4,509	1	,034
N of Valid Cases	434		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 14,63. Iz tabele 20 je v vrstici Pearson Chi-Square razvidno, da je  $p < 0,05$ , zato sprejmemo raziskovalno domnevo H1: Spremenljivki *področje zaposlitve* in *preferenca delovnega okolja* sta povezani.

### 3.4 Faktorska analiza

Faktorska analiza je multivariatna metoda, s pomočjo katere je mogoče iz večjega števila medsebojno povezanih merjenih spremenljivk oblikovati manjše število med seboj neodvisnih faktorjev (Tabachnick in Fidell, 2013). Ta pristop je zelo uporaben na področju ekonomskih in poslovnih ved, kjer imamo velikokrat opravka z večdimenzionalnimi spremenljivkami, ki se pogosto nanašajo na zaznave in mnenja anketiranih oseb (kupcev, potencialnih kupcev, podjetnikov in podobno).

Ustreznost faktorске analize preverjamo z Barlettovim testom sferičnosti, kjer preizkušamo ničelno domnevo, da je osnovna korelacijska matrika enaka matriki enote, kar pomeni, da ne bi obstajala povezanost med opazovanimi oziroma merjenimi

spremenljivkami in Keiser- Meyer-Olkinovo statistiko (KMO), ki mora biti večja od 0,5 (Fabrigar in Duane, 2012).

Komunalitete povedo, kolikšen delež variance  $i$ -te merjene spremenljivke je pojasnjen z dobljenimi faktorji. Komunaliteta mora biti večja od vrednosti 0,4 (Tabachnick in Fidell, 2013).

Glavni koraki factorske analize vključujejo izbiro metode za ekstrakcijo faktorjev, kot je na primer metoda glavnih komponent, in odločitev o številu faktorjev, ki v zadostni meri opišejo variabilnost merjenih spremenljivk in s katerimi merjene spremenljivke lahko nadomestimo. To lahko določimo s pomočjo lastnih vrednosti (angl. eigenvalues), scree testa ali z uporabo teoretičnih predpostavk o podatkovni strukturi.

Faktorji se običajno pojavljajo kot linearne kombinacije merjenih spremenljivk, kjer ima vsaka spremenljivka določeno factorsko obremenitev (angl. factor loading), ki označuje, koliko ta spremenljivka prispeva k določenemu faktorju. Pri izvajanju factorske analize je preverjanje lastne vrednosti ključnega pomena za odločanje o številu faktorjev, ki jih je smiselno uporabiti v nadaljni analizi, in velja, da mora biti lastna vrednost višja od 1. To pravilo izhaja iz predpostavke, da če ima faktor lastno vrednost večjo od 1, potem zajema več variance, kot bi jo zajemala ena spremenljivka sama po sebi. Faktor z majhno lastno vrednostjo pa ne prispeva bistveno k razlagi celotne variance v podatkih (Kaiser, 1960; Hayton idr., 2004).



## Naloga 1

V študiji, izvedeni na naključno izbranem vzorcu 260 podjetij, so raziskovalci želeli preučiti stopnjo, do katere so podjetja sprejela agilnost. Osredotočili so se na preučevanje nagnjenosti podjetij k razvijanju agilnih vodstvenih veščin. Za zbiranje podatkov so uporabili vprašalnik, ki vsebuje trditve, ocenjene na 5-stopenjski Likertovi lestvici, kjer pomeni 1 – sploh se ne strinjam in 5 – popolnoma se strinjam:

RAVV1: Vodstvo hitro in učinkovito odgovarja na spremembe v tržnem okolju.

RAVV2: Vodstvo podpira in spodbuja inovativnost na vseh ravneh podjetja.

RAVV3: V podjetju redno prepoznavamo in nagrajujemo prizadevanja za izboljšanje agilnosti.

RAVV4: Naša organizacijska struktura omogoča hitro prilagajanje novim izzivom.

RAVV5: Vodstvo ustvarja pogoje, ki spodbujajo delovno zavzetost zaposlenih.

Na osnovi podatkov v datoteki: *Faktorska analiza.sav* izvedite faktorsko analizo za večdimenzionalno spremenljivko *razvijanje agilnih vodstvenih veščin* in pojasnite rezultate.

Tabela 21: Korelacijska matrika

## Correlation Matrix

		Vodstvo hitro in učinkovito odgovarja na spremembe v tržnem okolju.	Vodstvo podpira in spodbuja inovativnost na vseh ravneh podjetja.	V podjetju redno prepoznavamo in nagrajujemo prizadevanja za izboljšanje agilnosti.	Naša organizacijska struktura omogoča hitro prilagajanje novim izzivom.	Vodstvo ustvarja pogoje, ki spodbujajo delovno zavzetost zaposlenih.
Correlation	Vodstvo hitro in učinkovito odgovarja na spremembe v tržnem okolju.	<b>1,000</b>	,937	,807	,818	,645
	Vodstvo podpira in spodbuja inovativnost na vseh ravneh podjetja.	,937	<b>1,000</b>	,791	,801	,645
	V podjetju redno prepoznavamo in nagrajujemo prizadevanja za izboljšanje agilnosti."	,807	,791	<b>1,000</b>	,854	,751
	Naša organizacijska struktura omogoča hitro prilagajanje novim izzivom.	,818	,801	,854	<b>1,000</b>	,655
	Vodstvo ustvarja pogoje, ki spodbujajo delovno zavzetost zaposlenih	,645	,645	,751	,655	<b>1,000</b>
Sig. (1-tailed)	Vodstvo hitro in učinkovito odgovarja na spremembe v tržnem okolju.		<,001	<,001	<,001	<,001
	Vodstvo podpira in spodbuja inovativnost na vseh ravneh podjetja.	,000		,000	,000	,000
	V podjetju redno prepoznavamo in nagrajujemo prizadevanja za izboljšanje agilnosti.	,000	,000		,000	,000
	Naša organizacijska struktura omogoča hitro prilagajanje novim izzivom.	,000	,000	,000		,000
	Vodstvo ustvarja pogoje, ki spodbujajo delovno zavzetost zaposlenih.	,000	,000	,000	,000	

Potek za izvedbo faktorske analize: kliknemo *Analyze*, izberemo *Dimension Reduction* ter nato *Factor*. V desno okence prenesemo tiste spremenljivke, za katere želimo izvesti faktorsko analizo. V našem primeru spremenljivke *RAVV1* do *RAVV5* prenesemo v desno okence. S klikom na gumb *Descriptives* se odpre novo pogovorno okno, v katerem pri *Statistics* izberemo *Univariate descriptives in Initial Solution*. V okencu *Correlation Matrix* izberemo *Coefficients, Significance levels in KMO and Barlett's Test of sphericity* ter kliknemo na *Continue*. Kliknemo gumb *Rotation*. Med ponujenimi metodami izberemo *Varimax*. S klikom na *Continue* se vrnemo v osnovno pogovorno okno. Kliknemo *Scores in* izberemo *Save as variables*. S klikom na *Continue* se vrnemo v osnovno pogovorno okno in kliknemo *OK*.

Odgovori in izpisi rezultatov:

Iz korelacijske matrike v tabeli 21 izhaja, da so spremenljivke med seboj statistično značilne in močno ter srednje močno pozitivno povezane.

Tabela 22: Kaiser-Meyer-Olkinova statistika in Bartlettov test sferičnosti za večdimenzionalno spremenljivko *razvijanje agilnih vodstvenih veščin*

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,841
Bartlett's Test of Sphericity	Approx. Chi-Square	1423,156
	df	10
	Sig.	<,001

Smiselnost uporabe faktorske analize preizkušamo z Bartlettovim testom sferičnosti. Z njim preizkušamo ničelno domnevo, da je osnovna korelacijska matrika enaka matriki enote, kar pomeni, da ne obstaja povezanost med merjenimi spremenljivkami. Poleg Bartlettovega testa sferičnosti se uporablja še Kaiser-Meyer-Olkinova statistika (KMO), za katero je značilno, da je uporaba faktorske analize smiselna pri vrednosti, ki je večja od 0,5 (Tabachnick in Fidell, 2013). V našem primeru KMO (0,841) in stopnja značilnosti Bartlettovega testa ( $p < 0,001$ ) kažeta na smiselnost uporabe faktorske analize (tabela 22).

Posamezne trditve je potrebno proučiti glede na vrednost komunalitet, ki morajo biti višje od 0,40 (ibid). Vrednosti vseh komunalitet v tabeli 23 za večdimenzionalno spremenljivko *razvijanje agilnih vodstvenih veščin* so višje od 0,40, zato nismo izločili nobene spremenljivke. Komunaliteta izraža delež celotne variance i-te spremenljivke, ki je pojasnjena z

oblikovanim faktorjem. Na primer, 87,4 % celotne variance spremenljivke *vodstvo hitro in učinkovito odgovarja na spremembe v tržnem okolju* je pojasnjene z oblikovanim faktorjem.

Tabela 23: Komunalitete za večdimenzionalno spremenljivko *razvijanje agilnih vodstvenih veščin*

Communalities

	Initial	Extraction
Vodstvo hitro in učinkovito odgovarja na spremembe v tržnem okolju.	1,000	,874
Vodstvo podpira in spodbuja inovativnost na vseh ravneh podjetja.	1,000	,860
V podjetju redno prepoznavamo in nagrajujemo prizadevanja za izboljšanje agilnosti.	1,000	,866
Naša organizacijska struktura omogoča hitro prilagajanje novim izzivom.	1,000	,839
Vodstvo ustvarja pogoje, ki spodbujajo delovno zavzetost zaposlenih.	1,000	,654

Extraction Method: Principal Component Analysis.

Tabela 24: Celotna pojasnjena varianca za večdimenzionalno spremenljivko *razvijanje agilnih vodstvenih veščin*

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,092	81,849	81,849	4,092	81,849	81,849
2	,456	9,117	90,966			
3	,260	5,197	96,163			
4	,130	2,602	98,765			
5	,062	1,235	100,000			

Extraction Method: Principal Component Analysis

Iz tabele 24 je razvidno, da smo iz petih merjenih spremenljivk pridobili tudi pet glavnih komponent, a le pri prvi glavni komponenti je lastna vrednost večja od 1. V nadaljevanju je ob upoštevanju tega kriterija (lastna vrednost večja od 1) zato upoštevan oziroma izločen samo en, to je prvi faktor. Ta faktor pojasnjuje 81,849 % variabilnosti merjenih spremenljivk skupaj.

Tabela 25 prikazuje, da so vse faktorske uteži,  $a_{ij}$  višje od 0,60. Kvadrat faktorske uteži pri  $i$ -ti spremenljivki in  $j$ -tem faktorju označuje delež pojasnjene variance  $i$ -te spremenljivke z  $j$ -tim faktorjem (delež pojasnjene variance  $i$ -te spremenljivke z  $j$ -tim faktorjem je prikazan v tabeli 20). Pri *razvijanju agilnih vodstvenih veščin* ima tako najpomembnejšo vlogo spremenljivka *vodstvo hitro in učinkovito odgovarja na spremembe v tržnem okolju*, pri kateri je vrednost faktorske uteži najvišja.

**Tabela 25: Faktorske uteži za večdimenzionalno spremenljivko *razvijanje agilnih vodstvenih veščin***

## Component Matrixa

	Component 1
Vodstvo hitro in učinkovito odgovarja na spremembe v tržnem okolju.	,935
Vodstvo podpira in spodbuja inovativnost na vseh ravneh podjetja.	,927
V podjetju redno prepoznavamo in nagrajujemo prizadevanja za izboljšanje agilnosti.	,931
Naša organizacijska struktura omogoča hitro prilagajanje novim izzivom.	,916
Vodstvo ustvarja pogoje, ki spodbujajo delovno zavzetost zaposlenih.	,808

Extraction Method: Principal Component Analysis.

a. 1 components extracted.



## Naloga 2

Na osnovi podatkov v datoteki *Faktorska analiza.sav* izvedite faktorsko analizo za večdimenzionalno spremenljivko *dejavniki za prehod na agilnost* (to so spremenljivke od *DPA1 do DPA6*) in pojasnite rezultate.

Odgovori in izpisi rezultatov:

**Tabela 26: Kaiser-Meyer-Olkinova statistika in Bartlettov test sferičnosti za večdimenzionalno spremenljivko *dejavniki za prehod na agilnost***

## KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	,747	
Bartlett's Test of Sphericity	Approx. Chi-Square	1226,302
	df	15
	Sig.	<,001

V tabeli 26 KMO (0,747) in stopnja značilnosti Bartlettovega testa ( $p < 0,001$ ) kažeta na smiselnost uporabe faktorske analize.

**Tabela 27: Komunalitete za večdimenzionalno spremenljivko *dejavniki za prehod na agilnost***  
Communalities

	Initial	Extraction
Agilnost je v podjetju potrebna za digitalno agendo.	1,000	,915
Agilnost je v podjetju potrebna za hitro in neprekinjeno izboljševanje zadovoljstva strank.	1,000	,790
Agilnost je v podjetju je potrebna za hitrejšo dobavo izdelkov, prilagojenih spreminjajočim se potrebam strank.	1,000	,842
Agilnost v našem podjetju je potrebna za hitro prilagajanje tržnim spremembam.	1,000	,840
Agilnost v podjetju je potrebna za izboljšanje transparentnosti in upravljanje prioritet v podjetju.	1,000	,666
Agilnost v našem podjetju je potrebna za povečano fleksibilnost.	1,000	,735

Extraction Method: Principal Component Analysis.



Vrednosti vseh komunalitet v tabeli 27 za večdimenzionalno spremenljivko *dejavniki za prehod na agilnost* so višje od 0,40, zato nismo izločili nobene spremenljivke. Na primer, če želimo pojasniti prvo spremenljivko: 91,5 % celotne variance spremenljivke *agilnost je v podjetju potrebna za digitalno agendo* je pojasnjene z oblikovanim faktorjem.

Tabela 28: Celotna pojasnjena varianca za večdimenzionalno spremenljivko *dejavniki za prehod na agilnost*

#### Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,783	63,046	63,046	3,783	63,046	63,046	3,655	60,910	60,910
2	1,004	16,739	79,786	1,004	16,739	79,786	1,133	18,876	79,786
3	,654	10,893	90,679						
4	,355	5,916	96,596						
5	,133	2,222	98,818						
6	,071	1,182	100,000						

Extraction Method: Principal Component Analysis.

Iz tabele 28 je razvidno, da se iz šestih osnovnih spremenljivk oblikujeta dva faktorja, pri katerih je lastna vrednosti večja od 1 in ki skupaj pojasnjujeta 79,786 % variance vseh osnovnih šestih spremenljivk skupaj, in sicer prvi faktor 63,046 %, drugi faktor 16,739 %.

Tabela 29: Rotirane faktorske uteži za večdimenzionalno spremenljivko *dejavniki za prehod na agilnost*

#### Rotated Component Matrixa

	Component	
	1	2
Agilnost je v podjetju potrebna za digitalno agendo.	,078	,953
Agilnost je v podjetju potrebna za hitro in neprekinjeno izboljševanje zadovoljstva strank.	,831	,315
Agilnost je v podjetju je potrebna za hitrejšo dobavo izdelkov, prilagojenih spreminjajočim se potrebam strank.	,896	,196
Agilnost v našem podjetju je potrebna za hitro prilagajanje tržnim spremembam.	,884	,244
Agilnost v podjetju je potrebna za izboljšanje transparentnosti in upravljanje prioritet v podjetju.	,816	-,011
Agilnost v našem podjetju je potrebna za povečano fleksibilnost.	,842	-,162

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization

a. Rotation converged in 3 iterations.

Za izboljšanje strukture oblikovanih faktorjev smo vrednosti uteži rotirali z metodo Varimax. Oba faktorja se poimenujeta glede na spremenljivke, ki so v vsakem faktorju vključene in ga zastopajo. Tako bi lahko prvi faktor poimenovali na primer *digitalna agilnost* (sem sodi spremenljivka *agilnost je v podjetju potrebna za digitalno agendo*), drugi faktor, kamor sodijo vse ostale spremenljivke, pa lahko poimenujemo na primer *tržna agilnost* (tabela 29).



## Naloga 3

Želeli smo ugotoviti, ali podjetja v Sloveniji težijo k agilnosti. Lastniki velikih podjetij v Sloveniji ( $n = 260$ ) so označili stopnjo strinjanja s posameznimi trditvami (kjer pomeni 1 – sploh se ne strinjam in 5 – popolnoma se strinjam), ki se nanašajo na agilno delovno okolje.

Podatki so v datoteki: *Faktorska analiza.sav*

Trditve za večdimenzionalno spremenljivko *agilno delovno okolje* so od *ADO1* do *ADO8*.

- a) Izvedite faktorsko analizo in pojasnite rezultate.
- b) Na osnovi kazalca Cronbachove alfe ugotovite, ali je merska lestvica zanesljiva.

Potek izvedbe izračuna kazalca Cronbachove alfe: kliknemo *Analyze* in *Scale* ter nato *Reliability Analysis*. V okencu *Model* kliknemo *Alpha*. V desno okence prenesemo spremenljivke in kliknemo *OK*.

Odgovori in izpisi rezultatov:

b)

**Tabela 30: Cronbachova alfa za faktor *agilno delovno okolje***

Reliability Statistics	
Cronbach's Alpha	N of Items
,952	8

Cronbachovo alfo uporabimo takrat, ko želimo preveriti, kako zanesljive so naše merske lestvice. Merimo jo z ustreznim kazalcem zanesljivost (angl. reliability analysis), s pomočjo Cronbachovega kazalca ali koeficienta  $\alpha$ . Zanesljivost merjenja je (Tabachnick in Fidell, 2013):

- zgledna, če je koeficient Cronbach alfa večji ali enak 0,80 ( $\alpha \geq 0,80$ ),
- zelo dobra, če je koeficient Cronbach alfa večji ali enak 0,70 in manjši od 0,80,
- zmerna, če je koeficient Cronbach alfa večji ali enak 0,60 in manjši od 0,70,
- komaj sprejemljiva, če je koeficient Cronbach alfa manjši od 0,60.

Tabela 30 kaže, da je vrednost Cronbachove alfe večja od 0,80 (0,952), zato lahko trdimo, da je faktor *Agilno delovno okolje* zglede zanesljivosti.

### 3.5 Enostavna linearna regresija

Enostavno linearno regresijo smo podrobno predstavili v 1.8 podpoglavju. V tem podpoglavju bomo predstavili uporabo programa SPSS za analizo podatkov z enostavno linearno regresijo. Namen poglavja je, da pokažemo, kako uporabiti program SPSS za izvedbo analize in interpretacijo rezultatov.

S pomočjo enostavne linearne regresije analiziramo odvisnost med eno odvisno ( $y$ ) in eno neodvisno ali pojasnjevalno spremenljivko ( $x_i$ ).

Moč linearne povezanosti med spremenljivkama glede na vrednost korelacijskega in determinacijskega koeficienta je prikazana v tabeli 31.

Tabela 31: Moč linearne povezanosti glede na vrednost korelacijskega in determinacijskega koeficienta

Korelacijski koeficient ( $r_{xy}$ )	Determinacijski koeficient ( $r^2_{xy}$ )	Moč linearne povezave
0	0	Korelacije ni
0–0,5	0–0,25	Slaba korelacija
0,51–0,79	0,26–0,64	Srednje močna korelacija
0,80–0,99	0,65–0,99	Močna korelacija
1	1	Popolna korelacija

Vir: Artenjak, 2003

Kakovost regresijskega modela kot celote preverjamo z F-testom, statistično značilen vpliv pojasnjevalnih spremenljivk pa s t-testom (oziroma ene pojasnjevalne spremenljivke,  $x_1$  v primeru enostavne regresije). Z F-testom preizkušamo domnevo:

H0: Determinacijski koeficient je enak 0 ( $r^2_{xy} = 0$ )

H1: Determinacijski koeficient je večji od 0 ( $r^2_{xy} > 0$ ).

Statistično značilen vpliv pojasnjevalne spremenljivke  $x_i$  testiramo s t-testom, pri čemer preizkušamo domnevo:

H0: Regresijski koeficient  $\beta_1$  je enak 0 ( $\beta_1 = 0$ )

H1: Regresijski koeficient  $\beta_1$  ni enak 0 ( $\beta_1 \neq 0$ ).



## Naloga 1

Ugotoviti želimo, ali *razvijanje agilnih vodstvenih veščin* (neodvisna spremenljivka) vpliva na *uspešnost podjetja* (odvisna spremenljivka). Lastniki podjetij so na 5-stopenjski Likertovi lestvici označili stopnjo strinjanja (1 – sploh se ne strinjam, 2 – ne strinjam se, 3 – delno se strinjam, 4 – strinjam se, 5 – popolnoma se strinjam) z navedenimi trditvami, ki so se nanašale na razvijanje agilnih vodstvenih veščin in uspešnost podjetja. Za večdimenzionalno spremenljivko *razvijanje agilnih vodstvenih veščin (RAVV)* je predhodno bila izvedena faktorska analiza, s katero smo izločili en faktor – *razvijanje agilnih vodstvenih veščin (x)*. Potrebno še je narediti faktorsko analizo za večdimenzionalno spremenljivko *uspešnost podjetja (y)*.

Odprite datoteko *Faktorska analiza.sav*.

Preverite rezultate faktorske analize za večdimenzionalno spremenljivko *uspešnost podjetja* (zajema spremenljivke od UP1 do UP5) in jih vsebinsko pojasnite. Vrednost dobljenega faktorja shranite v datoteko s podatki.

Nato izvedite enostavno linearno regresijo in pojasnite rezultate.

Postopek – izvedba enostavne linearne regresije: kliknemo na *Analyze*, nato *Regression* in izberemo *Linear*. Odpre se pogovorno okno, v katerem spremenljivko *uspešnost podjetja* prenesemo v desno okence, to je v polje *Dependent*, spremenljivko *razvijanje agilnih vodstvenih veščin* prenesemo v desno okence, to je v polje *Independent*. Nato kliknemo *Continue* in *OK*.

Odgovori in izpisi rezultatov:

**Tabela 32: Korelacijski in determinacijski koeficient (naloga 1)**

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,593 <sup>a</sup>	,352	,349	,80660785

a. Predictors: (Constant), FAKTOR1\_Razvijanje agilnih vodstvenih veščin

Vrednost korelacijskega koeficienta znaša 0,593, kar kaže na to, da med spremenljivkama *razvijanje agilnih vodstvenih veščin* in *uspešnost podjetja* obstaja srednje močna linerna povezanost. Vrednost determinacijskega koeficienta znaša 0,352. Determinacijski koeficient pojasnjuje, da smo 35,2 % celotne variance *uspešnosti podjetja* uspeli pojasniti z variabilnostjo neodvisne spremenljivke (*razvijanje agilnih vodstvenih veščin*). Standardna

napaka ocene odvisne spremenljivke (angl. Std. error of the estimate) je različna od 0, kar pomeni, da na odvisno spremenljivko *uspešnost podjetja* vplivajo še druge spremenljivke in slučajni vplivi (tabela 32).

Tabela 33: Rezultat – F-test (naloga 1)

ANOVA<sup>a</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	91,141	1	91,141	140,084	<,001 <sup>b</sup>
	Residual	167,859	258	,651		
	Total	259,000	259			

a. Dependent Variable: FAKTOR1\_Uspešnost podjetja

b. Predictors: (Constant), FAKTOR1\_Razvijanje agilnih vodstvenih veščin

Kakovost regresijskega modela ugotavljamo s F-testom, pri čemer lahko glede na vrednost  $p$  ( $p < 0,001$ ) trdimo, da je model kot celota kakovosten. Obstaja linearna odvisnost med neodvisno spremenljivko (*razvijanje agilnih vodstvenih veščin*) in odvisno spremenljivko (*uspešnost podjetja*). Ničelno domnevo:  $H_0: r^2_{xy} = 0$  zavrnemo in sprejmemo raziskovalno domnevo  $H_1: r^2_{xy} > 0$  (tabela 33).

Tabela 34: Rezultat – regresijski koeficienti in t-test (naloga 1)

Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4,787e-17 <sup>1</sup>	,050		,000	1,000
	FAKTOR1_Razvijanje agilnih vodstvenih veščin	,593	,050	,593	11,836	<,001

a. Dependent Variable: FAKTOR1\_Uspešnost podjetja

Vrednost t-testa in stopnja značilnosti ( $p < 0,001$ ) kažeta, da je regresijski koeficient  $\beta_1$  različen od nič (potrdimo raziskovalno domnevo:  $H_1: \beta_1 \neq 0$ ), kar pomeni, da pojasnjevalna spremenljivka *razvijanje agilnih vodstvenih veščin* statistično značilno vpliva na *uspešnost podjetja* (tabela 34).

<sup>1</sup> V tem primeru je številka 4,787 pomnožena z 10 na potenco -17, kar pomeni, da gre za zelo majhno številko, ki je blizu nič. V našem primeru je konstanta tako majhna, da jo lahko za praktične namene smatramo za nič. To je lahko posledica obsega podatkov ali načina, kako so bili podatki kodirani in vneseni v model. Pomembno je tudi omeniti, da čeprav je konstanta zelo majhna, to ne vpliva na kakovost modela ali pomembnost neodvisne spremenljivke, kar potrjuje tudi p-vrednost pri tej spremenljivki ( $p < 0,001$ ).



## Naloga 2

Imamo podatke za 80 naključno izbranih gospodinjstev. Vsako gospodinjstvo je poročalo o svoji povprečni mesečni porabi (v €) in skupnih letnih prihrankih (v €). Želimo ugotoviti, ali povprečna mesečna poraba gospodinjstev vpliva na njihove skupne letne prihranke.

Odprite datoteko: *Enostavna linearna regresijska analiza.sav*.

- Pojasnite korelacijski koeficient in determinacijski koeficient.
- Pojasnite standardno napako ocene odvisne spremenljivke.
- Preverite ustreznost modela kot celote in zapišite ustrezno postavljeno domnevo. Pojasnite rezultate.
- Zapišite ustrezno postavljene domneve za preverjanje statistične značilnosti regresijskih koeficientov v regresijski funkciji ter pojasnite rezultate.
- Zapišite enačbo regresijske funkcije z ocenjenimi regresijskimi koeficienti.

Odgovori in izpisi rezultatov:

**Tabela 35: Korelacijski in determinacijski koeficient (naloga 2)**

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,997 <sup>a</sup>	,993	,993	149,56223

a. Predictors: (Constant), Povprečna mesečna poraba v €

**Tabela 36: Rezultat – F-test (naloga 2)**

ANOVA<sup>a</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	258391472,585	1	258391472,585	11551,391	<,001 <sup>b</sup>
	Residual	1744771,216	78	22368,862		
	Total	260136243,800	79			

a. Dependent Variable: Skupni letni prihranki v €

b. Predictors: (Constant), Povprečna mesečna poraba v €

**Tabela 37: Rezultat – regresijski koeficienti in t-test (naloga 2)**

Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	11973,090	44,203		270,865	<,001
	Povprečna mesečna poraba v €	-2,974	,028	-,997	-107,477	<,001

a. Dependent Variable: Skupni letni prihranki v €

Enačba regresijske funkcije z ocenjenimi regresijskimi koeficienti:

V tabeli 37 lahko razberemo ocenjeno regresijsko funkcijo:  $\hat{y}_i = b_0 + b_1x_i$ .

$\hat{y}_i$  – ocenjena vrednost spremenljivke  $y$  pri  $i$ -ti opazovani vrednosti spremenljivke  $x$

$b_0$  in  $b_1$  – ocenjena regresijska koeficienta

$\hat{y} = 11973,090 - 2,974 x_1$  (kjer pomeni  $x_1$  povprečna mesečna poraba v €).

Vrednost regresijskega koeficienta pri neodvisni spremenljivki ( $b_1x_1$ ) lahko vsebinsko pojasnimo takole: če se povprečna mesečna poraba gospodinjstva poveča za 1 €, potem se v povprečju letni prihranki zmanjšajo za 2,974 €.

Vrednost pri konstanti ( $b_0$ ) lahko vsebinsko pojasnimo takole: če gospodinjstvo ne porabi ničesar v mesecu (kar je teoretično in v praksi skoraj nemogoče), bi njihovi letni prihranki v povprečju znašali 11.973,090 €.



### Naloga 3

Imamo podatke za 80 naključno izbranih gospodinjstev. Vsako gospodinjstvo je poročalo o njihovem letnem dohodku (v €) in skupnih letnih prihrankih (v €). Cilj je ugotoviti, ali so skupni letni prihranki (v €) odvisni od letnega dohodka gospodinjstva (v €).

Odprite datoteko: *Enostavna linearna Regresijska analiza.sav*

- Pojasnite korelacijski koeficient in determinacijski koeficient.
- Pojasnite standardno napako ocene odvisne spremenljivke.
- Preverite ustreznost modela kot celote in zapišite ustrezno postavljeno domnevo. Pojasnite rezultate.
- Zapišite ustrezno postavljene domneve za preverjanje statistične značilnosti regresijskih koeficientov v regresijski funkciji ter pojasnite rezultate.
- Zapišite enačbo regresijske funkcije z ocenjenimi regresijskimi koeficienti.

### 3.6 Multipla regresijska analiza

Multipla regresijsko analizo uporabimo takrat, kadar na odvisno spremenljivko ( $y$ ) vpliva več neodvisnih spremenljivk  $x_i$  ( $i = 1, 2, \dots, k$ ) in so izpolnjeni pogoji za izvedbo multiple regresijske analize. Multipli korelacijski koeficient  $R$  kaže jakost povezanosti med odvisno in  $k$  neodvisnimi spremenljivkami in je prikazan v absolutni vrednosti. Prilagojeni determinacijski koeficient  $R^2$  pa predstavlja delež variance odvisne spremenljivke, ki je pojasnjena z variabilnostjo v model vključenih neodvisnih spremenljivk (Tabachnick in Fidell, 2013).

Kakovost regresijskega modela kot celote preverjamo z F-testom kjer preizkušamo domnevi:

H0: Prilagojeni determinacijski koeficient je enak 0 ( $R^2 = 0$ )

H1: Prilagojeni determinacijski koeficient je večji od 0 ( $R^2 > 0$ ).

Statistično značilen vpliv pojasnjevalnih spremenljivk testiramo s t-testom, pri čemer preizkušamo domneve:

H0:  $\beta_i = 0$

H1:  $\beta_i \neq 0$

( $i = 1, 2, \dots, k$ ).



## Naloga 1

V podjetju, ki se ukvarja z izdelavo biološko razgradljive embalaže, želijo raziskati dejavnike, ki vplivajo na učinkovitost proizvodnje. Učinkovitost proizvodnje so merili kot količino kvalitetno izdelane embalaže na uro (v kilogramih). Želimo analizirati, ali na učinkovitost proizvodnje (kg/uro) vplivata dve neodvisni spremenljivki stroški materiala na uro (v €) in število zaposlenih na izmeno.

Podatki so v datoteki: *Multipla regresija\_naloga1.sav*

- Pojasnite multipli korelacijski koeficient in prilagojeni determinacijski koeficient.
- Preverite ustreznost modela kot celote in zapišite ustrezno postavljeno domnevo. Pojasnite rezultate.



- c) Zapišite ustrezno postavljene domneve za preverjanje statistične značilnosti regresijskih koeficientov v regresijski funkciji ter pojasnite rezultate.
- d) Zapišite enačbo regresijske funkcije z ocenjenimi regresijskimi koeficienti.

Odgovori in izpisi rezultatov:

**Tabela 38: Rezultat – multipli korelacijski in prilagojeni determinacijski koeficient**

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,999 <sup>a</sup>	,998	,997	1,96886

a. Predictors: (Constant), število zaposlenih na izmeno, stroški materiala na uro (€)

Vrednost multiplega korelacijskega koeficienta znaša 0,999, kar kaže na to, da med odvisno spremenljivko *učinkovitost proizvodnje (kg/uro)* in neodvisnima spremenljivkama *stroški materiala na uro (v €)* in *število zaposlenih na izmeno* obstaja močna povezanost. Vrednost prilagojenega multiplega determinacijskega koeficienta znaša 0,997. Prilagojen determinacijski koeficient pojasnjuje, da je 99,7 % celotne variance *učinkovitost proizvodnje (kg/uro)* pojasnjene z regresijskim modelom (s spremenljivkama *stroški materiala na uro ( $x_1$ )* in *število zaposlenih na izmeno ( $x_2$ )*) (tabela 38).

**Tabela 39: Rezultat – F-test (ANOVA)**

ANOVA<sup>a</sup>

	Model	Sum of Squares	df	Mean Square	F	Sig.
1	Regression	44020,303	2	22010,152	5677,966	<,001 <sup>b</sup>
	Residual	104,663	27	3,876		
	Total	44124,967	29			

a. Dependent Variable: Učinkovitost proizvodnje (kg/uro)

b. Predictors: (Constant), število zaposlenih na izmeno, stroški materiala na uro (€)

Kakovost regresijskega modela kot celote smo preverjali z F-testom, pri čemer lahko glede na vrednost  $p$  ( $p < 0,001$ ) trdimo, da je model kakovosten. Navedeno pomeni, da obstaja odvisnost med odvisno spremenljivko (*učinkovitost proizvodnje*) in vsaj eno neodvisno spremenljivko (tabela 39).

Z F-testom preizkušamo domnevi:

H0: Prilagojeni determinacijski koeficient je enak 0 ( $R^2 = 0$ )

H1: Prilagojeni determinacijski koeficient je večji od 0 ( $R^2 > 0$ ) .

Na osnovi rezultatov lahko zavrնemo ničelno domnevo, da je  $R^2 = 0$ , saj je  $p < 0,05$  (tabela 39).

Tabela 40: Rezultat – regresijski koeficienti in t-test

Model		Coefficients <sup>a</sup>				
		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3,407	1,224		-2,783	,010
	Stroški materiala na uro (€)	,358	,039	,640	9,129	<,001
	Število zaposlenih na izmeno	3,261	,633	,361	5,156	<,001

a. Dependent Variable: Učinkovitost proizvodnje (kg/uro)

Pri tabeli 40 za preverjanje statistične značilnosti regresijskih koeficientov v regresijski funkciji napišemo naslednje domneve:

$X_1$  (Stroški materiala na uro):

$X_2$  (Število zaposlenih na izmeno):

$H_0: \beta_1 = 0$

$H_0: \beta_2 = 0$

$H_1: \beta_1 \neq 0$

$H_1: \beta_2 \neq 0$

T-test in stopnja značilnosti za oba regresijska koeficienta ( $p < 0,05$ ) kažeta, da obe pojasnjevalni spremenljivki statistično značilno vplivata na odvisno spremenljivko (zato sprejmemo raziskovalni domnevi  $H_1: \beta_1 \neq 0$  in  $H_1: \beta_2 \neq 0$ ).

Dobljena enačba regresijske funkcije z ocenjenimi vrednostmi regresijskih koeficientov na osnovi uporabljenih vzorčnih podatkov je (tabela 40):

$$\hat{y} = -3,407 + 0,358 x_1 + 3,261 x_2$$

Ocenjeni vrednosti regresijskih koeficientov povesta, za koliko enot se v povprečju spremeni vrednost odvisne spremenljivke, če se vrednost posamezne pojasnjevalne spremenljivke spremeni za eno enoto, vrednost druge pojasnjevalne spremenljivke pa ostane pri tem nespremenjena (ne obstaja multikolineranost med pojasnjevalnima spremenljivkama).

Ocenjeni regresijski koeficient pri  $x_1$  pove, da se učinkovitost proizvodnje (kg/uro) v povprečju poveča za 0,358 kg/uro, če se stroški materiala na uro (€) povečajo za 1 € in pri tem vrednost spremenljivke število zaposlenih na izmeno ostane nespremenjena.



## Naloga 2

Banka želi raziskati dejavnike, ki vplivajo na zadovoljstvo strank. Zadovoljstvo strank je merjeno na lestvici od 1 do 10, kjer 10 pomeni najvišje zadovoljstvo. Vodstvo banke želi ugotoviti, ali na zadovoljstvo strank vpliva število bančnih storitev, ki jih stranka uporablja (npr. računi, krediti, spletne storitve) in povprečni čas čakanja na storitev v minutah.

Podatki so v datoteki: *Multipla regresija\_naloga2.sav*

- Pojasnite multipli korelacijski koeficient in prilagojeni determinacijski koeficient.
- Preverite ustreznost modela kot celote in zapišite ustrezno postavljeno domnevo. Pojasnite rezultate.
- Zapišite ustrezno postavljene domneve za preverjanje statistične značilnosti regresijskih koeficientov v regresijski funkciji ter pojasnite rezultate.
- Zapišite enačbo regresijske funkcije z ocenjenimi regresijskimi koeficienti.



## Naloga 3

Bolnišnica želi raziskati, kako dva različna dejavnika vplivata na čas okrevanja pacientov po ortopedskih operacijah. Odvisna spremenljivka bo *čas okrevanja*, merjen v dnevih, od operacije do odpusta. Dve neodvisni spremenljivki bosta *starost pacienta* in *intenzivnost fizioterapije* (merjena s številom terapevtskih obravnav na teden).

Podatki so v datoteki: *Multipla regresija\_naloga3.sav*

- Pojasnite multipli korelacijski koeficient in prilagojeni determinacijski koeficient.
- Preverite ustreznost modela kot celote in zapišite ustrezno postavljeno domnevo. Pojasnite rezultate.
- Zapišite ustrezno postavljene domneve za preverjanje statistične značilnosti regresijskih koeficientov v regresijski funkciji ter pojasnite rezultate.
- Zapišite enačbo regresijske funkcije z ocenjenimi regresijskimi koeficienti.



## Naloga 4

Tehnološko podjetje razvija algoritme strojnega učenja za prepoznavanje vzorcev v velikih podatkovnih zbirkah. Želijo raziskati, kako dve spremenljivki vplivata na natančnost (angl. accuracy) algoritma, izraženo v odstotkih. Dve neodvisni spremenljivki bosta *količina učnih podatkov* (v GB) in *število plasti v nevronske mreži*. Podatki za 30 testiranj so torej:

- količina učnih podatkov (GB): koliko gigabajtov podatkov je bilo uporabljenih za treniranje algoritma;
- število plasti: koliko plasti ima uporabljena nevronska mreža;
- natančnost algoritma (%): odstotek uspešno prepoznanih vzorcev.

Podatki so v datoteki: *Multipla regresija\_naloga4.sav*

- a) Pojasnite multipli korelacijski koeficient in prilagojeni determinacijski koeficient.
- b) Preverite ustreznost modela kot celote in zapišite ustrezno postavljeno domnevo. Pojasnite rezultate.
- c) Zapišite ustrezno postavljene domneve za preverjanje statistične značilnosti regresijskih koeficientov v regresijski funkciji ter pojasnite rezultate.
- d) Zapišite enačbo regresijske funkcije z ocenjenimi regresijskimi koeficienti.

## 4 Rešitve računskih nalog

### 4.1 Urejanje in prikazovanje podatkov

#### Naloga 1a)

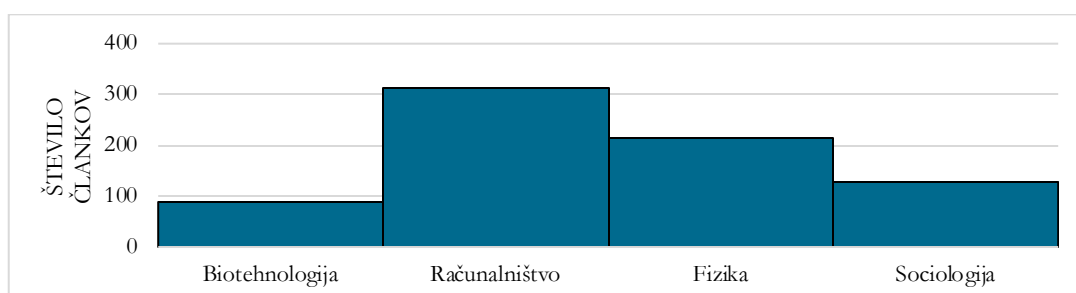
Statistična enota: posamezen objavljen znanstveni članek.

Statistične spremenljivke:

- članki v konferenčnih zbornikih (številka, nezvezna spremenljivka),
- članki v znanstvenih revijah (številka, nezvezna spremenljivka),
- področje raziskav (opisna spremenljivka).

#### Naloga 1b)

Prikaz stvarne statistične vrste s stolpci.



**Naloga 1c)**

Izračun relativne strukture objavljenih člankov v znanstvenih revijah po raziskovalnih področjih:

Področje raziskav	Članki v znanstvenih revijah $f_k$	Strukturni odstotek $f_{k\%} = \frac{fk}{n} \cdot 100$
Biotehnologija	156	$\frac{156}{835} \cdot 100 = 18,68 \%$
Računalništvo	205	24,55 %
Fizika	198	23,71 %
Sociologija	276	33,05 %
Skupaj	835	100 %

Grafični prikaz relativne strukture objavljenih člankov v znanstvenih revijah po raziskovalnih področjih:

**Naloga 2 b)**

Cenovni razred (v 1.000 \$)	Število prodanih avtomobilov	$y_{k, \min}$	$y_{k, \max}$	$i_k$
Od 21 do 30	36	20,5	30,5	10
Od 31 do 40	30	30,5	40,5	10
Od 41 do 50	25	40,5	50,5	10
Od 51 do 60	14	50,5	60,5	10
Od 61 do 70	6	60,5	70,5	10
Skupaj	111			

**Naloga 2 c)**

Enačba za strukturni odstotek:  $f_{\%k} = \frac{fk}{N} \cdot 100$

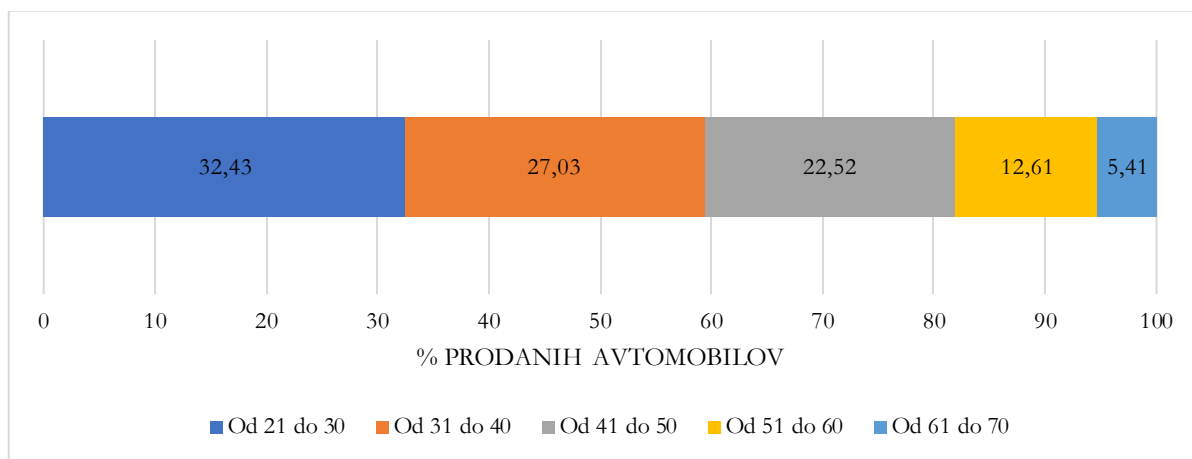
Cenovni razred (v 1.000 \$)	Število prodanih avtomobilov ( $f_k$ )	$f_{\%k}$
Od 21 do 30	36	32,43
Od 31 do 40	30	27,03
Od 41 do 50	25	22,52
Od 51 do 60	14	12,61
Od 61 do 70	6	5,41
<b>Skupaj</b>	111	100 %

Na primer:

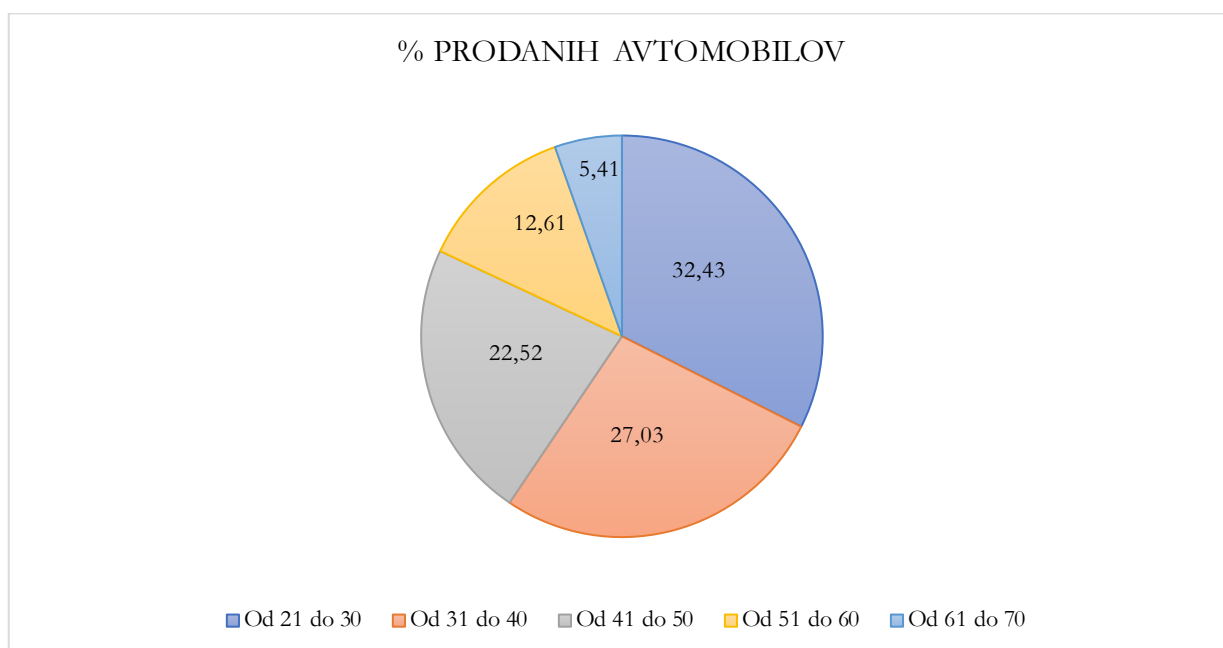
$$f_{\%k} = \frac{36}{111} \cdot 100\% = 32,43 \%$$

$$f_{\%k} = \frac{30}{111} \cdot 100\% = 27,03 \%$$

Struktura prodanih avtomobilov glede na cenovni razred v strukturnem stolpcu:

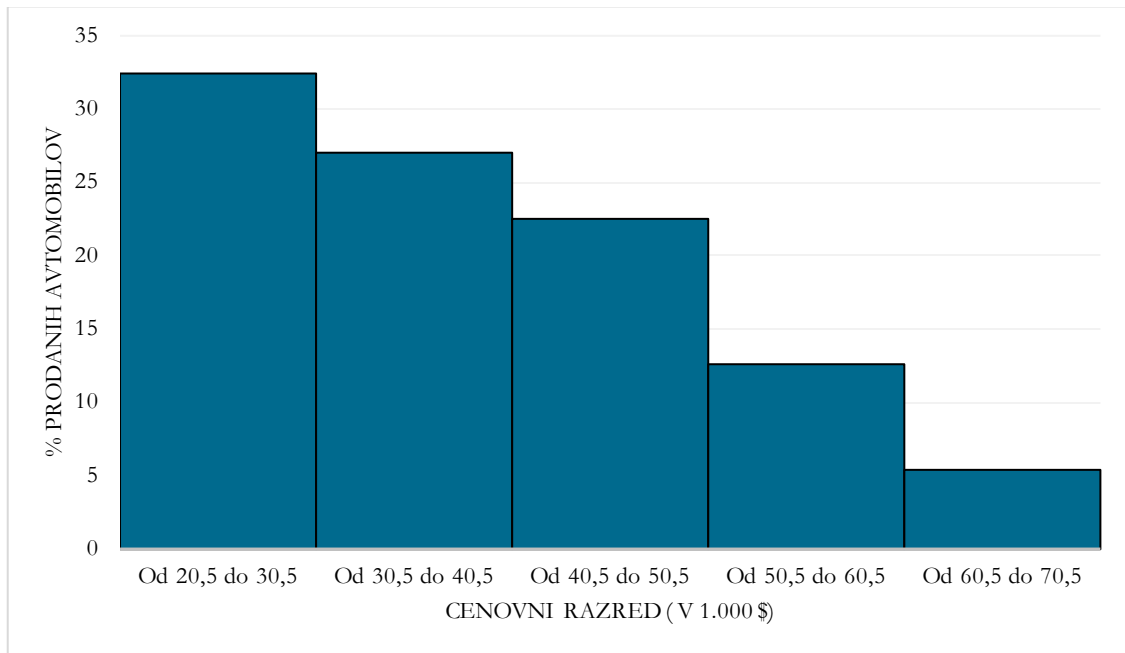


Strukturni krog:



**Naloga 2 d)**

Frekvenčni histogram:

**Naloga 3 b)**

Starost zaposlenih	Zaposleni ( $f_k$ )	Kumulativna frekvenčna porazdelitev $F_k$
Od 20 do pod 30	16	16
Od 30 do pod 40	20	36
Od 40 do pod 50	28	64
Od 50 do pod 60	23	87
Od 60 do pod 70	11	98
<b>Skupaj</b>	<b>98</b>	

**Naloga 3 c), d)**Enačba za strukturni odstotek:  $f\%k = \frac{fk}{N} \cdot 100$ 

Starost zaposlenih	Zaposleni ( $f_k$ )	Kumulativna frekvenčna porazdelitev $F_k$	$f\%k$	$F\%k$
Od 20 do pod 30	16	16	16,3	16,3
Od 30 do pod 40	20	36	20,4	36,7
Od 40 do pod 50	28	64	28,6	65,3
Od 50 do pod 60	23	87	23,5	88,8
Od 60 do pod 70	11	98	11,2	100
<b>Skupaj</b>	<b>98</b>		<b>100 %</b>	

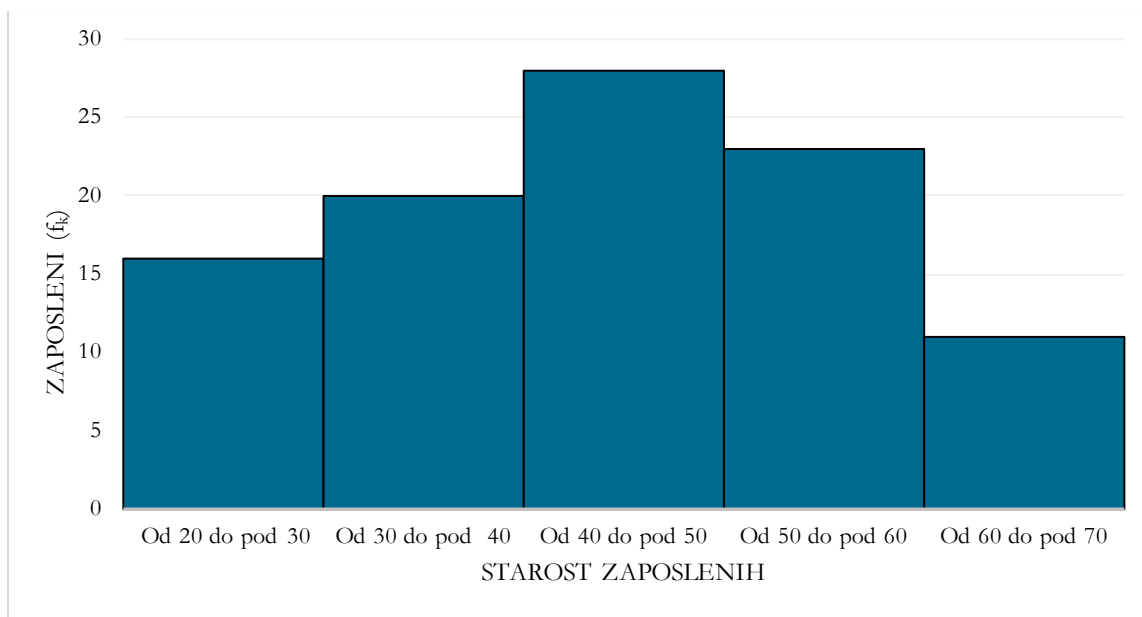


Odgovor pri točki c: 20,4 % zaposlenih je v podjetju A starih od 30 do 40 let.

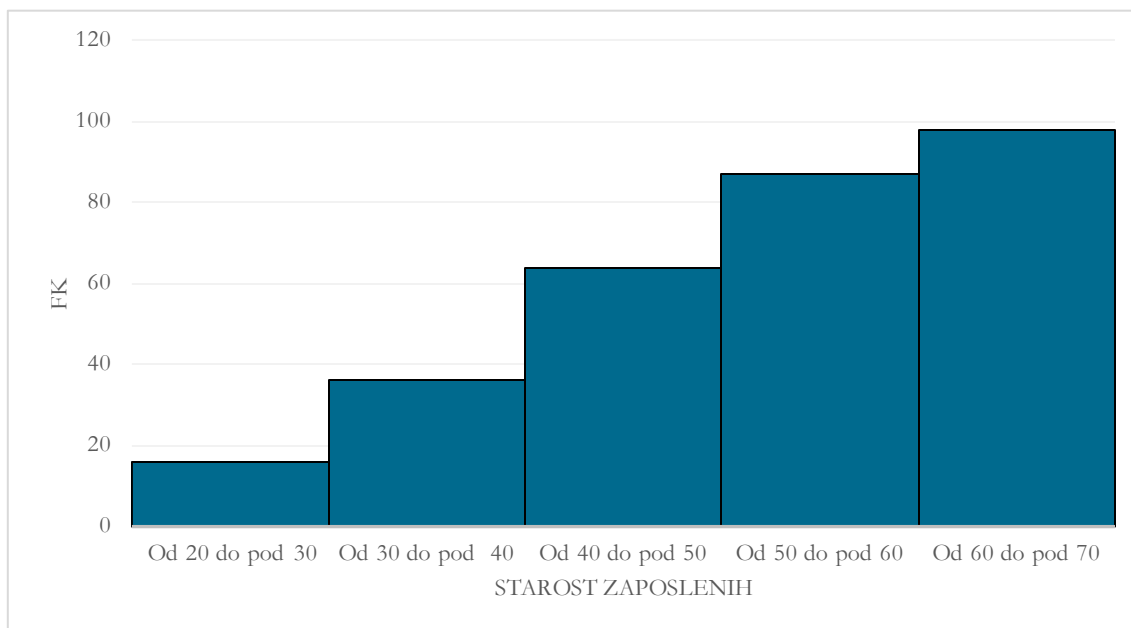
Odgovor pri točki d: 88,8 % zaposlenih je v podjetju A starih do 60 let.

### Naloga 3 e)

Histogram frekvenčne porazdelitve:



Histogram kumulativne frekvenčne porazdelitve:



**Naloga 4 b)**

Število dni dopusta	$f_k$	$F_k$	$f\%_{0k}$	$F\%_{0k}$
Od 5 do pod 10	6	6	5,22	5,22
Od 10 do pod 15	17	23	14,78	20
Od 15 do pod 20	38	61	33,04	53,04
Od 20 do pod 25	33	94	28,70	81,74
Od 25 do pod 30	12	106	10,43	92,17
Od 30 do pod 35	9	115	7,83	100
<b>Skupaj</b>	115		100 %	

**Naloga 4 c)**

33,04 % zaposlenih je imelo od 15 do pod 20 dni dopusta.

**Naloga 4 d)**

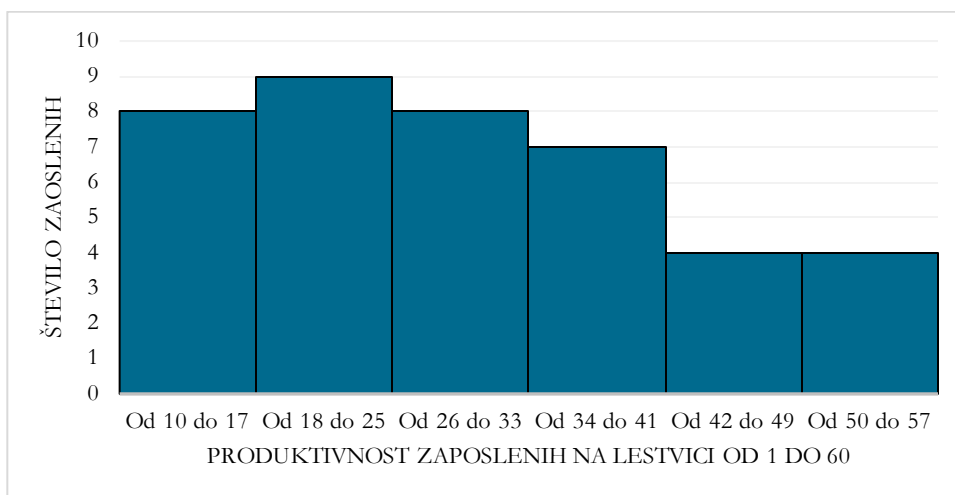
81,74 % zaposlenih je imelo do pod 25 dni dopusta.

**Naloga 5 b)**

Produktivnost	$f_k$	$F_k$
Od 10 do 17	8	8
Od 18 do 25	9	17
Od 26 do 33	8	25
Od 34 do 41	7	32
Od 42 do 49	4	36
Od 50 do 57	4	40

**Naloga 5 c)**

Histogram frekvenčne porazdelitve

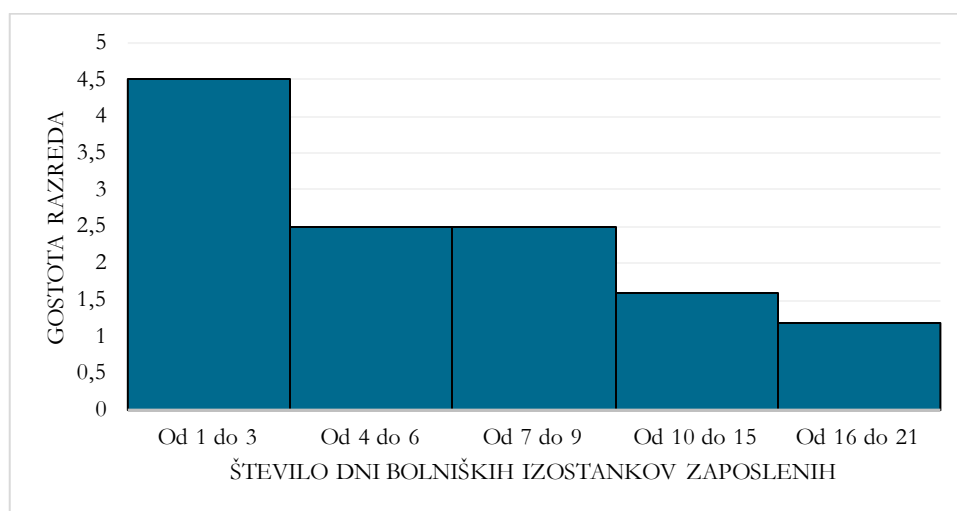


## Naloga 6 b)

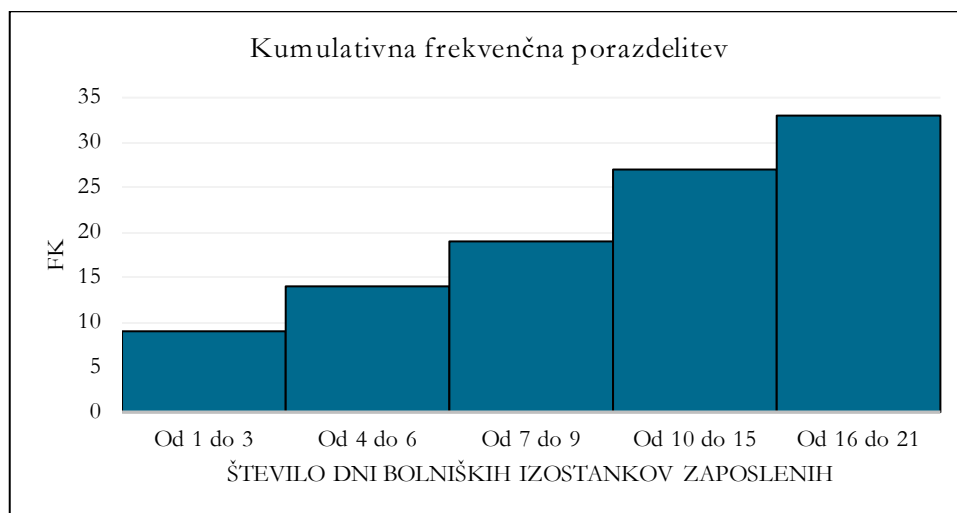
Število dni bolniških izostankov	Frekvenca $f_k$	Kumulativna frekvenčna porazdelitev $F_k$	Širina razreda $i_k$	Gostota razreda $g_k = \frac{f_k}{i_k}$	$f\%_k$	$F\%_k$
Od 1 do 3	9	9	2	4,5	27,2	27,2
Od 4 do 6	5	14	2	2,5	15,2	42,4
Od 7 do 9	5	19	2	2,5	15,2	57,6
Od 10 do 15	8	27	5	1,6	24,2	81,8
Od 16 do 21	6	33	5	1,2	18,2	100%
<b>Skupaj</b>	33				100%	

## Naloga 6 c)

Pri frekvenčni porazdelitvi z različno širokimi razredi na y osi prikažemo gostoto.



Prikaz kumulativne frekvenčne porazdelitve:



**Naloga 6 d)**

15,15 % zaposlenih je imelo od 7 do 9 dni bolniških izostankov.

**Naloga 6 e)**

81,8 % zaposlenih je imelo do 15 dni bolniških izostankov.

**Naloga 7 a)**

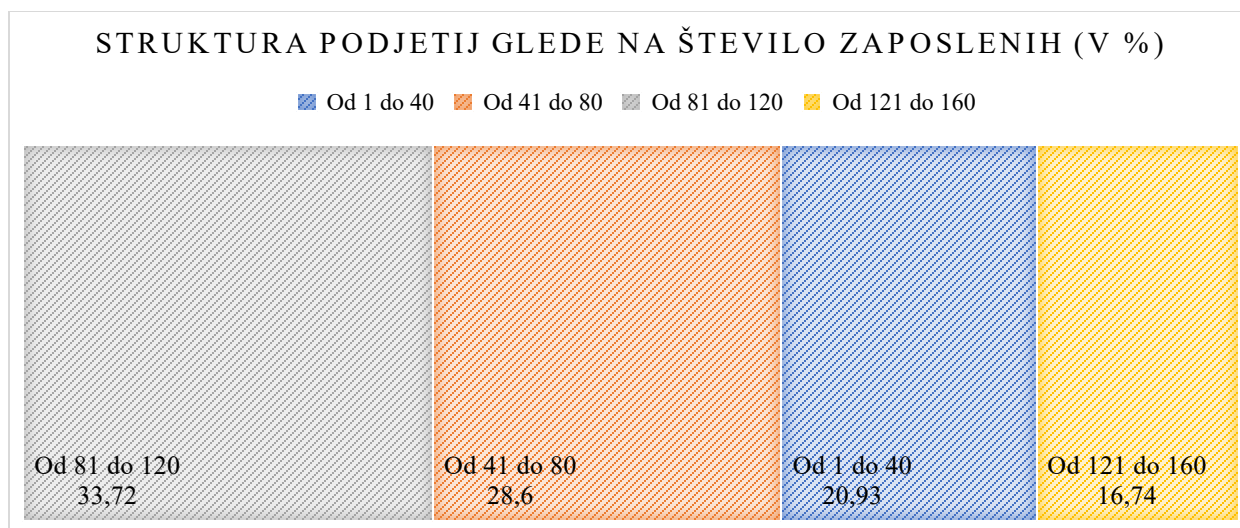
Podani so podatki za 430 velikih in srednje velikih podjetij glede na število zaposlenih:

Število zaposlenih	Število podjetij $f_k$	$y_{k,min}$	$y_{k,max}$	Širina razreda $i_k = y_{k,max} - y_{k,min}$
Od 1 do 40	90	0,5	40,5	40
Od 41 do 80	123	40,5	80,5	40
Od 81 do 120	145	80,5	120,5	40
Od 121 do 160	72	120,5	160,5	40

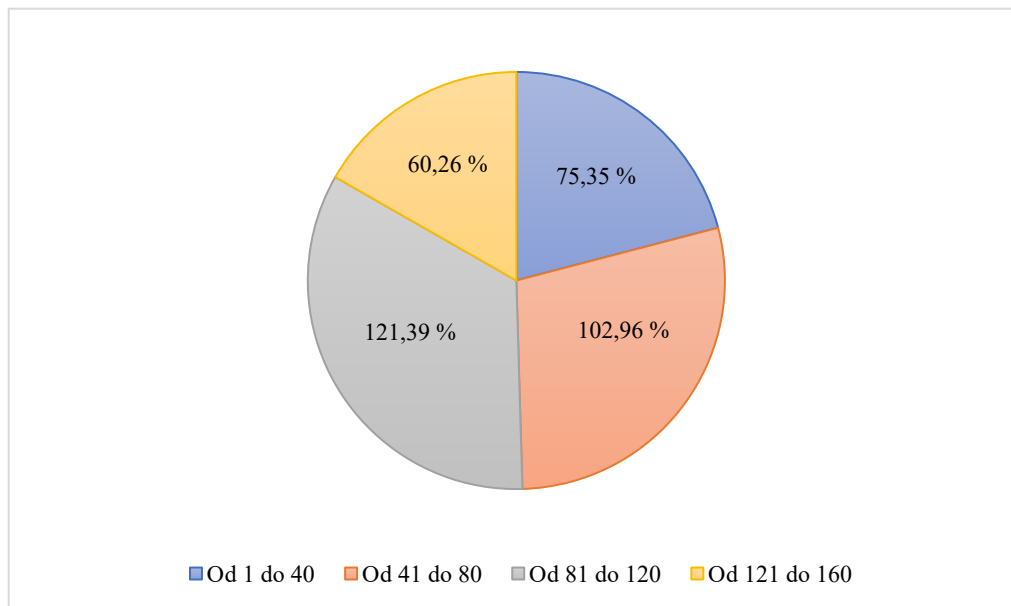
**Naloga 7 b)**

Število zaposlenih	Število podjetij $f_k$	$f\%_k$	Strukturni krog $f\%_k \cdot 3,6$
Od 1 do 40	90	20,93	75,35
Od 41 do 80	123	28,60	102,96
Od 81 do 120	145	33,72	121,39
Od 121 do 160	72	16,74	60,26
<b>Skupaj</b>	<b>430</b>	<b>100%</b>	<b>360</b>

Prikaz strukture podjetij glede na število zaposlenih v strukturnem stolpcu:



Prikaz strukture podjetij glede na število zaposlenih v strukturnem krogu:



## 4.2 Deskriptivna statistika

### Naloga 8b)

Variacijski razmik:  $VR = y_{\max} - y_{\min}$

$$VR = 7 - 1,8 = 5,2 \%$$

Odgovor: Učenci se glede na odstotek izboljšanja uspešnosti pri določenem predmetu po uvedbi novega učnega programa med seboj razlikujejo za največ 5,2 %.

### Naloga 8c)

Za izračun povprečnega odstotka uspešnosti učencev pri določenem predmetu po uvedbi novega učnega programa uporabimo enačbo (*aritmetična sredina iz nerazvrščenih vrednosti*):

$$\bar{y} = \frac{1}{N}(y_1 + y_2 + \dots + y_N) = \frac{1}{N} \sum_{i=1}^N y_i$$

$$\bar{y} = \frac{1}{10} \cdot (1,8 + 4,5 + 3,2 + 6,1 + 2,5 + \dots + 7) = \frac{41}{10} = 4,1 \%$$

Odgovor: Povprečen odstotek izboljšanja uspešnosti učencev pri določenem predmetu po uvedbi novega učnega programa za 10 srednjih šol znaša 4,1 %.

**Naloga 8 d)**

Izračun koeficienta variabilnosti v odstotku:

$$KV\% = \frac{\sigma}{\bar{y}} \times 100$$

$$\bar{y} = 4,1 \%$$

Izračunamo varianco iz nerazvrščenih vrednosti po enačbi:

$$VAR = \sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2$$

$$\sigma^2 = \frac{1}{10} \cdot [(1,8 - 4,1)^2 + (4,5 - 4,1)^2 + (3,2 - 4,1)^2 + (6,1 - 4,1)^2 + \dots + (7 - 4,1)^2] = 2,56 (\%)^2$$

Izračunamo standardni odklon po enačbi:

$$SD = \sigma = \sqrt{VAR} = \sqrt{\sigma^2},$$

$$\sigma = \sqrt{2,56} = 1,6 \%$$

$$KV\% = \frac{1,6}{4,1} \cdot 100 = 39,02 \%$$

Odgovor: Delež standardnega odklona v aritmetični sredini znaša 39,02 %.

**Naloga 8e)**

Podatke uredimo v ranžirno vrsto:

<b>R<sub>i</sub></b>	1	2	3	4	5	6	7	8	9	10
<b>y<sub>i</sub></b>	1,8	2,5	2,9	3	3,2	4,5	4,8	5,2	6,1	7

Uporabimo enačbo za kvartilni razmik:  $Q = Q_3 - Q_1$  (kjer je  $Q_3 = 75\%$  in  $Q_1 = 25\%$ ), vendar je potrebno najprej izračunati:

relativni rang  $P_i$  je znan:  $P_i = 0,25$  (kjer je  $Q_1 = 25\%$ ), zato pri izračunu uporabimo enačbe (*kvantili iz nerazvrščenih vrednosti*):

$$R_i = N \times P_i + 0,5$$

$$R_0 \leq R_i < R_1$$

$$y_i = y_0 + \frac{R_i - R_0}{R_1 - R_0} \times (y_1 - y_0)$$

Izračunamo rang  $R_i$ :

$$R_i = 10 \cdot 0,25 + 0,5 = 3$$

Nato odčitamo vrednost  $y_i$ :

$$y_i = 2,9 \%$$

Odgovor: 25 % srednjih šol je imelo odstotek izboljšanja uspešnosti učencev pri določenem predmetu po uvedbi novega učnega programa 2,9 % ali manj, 75 % srednjih šol pa več kot 2,9 %.

$$P_i = 0,75 \text{ (kjer je } Q_3 = 75 \%)$$

$$R_i = 10 \cdot 0,75 + 0,5 = 8$$

Nato odčitamo vrednost  $y_i$ :

$$y_i = 5,2 \%$$

Odgovor: 75 % srednjih šol je imelo odstotek izboljšanja uspešnosti učencev pri določenem predmetu po uvedbi novega učnega programa 5,2 % ali manj, 25 % srednjih šol pa več kot 5,2 %.

Kvartilni razmik:

$$Q = Q_3 - Q_1 = 5,2 - 2,9 = 2,3 \%$$

Odgovor: 50% srednjih šol, ki glede na odstotek izboljšanja uspešnosti učencev pri določenem predmetu po uvedbi novega učnega programa ležijo na sredini ranžirne vrste, se razlikuje za največ 2,3 %.

**Naloga 9a)**

Rešitve za organizacijo A:

<b>R<sub>i</sub></b>	1	2	3	4	5	6	7	8	9	10	11	12
<b>y<sub>i</sub></b>	22	26	29	30	33	34	38	39	40	41	43	45

a)

Izračun odstotka zaposlenih, ki so za izdelek porabili manj kot 36 minut:

$y_i$  je znan ( $y_i = 36$ ), zato pri izračunu uporabimo enačbe (kvantilni rangi iz nerazvrščenih vrednosti):

$$y_i = 36$$

$$y_0 \leq y_i < y_1$$

$$R_i = R_0 + \frac{y_i - y_0}{y_1 - y_0} \times (R_1 - R_0)$$

$$P_i = \frac{R_i - 0,5}{N}$$

V ranžirni vrsti poiščemo položaj dane vrednosti:

$$y_i = 36$$

$$y_0 \leq y_i < y_1$$

$$y_0 = 34 \leq y_i = 36 < y_1 = 38$$

Iz vrednosti členov poiščemo ustrezajoče range:

$$R_0 \leq R_i < R_1$$

$$R_0 = 6 \leq R_i < R_1 = 7$$

Izračunamo rang  $R_i$ :

$$R_i = 6 + \frac{36 - 34}{38 - 34} \cdot (7 - 6) = 6,5$$

Izračunamo relativni rang  $P_i$ :

$$P_i = \frac{6,5 - 0,5}{12} = 0,5$$



Odgovor: 50 % zaposlenih je za izdelavo enega izdelka porabilo 36 minut ali manj.

*Izračun odstotka zaposlenih, ki so za izdelek porabili več kot 40 minut:*

V ranžirni vrsti poiščemo položaj dane vrednosti:

$$y_i = 40$$

$$y_0 \leq y_i < y_1$$

$$y_0 = 39 \leq y_i = 40 < y_1 = 41$$

Iz vrednosti členov poiščemo ustrezajoče range:

$$R_0 \leq R_i < R_1$$

$$R_0 = 8 \leq R_i < R_1 = 10$$

Izračunamo rang  $R_i$ :

$$R_i = 8 + \frac{40 - 39}{40 - 39} \cdot (10 - 8) = 10$$

Izračunamo relativni rang  $P_i$ :

$$P_i = \frac{10 - 0,5}{12} = 0,79$$

Odgovor: 79 % zaposlenih je za izdelavo enega izdelka porabilo 40 minut ali manj, 21 % zaposlenih pa več kot 40 minut.

b)

*Izračun časa, ki ga je porabilo 30 % zaposlenih z največjo (oziroma najmanjšo) porabo časa za en izdelek:*

Relativni rang  $P_i$  je znan:  $P_i = 0,3$

$$R_i = N \times P_i + 0,5$$

$$R_0 \leq R_i < R_1$$

$$y_i = y_0 + \frac{R_i - R_0}{R_1 - R_0} \times (y_1 - y_0)$$

Izračunamo rang  $R_i$ :

$$R_i = 12 \cdot 0,3 + 0,5 = 4,1$$

Nato določimo vrednosti:

$$R_0 \leq R_i < R_1$$

$$R_0 = 4 \leq R_i = 4,1 < R_1 = 5$$

$$y_0 \leq y_i < y_1$$

$$y_0 = 30 \leq y_i < y_1 = 33$$

Nato izračunamo vrednost  $y_i$ :

$$y_i = 30 + \frac{4,1 - 4}{5 - 4} \cdot (33 - 30) = 30,3 \text{ minut}$$

Odgovor: 30 % zaposlenih je za izdelavo enega izdelka porabilo 30,3 minut ali manj, 70 % zaposlenih pa več kot 30,3 minut.

c) *Izračun decilnega razmika:*

Uporabimo enačbo za decilni razmik:  $D = D_9 - D_1$  (kjer je  $D_9 = 90 \%$  in  $D_1 = 10 \%$ ).

Relativni rang  $P_i$  je znan:  $P_i = 0,1$  (kjer je  $D_1 = 10 \%$ ), zato pri izračunu uporabimo enačbo:

$$R_i = N \times P_i + 0,5$$

$$R_0 \leq R_i < R_1$$

$$y_i = y_0 + \frac{R_i - R_0}{R_1 - R_0} \times (y_1 - y_0)$$

Izračunamo rang  $R_i$ :

$$R_i = 12 \cdot 0,1 + 0,5 = 1,7$$

Nato določimo vrednosti:

$$R_0 \leq R_i < R_1$$

$$R_0 = 1 \leq R_i = 1,7 < R_1 = 2$$

$$y_0 \leq y_i < y_1$$

$$y_0 = 22 \leq y_i < y_1 = 26$$

Nato izračunamo vrednost  $y_i$ :

$$y_i = 22 + \frac{1,7 - 1}{2 - 1} \cdot (26 - 22) = 24,8 \text{ minut}$$

Odgovor: 10 % zaposlenih je porabilo za izdelavo enega izdelka 24,8 minut ali manj, 90 % zaposlenih pa več kot 24,8 minut.

$$P_i = 0,9 \text{ (kjer je } D_9 = 90 \text{ \%)}$$

$$R_i = 12 \cdot 0,9 + 0,5 = 11,3$$

$$R_0 \leq R_i < R_1$$

$$R_0 = 11 \leq R_i = 11,3 < R_1 = 12$$

$$y_0 \leq y_i < y_1$$

$$y_0 = 43 \leq y_i < y_1 = 45$$

Nato izračunamo vrednost  $y_i$ :

$$y_i = 43 + \frac{11,3 - 11}{12 - 11} \cdot (45 - 43) = 43,6 \text{ minut}$$

Odgovor: 90 % zaposlenih je porabilo za izdelavo enega izdelka 43,6 minut ali manj, 10 % zaposlenih pa več kot 43,6 minut.

Decilni razmik:

$$D = D_9 - D_1 = 43,6 - 24,8 = 18,8 \text{ minut}$$

Odgovor: 80% zaposlenih, ki glede na porabljen čas za izdelavo enega izdelka ležijo na sredini ranžirne vrste, se razlikuje za največ 18,8 minut.

d) *Izračun povprečno porabljenega časa za en izdelek:*

$$\bar{y} = \frac{1}{N}(y_1 + y_2 + \dots + y_N) = \frac{1}{N} \sum_{i=1}^N y_i$$

$$\bar{y} = \frac{1}{12} \cdot (22 + 26 + 29 + 30 + 33 + \dots + 45) = \frac{420}{12} = 35 \text{ minut}$$

Odgovor: Zaposleni v organizaciji A so v povprečju porabili za izdelavo enega izdelka 35 minut.

e) Čas za en izdelek, ki ga je porabilo 50 % zaposlenih z najmanjšo porabo časa za en izdelek:

Izračun mediane ( $Me = 50\%$ ):

$$P_i = 0,5$$

$$R_i = 12 \cdot 0,5 + 0,5 = 6,5$$

$$R_0 \leq R_i < R_1$$

$$R_0 = 6 \leq R_i = 6,5 < R_1 = 7$$

$$y_0 \leq y_i < y_1$$

$$y_0 = 34 \leq y_i < y_1 = 38$$

Nato izračunamo vrednost  $y_i$ :

$$y_i = 34 + \frac{6,5 - 6}{7 - 6} \cdot (38 - 34) = 36 \text{ minut}$$

Odgovor: 50 % zaposlenih je imelo porabo časa za izdelavo enega izdelka 36 minut ali manj, 50 % zaposlenih pa več kot 36 minut.

### Naloga 9b)

a) Izračun variacijskega razmika:

Uporabimo enačbo za variacijski razmik:  $VR = y_{\max} - y_{\min}$

Organizacija A:

$$VR = 45 - 22 = 23 \text{ minut}$$

Odgovor: Zaposleni se v organizaciji A glede na porabljen čas za izdelavo izdelka med seboj razlikujejo za največ 23 minut.

Organizacija B:

$$VR = 46 - 22 = 24 \text{ minut}$$

Odgovor: Zaposleni se v organizaciji B glede na porabljen čas za izdelavo izdelka med seboj razlikujejo za največ 24 minut.

b) Izračun variance in standardnega odklona:

$$\begin{aligned} \text{Organizacija A:} \quad \sigma^2 &= \frac{15266}{12} - 35^2 = 47,17 \text{ minut}^2 \\ \sigma &= 6,87 \text{ minut} \end{aligned}$$

$$\begin{aligned} \text{Organizacija B:} \quad \sigma^2 &= \frac{817296}{730} - 33,05^2 = 27,28 \text{ minut}^2 \\ \sigma &= 5,22 \text{ minut} \end{aligned}$$

Pri organizaciji A smo uporabili enačbo (*varianca iz nerazvrščenih vrednosti*):

$$VAR = \sigma^2 = \frac{1}{N} \sum_{i=1}^N y_i^2 - \bar{y}^2,$$

Pri organizaciji B smo uporabili enačbo (*varianca iz razvrščenih vrednosti*):

$$\sigma^2 = \frac{1}{N} \sum_{k=1}^r f_k y_k^2 - \bar{y}^2$$

c) Izračun koeficienta variabilnosti:

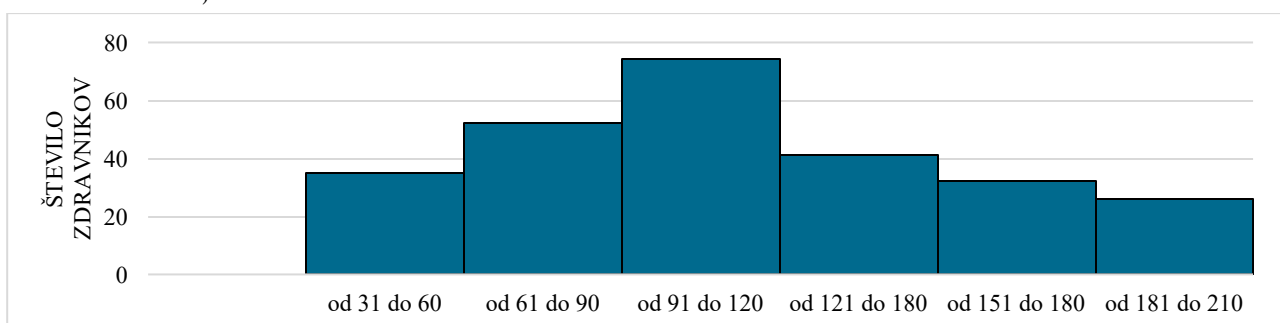
$$\text{Organizacija A:} \quad KV \% = \frac{6,87}{35} \cdot 100 = 19,6 \%$$

$$\text{Organizacija B:} \quad KV \% = \frac{5,22}{33,05} \cdot 100 = 15,8 \%$$

V organizaciji A se zaposleni glede na čas (v minutah), ki so ga porabili za izdelavo enega izdelka, med seboj bolj razlikujejo.

### Naloga 10 b)

Porazdelitev je asimetrična v desno stran.



**Naloga 10 c)**

$$\bar{y} = \frac{29260}{260} = 112,54 \text{ pregledov}$$

$$\sigma^2 = \frac{3815495}{260} - 112,54^2 = 2.010,08 \text{ pregledov}^2$$

$$\sigma = 44,83 \text{ pregledov}$$

$$\text{KV}\% = 39,83 \%$$

Delež standardnega odklona v aritmetični sredini znaša 39,83 %.

**Naloga 11 a)**

$$\text{VR} = 49 \text{ enot}$$

**Naloga 11 c)**

$$\bar{y} = 23,06 \text{ enot}$$

**Naloga 11 d)**

$$\sigma = 11,57 \text{ enot}$$

**Naloga 11 e)**

$$\text{KV}\% = 50,2 \%$$

**Naloga 12 b)**

Podatke uredimo v ranžirno vrsto.

$R_i$	1	2	3	4	5	6	7
$v_i$	75	87	150	230	380	450	500

Izračun vrednosti naložb za 25 % startup podjetij z najnižjo vrednostjo naložb (računamo  $Q_1 = 25\%$ ):

$$P_i = 0,25$$

$$R_i = 7 \cdot 0,25 + 0,5 = 2,3$$

$$R_0 \leq R_i < R_1$$

$$R_0 = 2 \leq R_i = 2,3 < R_1 = 3$$

$$y_0 \leq y_i < y_1$$

$$y_0 = 87 \leq y_i < y_1 = 150$$

Nato izračunamo vrednost  $y_i$ :

$$y_i = 87 + \frac{2,3 - 2}{3 - 2} \cdot (150 - 87) = 105,9 \text{ d.e.}$$

Odgovor: 25 % startup podjetij je imelo vrednost naložb 105,9 d.e. ali manj, 75 % startup podjetij pa več kot 105,9 d.e.

*Izračun vrednosti naložb za 25 % startup podjetij z najvišjo vrednostjo naložb (računamo  $Q_3 = 75$  %):*

$$P_i = 0,75$$

$$R_i = 7 \cdot 0,75 + 0,5 = 5,8$$

$$R_0 \leq R_i < R_1$$

$$R_0 = 5 \leq R_i = 5,8 < R_1 = 6$$

$$y_0 \leq y_i < y_1$$

$$y_0 = 380 \leq y_i < y_1 = 450$$

Nato izračunamo vrednost  $y_i$ :

$$y_i = 380 + \frac{5,8 - 5}{6 - 5} \cdot (450 - 380) = 436 \text{ d.e.}$$

Odgovor: 75 % startup podjetij je imelo vrednost naložb 436 d.e. ali manj, 25 % startup podjetij pa več kot 436 d.e.

### Naloga 12 c)

Kvartilni razmik:

$$Q = Q_3 - Q_1 = 436 - 105,9 = 330,1 \text{ d.e.}$$

Odgovor: 50 % startup podjetij, ki se glede na vrednosti naložb razvrščajo na sredino ranžirne vrste, se razlikuje za največ 330,1 d.e.

**Naloga 13 a)**

Ranžirna vrsta:

$R_i$	1	2	3	4	5	6	7	8	9
$y_i$	9	12	16	41	60	80	100	120	150

Izračun mediane ( $Me = 50 \%$ ):

$$P_i = 0,5$$

$$R_i = 9 \cdot 0,5 + 0,5 = 5$$

$$Me = 60 \text{ ton}$$

Mediana vrednosti zmanjšanja emisij CO<sub>2</sub> zaradi uporabe javnega prevoza s strani zaposlenih znaša 60 ton na leto. To pomeni, da polovica zaposlenih dosega zmanjšanje emisij CO<sub>2</sub> manjše ali enako 60 ton na leto, druga polovica zaposlenih pa večje od 60 ton na leto.

**Naloga 13 b)**

$$\text{Podjetje: } KV \% = \frac{47,90}{65,33} \cdot 100 = 73,3 \%$$

$$\text{Okoljevarstvena organizacija A: } KV \% = \frac{2650}{7530} \cdot 100 = 35,2 \%$$

V podjetju je zmanjšanje emisij CO<sub>2</sub> med posamezniki bolj raznoliko kot pa v okoljevarstveni organizaciji A.

**Naloga 14b)**

Ranžirna vrsta:

$R_i$	1	2	3	4	5	6	7	8	9
$y_i$	75	98	160	180	195	220	246	260	290

Enačba za koeficient asimetrije na podlagi mediane:

$$KA_{Me} = \frac{3(\bar{y} - Me)}{\sigma}$$



Mediana:

$$P_i = 0,5$$

$$R_i = N \cdot P_i + 0,5 = 9 \cdot 0,5 + 0,5 = 5$$

$$Me = 195 \text{ d.e.}$$

50 % vlagateljev je v enem mesecu vložilo v trajnostne finančne produkte 195 d.e. ali manj, 50 % vlagateljev pa več kot 195 d.e.

$$\bar{y} = \frac{1}{9} \cdot (75 + 98 + \dots + 290) = 191,56 \text{ d.e.}$$

$$VAR = \sigma^2 = \frac{1}{N} \sum_{i=1}^N y_i^2 - \bar{y}^2,$$

$$\sigma^2 = \frac{1}{9} \cdot [(75^2 + 98^2 + \dots + 290^2)] - 191,56^2 = \frac{1}{9} \cdot 371.870 - 191,56^2 = 4.623,66 \text{ (d.e.)}^2$$

$$\sigma = \sqrt{4.623,66} = 67,997 \text{ d.e.}$$

Izračun koeficienta asimetrije na podlagi mediane:

$$KA_{Me} = \frac{3(\bar{y} - Me)}{\sigma}$$

$$KA_{Me} = \frac{3(191,56 - 195)}{67,997} = -0,152$$

Porazdelitev je šibka in asimetrična v levo.

*Ponovitev:*

$\bar{y} > Me$  pomeni, da je porazdelitev asimetrična v desno (več kot polovica vrednosti je manjših od aritmetične sredine).

$\bar{y} < Me$  pomeni, da je porazdelitev asimetrična v levo (manj kot polovica vrednosti je manjših od aritmetične sredine).

$\bar{y} = Me$  pomeni, da je porazdelitev simetrična.

### Naloga 15 a)

Podatke uredimo v ranžirno vrsto:

$R_i$	1	2	3	4	5	6	7
$y_i$	2	4	7	10	12	16	18

$y_i$  je znan ( $y_i = 9$ ), zato pri izračunu uporabimo enačbe (*kvantilni rangi iz nerazvrščenih vrednosti*):

$$y_0 \leq y_i < y_1$$

$$R_i = R_0 + \frac{y_i - y_0}{y_1 - y_0} \times (R_1 - R_0)$$

$$P_i = \frac{R_i - 0,5}{N}$$

V ranžirni vrsti poiščemo položaj dane vrednosti:

$$y_i = 9$$

$$y_0 \leq y_i < y_1$$

$$y_0 = 7 \leq y_i = 9 < y_1 = 10$$

Iz vrednosti členov poiščemo ustrezajoče range:

$$R_0 \leq R_i < R_1$$

$$R_0 = 3 \leq R_i < R_1 = 4$$

Izračunamo rang  $R_i$ :

$$R_i = 3 + \frac{9 - 7}{10 - 7} \cdot (4 - 3) = 3,7$$

Izračunamo relativni rang  $P_i$ :

$$P_i = \frac{3,7 - 0,5}{7} = 0,46$$

Odgovor: 46 % investicijskih projektov je imelo donosnost 9 % ali manj.

### Naloga 15 b)

V ranžirni vrsti poiščemo položaj dane vrednosti:

$$y_i = 14$$

$$y_0 \leq y_i < y_1$$

$$y_0 = 12 \leq y_i = 14 < y_1 = 16$$

Iz vrednosti členov poiščemo ustrežajoče range:

$$R_0 \leq R_i < R_1$$

$$R_0 = 5 \leq R_i < R_1 = 6$$

Izračunamo rang  $R_i$ :

$$R_i = 5 + \frac{14 - 12}{16 - 12} \cdot (6 - 5) = 5,5$$

Izračunamo relativni rang  $P_i$ :

$$P_i = \frac{5,5 - 0,5}{7} = 0,71$$

Odgovor: 71 % investicijskih projektov je imelo donosnost 14 % ali manj, 29 % investicijskih projektov je imelo donosnost večje od 14 %.

### Naloga 15 c)

Izračun mediane ( $Me = 50\%$ ):

$$P_i = 0,5$$

$$R_i = 7 \cdot 0,5 + 0,5 = 4$$

Mediana: 10 %

Odgovor: 50 % investicijskih projektov v podjetju Y je imelo donosnost več kot 10 %.

### Naloga 15 d)

Uporabimo enačbo za decilni razmik:  $D = D_9 - D_1$  (kjer je  $D_9 = 90\%$  in  $D_1 = 10\%$ ), vendar je potrebno najprej izračunati:

Relativni rang  $P_i$  je znan:  $P_i = 0,9$  (kjer je  $D_9 = 90\%$ ), zato pri izračunu uporabimo enačbe (*kvantili iz nerazvrščenih vrednosti*):

$$R_i = N \times P_i + 0,5$$

$$R_0 \leq R_i < R_1$$

$$y_i = y_0 + \frac{R_i - R_0}{R_1 - R_0} \times (y_1 - y_0)$$

Izračunamo rang  $R_i$ :

$$R_i = 7 \cdot 0,9 + 0,5 = 6,8$$

Nato določimo vrednosti:

$$R_0 \leq R_i < R_1$$

$$R_0 = 6 \leq R_i = 6,8 < R_1 = 7$$

$$y_0 \leq y_i < y_1$$

$$y_0 = 16 \leq y_i < y_1 = 18$$

Nato izračunamo vrednost  $y_i$ :

$$y_i = 16 + \frac{6,8 - 6}{7 - 6} \cdot (18 - 16) = 17,6 \%$$

Odgovor: 90 % investicijskih projektov je imelo donosnost 17,6 % ali manj, 10 % investicijskih projektov pa več kot 17,6 %.

$$P_i = 0,1 \text{ (kjer je } D_1 = 10 \%)$$

$$R_i = 7 \cdot 0,1 + 0,5 = 1,2$$

$$R_0 \leq R_i < R_1$$

$$R_0 = 1 \leq R_i = 1,2 < R_1 = 2$$

$$y_0 \leq y_i < y_1$$

$$y_0 = 2 \leq y_i < y_1 = 4$$

Nato izračunamo vrednost  $y_i$ :

$$y_i = 2 + \frac{1,2 - 1}{2 - 1} \cdot (4 - 2) = 2,4 \%$$

Odg. 10 % investicijskih projektov je imelo donosnost 2,4 % ali manj, 90 % investicijskih projektov pa več kot 2,4 %.

Decilni razmik:

$$D = D_9 - D_1 = 17,6 - 2,4 = 15,2 \%$$

Odgovor: 80 % investicijskih projektov, ki glede na odstotek donosnosti ležijo na sredini ranžirne vrste, se razlikuje za največ 15,2 %.

### Naloga 15 e)

Za izračun povprečnega odstotka donosnosti uporabimo enačbo (*aritmetična sredina iz nerazvrščenih vrednosti*):

$$\bar{y} = \frac{1}{N}(y_1 + y_2 + \dots + y_N) = \frac{1}{N} \sum_{i=1}^N y_i$$

$$\bar{y} = \frac{1}{7} \cdot (2 + 4 + 7 + 10 + 12 + 16 + 18) = 9,9 \%$$

Odgovor: Povprečen odstotek donosnosti za sedem investicijskih projektov v podjetju Y znaša 9,9 %.

### Naloga 15 f)

Izračunamo varianco iz nerazvrščenih vrednosti po enačbi:

$$VAR = \sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2$$

$$\sigma^2 = \frac{1}{7} \cdot [(2 - 9,9)^2 + (4 - 9,9)^2 + (7 - 9,9)^2 + (10 - 9,9)^2 + \dots + (18 - 9,9)^2] = 30,41 \%^2$$

Izračunamo standardni odklon po enačbi:

$$SD = \sigma = \sqrt{VAR} = \sqrt{\sigma^2},$$

$$\sigma = \sqrt{30,41} = 5,5 \%$$

### Naloga 15 g)

Izračunamo koeficient variabilnosti v odstotku za oba podjetja po enačbi:

$$KV\% = \frac{\sigma}{\bar{y}} \times 100$$

Podjetje Y:

$$\bar{y} = 9,9 \%$$

$$\sigma = 5,5 \%$$

$$KV\% = \frac{5,5}{9,9} \cdot 100 = 55,56 \%$$

Odgovor: Delež standardnega odklona v aritmetični sredini znaša 55,56 %.

Podjetje Z:

$$\bar{y} = 12 \%$$

$$\sigma = 6,3 \%$$

$$KV\% = \frac{6,3}{12} \cdot 100 = 52,5 \%$$

Odgovor: V podjetju Y se investicijski projekti glede na odstotek donosnosti med seboj bolj razlikujejo.

### 4.3 Relativna števila: indeksna števila, povprečna vrednost v časovni vrsti

**Naloga 16a)**

Leto	$I_{t/2018}$	Koeficient dinamike $K_t$	$Y_t$
2018	100	/	200
2019	105	1,05	210
2020	95	0,90	189
2021	110	1,16	219,24
2022	120	1,09	238,97

Za koeficient dinamike uporabimo enačbo:

$$K_t = \frac{y_t}{y_{t-1}} \quad K_1 = / \quad T = 2, \dots, T$$

Na primer:

$$K_{t/2019} = \frac{105}{100} = 1,05$$

$$K_{t/2020} = \frac{95}{105} = 0,90$$

**Naloga 16b)**

Uporabimo enačbo za povprečni koeficient dinamike:

$$K = \sqrt[t-t_0]{\frac{I_{T/0}}{I_{I/0}}}$$

$$K = \sqrt[5-1]{\frac{120}{100}} = \sqrt[4]{1,2} = 1,047$$

Uporabimo enačbo za povprečno stopnjo rasti:

$$S = (K - 1) \cdot 100$$

$$S = (1,047 - 1) \cdot 100 = 4,7 \%$$

Odgovor: Povprečna letna stopnja rasti prodaje produkta v obdobju petih let znaša 4,7 %, kar pomeni, da je podjetje v povprečju dosegalo 4,7-odstotno letno rast prodaje svojega glavnega produkta v navedenem obdobju.

**Naloga 17 b), c)**

Leto	S <sub>t</sub> %	Verižni indeks (V <sub>t</sub> ) V <sub>t</sub> = S <sub>t</sub> + 100	Koeficient dinamike (K <sub>t</sub> ) K <sub>t</sub> = $\frac{S_t}{100} + 1$	I <sub>t/4</sub>	Y <sub>t</sub> (v GWh)
1	/	/	/	98,62	1183,44
2	+3,5	103,5	1,035	102,07	1224,86
3	+1,0	101	1,01	103,09	1237,11
4	-3,0	97	0,97	100	1200
5	+4,5	104,5	1,045	104,5	1254
6	-2,5	97,5	0,975	101,89	1222,65
7	+6,0	106	1,06	108,00	1296,01

**Naloga 18**

Leto	Y <sub>t</sub>
2018	1550
2024	890
2032	?

Izračunamo povprečni koeficient dinamike po enačbi:

$$K = \sqrt[t-t_0]{\frac{Y_t}{Y_1}}$$

$$K = \sqrt[5]{\frac{890}{1550}} = 0,895$$

Uporabimo enačbo za povprečno stopnjo rasti:

$$S = (K - 1) \cdot 100$$

$$S = (0,895 - 1) \cdot 100 = -10,5 \%$$

Ocena prodaje za leto 2032:

$$Y_{2032} = Y_{2024} \cdot K^8 = 890 \cdot 0,895^8 = 366,42 \text{ količinskih enot}$$

### Naloga 19a)

Verižni indeks za 5. časovno enoto, v %:

$$V_t = 100 \times \frac{Y_t}{Y_{t-1}} \quad \text{za } t = 2, 3, \dots, T$$

$$V_5 = \frac{300}{250} \cdot 100 = 120 \%$$

### Naloga 19b)

Indeksi s stalno osnovo iz časovnih vrst:

$$I_{t/0} = 100 \times \frac{Y_t}{Y_0} \quad \text{za } t = 1, 2, \dots, T$$

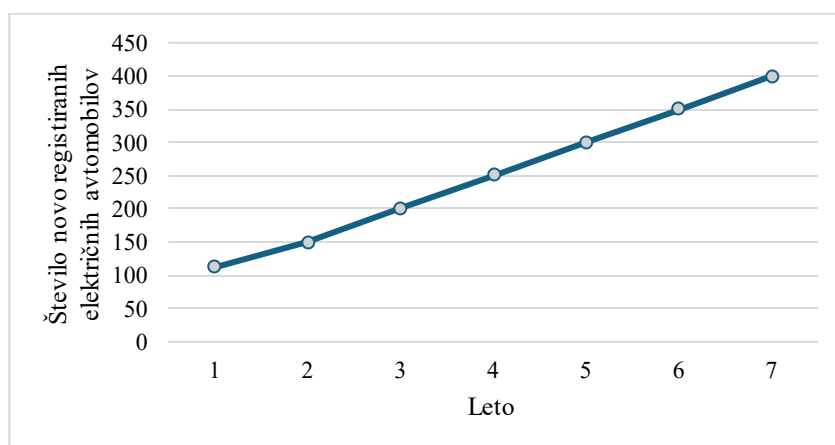
Indeks z osnovo 1, za 7. časovno enoto, v %:

$$V_{7/1} = \frac{400}{112} \cdot 100 = 357,14 \%$$



**Naloga 19c)**

Grafični prikaz:

**Naloga 20 a), b)**

Leto	St%	$K_t$	$I_{t/2}$ (v %)	$Y_t$ (v EUR)
1	/	/	92,25	183.279,98
2	+8,4	1,084	100	198.675,50
3	+5,7	1,057	105,7	210.000
4	-2,3	0,977	103,27	205.170

**Naloga 20 c)**

Uporabimo enačbo za povprečni koeficient dinamike:

$$K = \sqrt[t-1]{\frac{I_{T/o}}{I_{1/o}}}$$

$$K = \sqrt[3]{\frac{103,27}{92,25}} = 1,038$$

Nato izračunamo povprečno stopnjo rasti:

$$S = (K - 1) \cdot 100$$

$$S = (1,038 - 1) \cdot 100 = 3,8 \%$$

**Naloga 20 d)**

Napoved za 7. časovno enoto, v EUR:

$$Y_7 = Y_4 \cdot K^3 = 205.170 \cdot 1,038^3 = 229.459,43$$

**Naloga 21 a)**

Leto	$V_t$	$K_t = \frac{v_t}{100}$	$Y_t$
2021	/	/	1393,16
2022	130	1,3	1811,11
2023	90	0,9	1630
2024	110	1,1	1793
2025	125	1,25	2241,25

Povprečni koeficient dinamike:

$$K = \sqrt[4]{\frac{2241,25}{1393,16}} = 1,126$$

Povprečna stopnja rasti:

$$S = (1,126 - 1) \cdot 100 = 12,6 \%$$

**Naloga 21 b)**

Izračunati moramo povprečni koeficient dinamike po enačbi:  $K = \frac{V}{100}$

Povprečni verižni indeks:

$$V = \sqrt[T-1]{V_2 \cdot V_3 \cdot \dots \cdot V_T}$$

$$V = \sqrt[4]{130 \cdot 90 \cdot 110 \cdot 125} = 112,62$$

$$K = \frac{V}{100} = \frac{112,62}{100} = 1,1262$$

Napoved vrednost spremenljivke v letu 2028:

$$Y_{2028} = Y_{2025} \cdot K^3 = 2241,25 \cdot 1,1262^3 = 3201,38 \text{ enot}$$

**Naloga 22**

Leto	$V_t$	$K_t$	$Y_t$
2020	/	/	592,86
2021	70	0,7	415
2022	100	1	415
2023	85	0,85	352,75
2024	95	0,95	335,11
2025	110	1,1	368,62

**Naloga 23 b)**

Uporabimo enačbo za povprečni koeficient dinamike:

$$K = \sqrt[r]{\frac{Y_r}{Y_1}}$$

$$K = \sqrt[4]{\frac{16}{5}} = 1,337$$

$$S = (K - 1) \cdot 100 = (1,337 - 1) \cdot 100 = 33,7 \%$$

Število raziskovalnih projektov se je v petih letih povečevalo povprečno za 33,7 % na leto.

$$Y_8 = 16 \cdot K^3 = 16 \cdot 1,337^3 = 38,24 \text{ raziskovalnih projektov}$$

**Naloga 24**

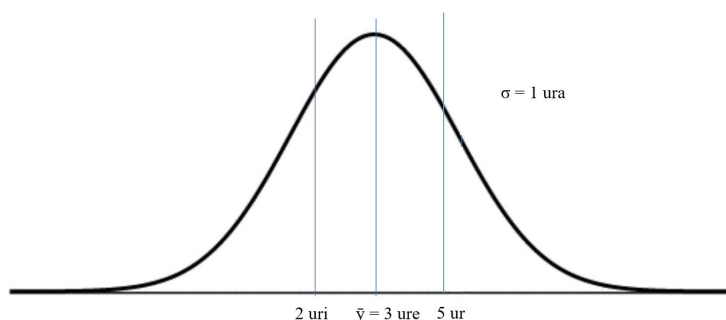
Mesec	I	II	III	IV	V
Vrednost proizvodnje v 10 <sup>6</sup> EUR	652	730	840	752	–
Štev. zaposlenih na začetku meseca	214	240	226	208	200
Štev. zaposlenih na sredini meseca	227	233	217	204	–

Povprečni statistični koeficient – povprečna vrednost proizvodnje na 10 zaposlenih – za opazovano obdobje je enak (v 10<sup>6</sup> EUR):

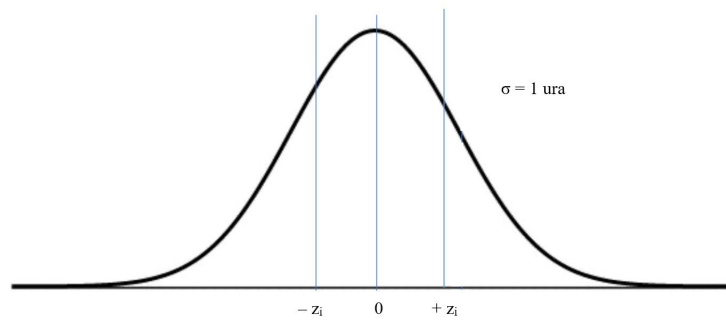
$$K = \frac{652 + 730 + 840 + 752}{227 + 233 + 217 + 204} \cdot 10 = 33,76$$

**4.4 Normalna porazdelitev****Naloga 36 a)**

Prikaz normalne porazdelitve:



Prikaz standardizirane normalne porazdelitve:



$$z_1 = \frac{y - \bar{y}}{\sigma} = \frac{2 - 3}{1} = -1$$

$$z_2 = \frac{5 - 3}{1} = 2$$

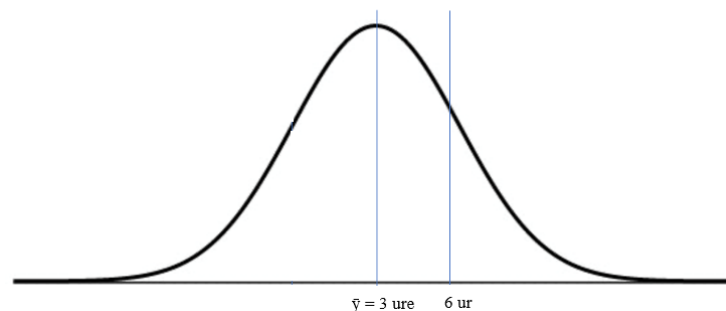
$$P(2 \text{ ure} < y < 5 \text{ ur}) = H(-1,0) + H(2,0) = 0,3413 + 0,4772 = 0,8185 = 81,85 \%$$

Vrednost  $H(z_i)$  odčitamo iz tabele *ploščine  $H(z)$  za standardizirano normalno porazdelitev*.

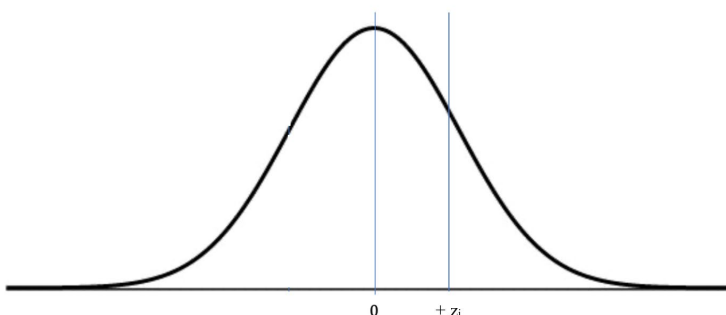
Verjetnost, da je v naključno izbranem gospodinjstvu družinski avtomobil v uporabi med 2 in 5 urami na dan, je enaka 81,85 %.

### Naloga 36 b)

Normalna porazdelitev:



Standardizirana normalna porazdelitev:



$$z_1 = \frac{y - \bar{y}}{\sigma} = \frac{6 - 3}{1} = 3$$

$$H(3) = 0,4987$$

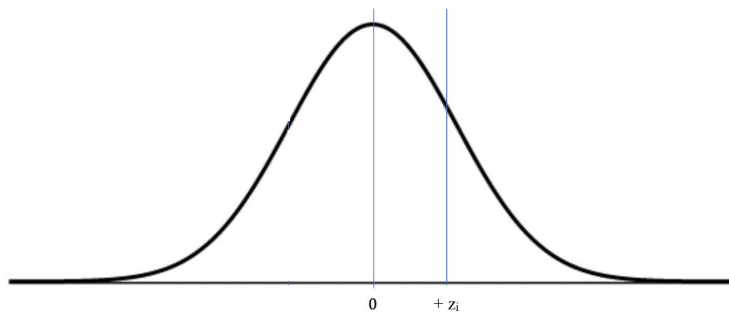
$$P(y > 6) = 0,5 - H(3) = 0,5 - 0,4987 = 0,0013 \text{ oz. } 0,13 \%$$

V 0,13 % gospodinjstev se družinski avtomobil uporablja več kot 6 ur dnevno.

### Naloga 36 c)

V tem primeru je podana verjetnost:

$$P(y > y_i) = 35 \%$$



$$H(z_i) = 0,15$$

Iz tabele *ploščine*  $H(z)$  za *standardizirano normalno porazdelitev* odčitamo vrednost  $z_i$ :

$$H(0,39) = 0,1517$$

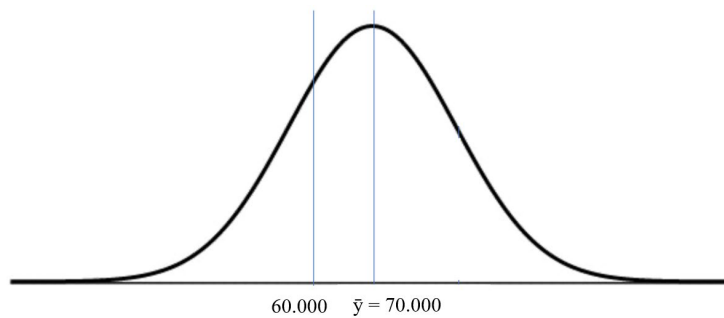
Uporabimo enačbo:  $y_i = z_i \cdot \sigma + \bar{y}$

$$y_i = 0,39 \cdot 1 + 3 = 3,39 \text{ ure}$$

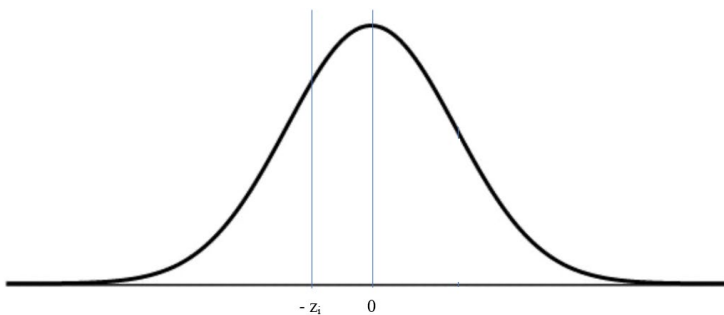
V 35 % gospodinjstev, kjer je čas uporabe družinskega avtomobila na dan najdaljši, je ta čas več kot 3,39 ure na dan.

**Naloga 37**

Normalna porazdelitev:



Standardizirana normalna porazdelitev:



$$z_i = \frac{y - \bar{y}}{\sigma} = \frac{60000 - 70000}{8000} = -1,25$$

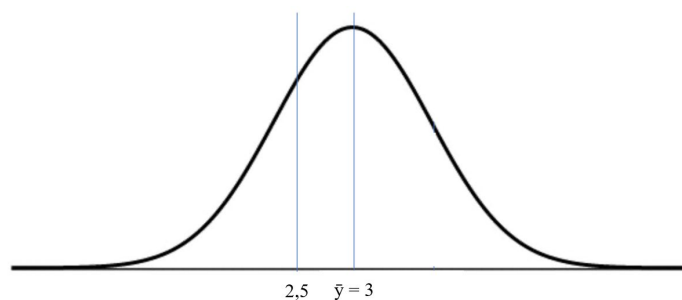
$$P(y > 60.000 \text{ eur}) = 0,5 + H(-1,25) = 0,5 + 0,3944 = 0,8944 = 89,44 \%$$

Vrednost  $H(z_i)$  odčitamo iz tabele *ploščine*  $H(z)$  za standardizirano normalno porazdelitev.

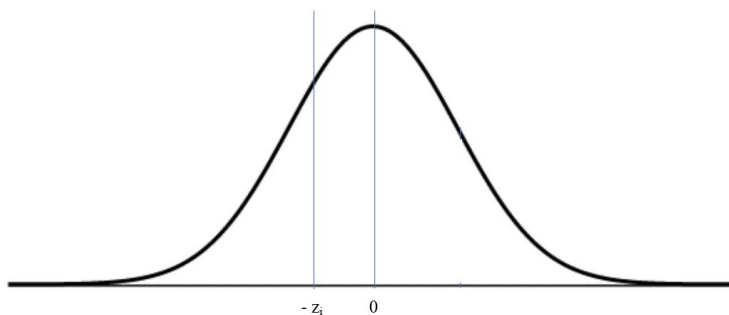
Verjetnost, da bo naključno izbrani samostojni podjetnik imel letni prihodek večji od 60.000 evrov, je 89,44 %.

**Naloga 38**

Normalna porazdelitev:



Standardizirana normalna porazdelitev:



$$z_i = \frac{y - \bar{y}}{\sigma} = \frac{2,5 - 3}{1,5} = -0,33$$

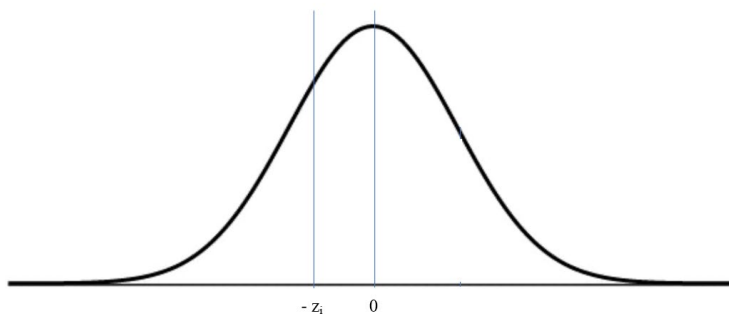
$$P(y < 2,5\%) = 0,5 - H(-0,33) = 0,5 - 0,1293 = 0,3707 = 37,07\%$$

Vrednost  $H(z_i)$  odčitamo iz tabele *ploščine  $H(z)$  za standardizirano normalno porazdelitev*.

Verjetnost, da bo letna rast BDP te države v določenem letu manjša od 2,5 %, je 37,07 %.

### Naloga 39

Standardizirana normalna porazdelitev:



$$P(y < y_i) = 35\%$$

$$H(z_i) = 0,15$$

Vrednost  $H(z_i)$  poiščemo v tabeli *ploščine  $H(z)$  za standardizirano normalno porazdelitev*, kjer poiščemo ploščino, ki je po vrednosti najbližje (tj. 0,1517), ter odčitamo pripadajočo standardizirano vrednost v prvem stolpcu ter prvi vrstici tabele, kar pomeni, da je  $z_i = 0,49$ , ki ji dodamo negativni predznak, saj leži iskana vrednost na levo od aritmetične sredine.

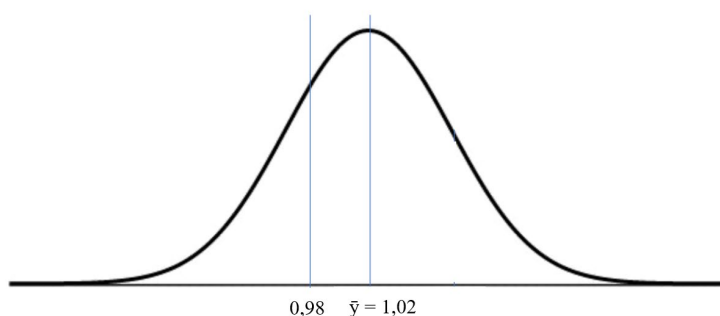
$$y_i = \bar{y} + z_i \cdot \sigma$$

$$y_i = 7 - 0,49 \cdot 1,2 = 6,4 \text{ ur}$$

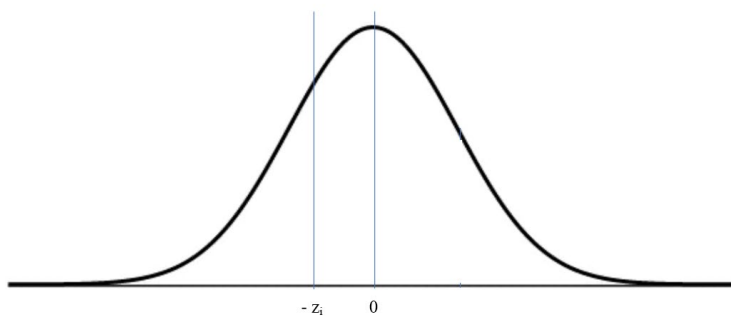
Posamezniki, ki spadajo med 35 % posameznikov z najkrajšim trajanjem spanja, so spali 6,4 ur.

### Naloga 40 a)

Normalna porazdelitev:



Standardizirana normalna porazdelitev:



$$z_i = \frac{y - \bar{y}}{\sigma} = \frac{0,98 - 1,02}{0,04} = -1$$

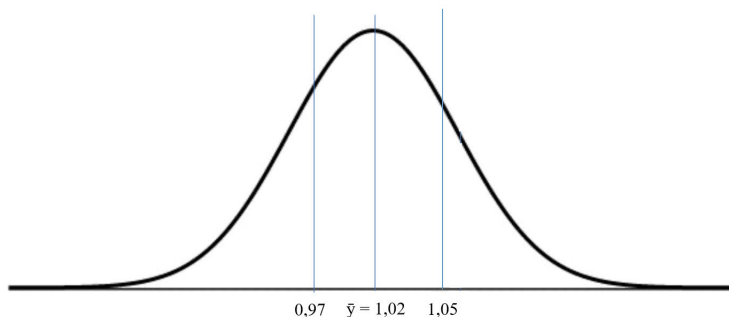
$$P(y \leq 0,98 \text{ sekund}) = 0,5 - H(-1) = 0,5 - 0,3413 = 0,1587 = 15,87 \%$$

Odgovor: Verjetnost, da bo naključno izbrani algoritem potreboval za odgovor največ 0,98 sekund je enaka 0,1587 oziroma 15,87 %.

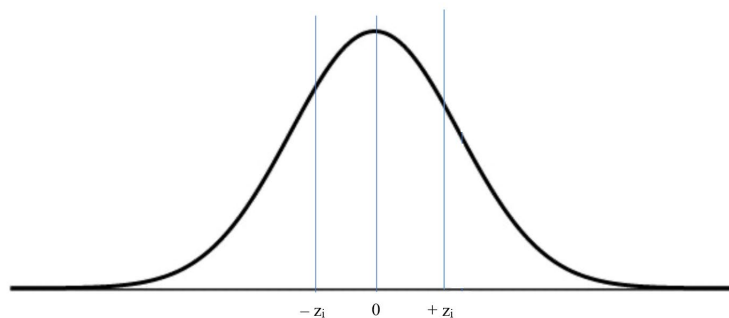


**Naloga 40b)**

Normalna porazdelitev:



Standardizirana normalna porazdelitev:



$$z_1 = \frac{y - \bar{y}}{\sigma} = \frac{0,97 - 1,02}{0,04} = -1,25$$

$$z_2 = \frac{y - \bar{y}}{\sigma} = \frac{1,05 - 1,02}{0,04} = 0,75$$

$$P(0,97 \text{ sekund} \leq y \leq 1,05 \text{ sekund}) = H(-1,25) + H(0,75) = 0,3944 + 0,2734 = 0,6678 = 66,78 \%$$

Odgovor: Verjetnost, da bo naključno izbrani algoritem potreboval za odgovor med 0,97 in 1,05 sekund, je 0,6678 oziroma 66,78 %.

**Naloga 40c)**

$$P(y_1 \leq y \leq y_2) = 50 \%$$

$$H(z_i) = \frac{0,50}{2} = 0,25$$

Vrednost  $H(z_i)$  odčitamo iz tabele *ploščine*  $H(z)$  za *standardizirano normalno porazdelitev*, kjer poiščemo najbližje število 0,25 (tj. 0,2486), kar pomeni, da je  $z_i = \pm 0,67$ .

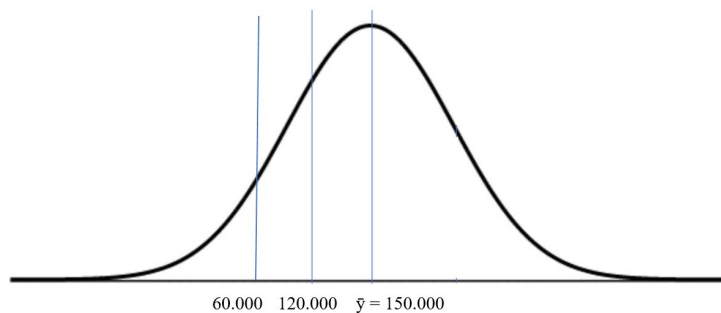
Uporabimo enačbo:  $y_i = \bar{y} + z_i \cdot \sigma$

$$y_1 = 1,02 - 0,67 \cdot 0,04 = 0,9932$$

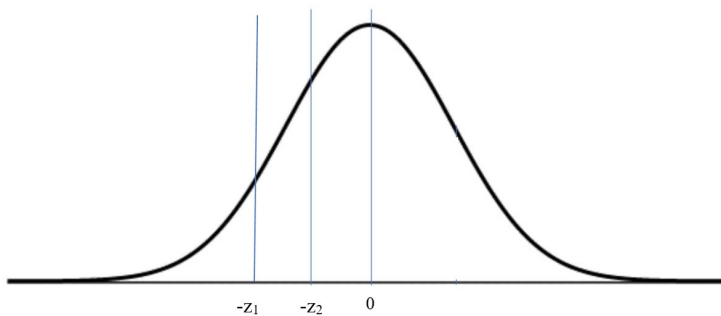
$$y_2 = 1,02 + 0,67 \cdot 0,04 = 1,0468$$

## Naloga 41

Normalna porazdelitev:



Standardizirana normalna porazdelitev:



$$z_1 = \frac{y - \bar{y}}{\sigma} = \frac{60.000 - 150.000}{40.000} = -2,25$$

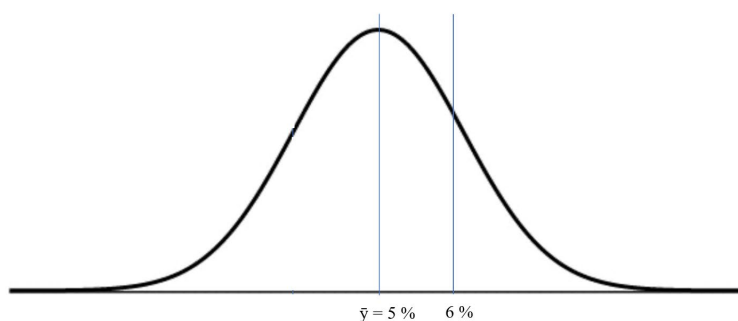
$$z_2 = \frac{y - \bar{y}}{\sigma} = \frac{120.000 - 150.000}{40.000} = -0,75$$

$$P(60.000\text{€} \leq y \leq 120.000\text{€}) = H(-2,25) - H(-0,75) = 0,4878 - 0,2734 = 0,2144 = 21,44\%$$

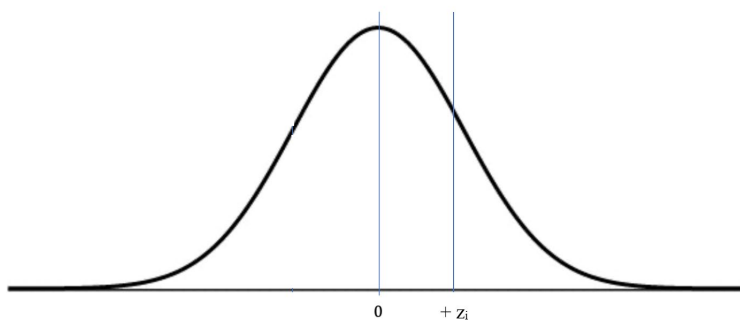
Verjetnost, da je prodajna cena naključno izbranega stanovanja med 60.000 in 120.000 evrov, je 21,44 %.

### Naloga 42

Normalna porazdelitev:



Standardizirana normalna porazdelitev:



$$z_i = \frac{y - \bar{y}}{\sigma} = \frac{6 - 5}{2} = 0,5$$

$$P(y > 6\%) = 0,5 - H(0,5) = 0,5 - 0,1915 = 0,3085 = 30,85\%$$

Vrednost  $H(z_i)$  odčitamo iz tabele *ploščine  $H(z)$  za standardizirano normalno porazdelitev*.

Verjetnost, da bo v podjetju X nov sistem napovedal naraščanje cen z uspešnostjo, ki je večja od 6 %, je 30,85 %.

### Naloga 43a)

$$P(y > y_i) = 40\%$$

$$H(z_i) = 0,1$$

Vrednost  $H(z_i)$  odčitamo iz tabele *ploščine*  $H(z)$  za *standardizirano normalno porazdelitev*, kjer poiščemo najbližje število 0,1 (tj. 0,0987), kar pomeni, da je  $z_i = 0,25$ .

Uporabimo enačbo:  $y_i = \bar{y} + z_i \cdot \sigma$

$$y_i = 90 + 0,25 \cdot 35 = 98,75 \text{ minut}$$

### Naloga 43b)

$$P(y < y_i) = 30 \%$$

$$H(z_i) = 0,2$$

Vrednost  $H(z_i)$  odčitamo iz tabele *ploščine*  $H(z)$  za *standardizirano normalno porazdelitev*, kjer poiščemo najbližje število 0,2 (tj. 0,1985), kar pomeni, da je  $z_i = -0,52$ .

$$y_i = \bar{y} + z_i \cdot \sigma$$

$$y_i = 90 - 0,52 \cdot 35 = 71,8 \text{ minut}$$

### Naloga 43c)

$$z_1 = \frac{y - \bar{y}}{\sigma} = \frac{50 - 90}{35} = -1,14$$

$$z_2 = \frac{105 - 90}{35} = 0,43$$

$$P(50 \text{ minut} < y < 105 \text{ minut}) = H(-1,14) + H(0,43) = 0,3729 + 0,1664 = 0,5393 = 53,93 \%$$

### Naloga 43d)

$$z_1 = \frac{y - \bar{y}}{\sigma} = \frac{100 - 90}{35} = 0,29$$

$$z_2 = \frac{130 - 90}{35} = 1,14$$

$$P(100 \text{ minut} < y < 130 \text{ minut}) = H(z_2) - H(z_1) = H(1,14) - H(0,29) = 0,3729 - 0,1141 = 0,2588 = 25,88 \%$$

## 4.5 Osnove vzorčenja in osnove preizkušanja domnev

### Naloga 47a)

Porazdelitev parametrov izračunanih iz vseh možnih kombinacij enot v vzorcih (vseh možnih vzorcev):

$$K_n^N = \binom{N}{n} = \frac{N!}{(N-n)!n!}$$

$$K = \frac{N}{n} = \frac{4!}{(4-2)! \cdot 2!} = 6 \quad (\text{število vseh možnih vzorcev})$$

### Naloga 47b)

4 študenti: A, B, C, D    Vzorec: 2 študenta    Ocene: A = 3, B = 5, C = 4, D = 2

Vzorec	Enote v vzorcu	Vrednosti v vzorcu	Vzorčna aritmetična sredina ( $\bar{y}$ )
1	A, B	3; 5	$\frac{3+5}{2} = 4$
2	A, C	3; 4	$\frac{3+4}{2} = 3,5$
3	A, D	3; 2	$\frac{3+2}{2} = 2,5$
4	B, C	5; 4	$\frac{5+4}{2} = 4,5$
5	B, D	5; 2	$\frac{5+2}{2} = 3,5$
6	C, D	4; 2	$\frac{4+2}{2} = 3$
			Skupaj = 21

### Naloga 47c)

Aritmetična sredina iz vseh vzorčnih ocen aritmetične sredine:  $\bar{y} = \frac{21}{6} = 3,5$

Aritmetična sredina iz podatkov v statistični množici (A, B, C, D):  $\frac{3+5+4+2}{4} = 3,5$

Najpomembnejša značilnost porazdelitve vzorčnih ocen aritmetične sredine je, da je aritmetična sredina vseh vzorčnih ocen aritmetične sredine iz vseh možnih vzorcev **enaka** aritmetični sredini iz statistične množice.

**Naloga 47e)**

Vzorčne variance ( $\sigma^2$ ) za vse možne vzorce izračunamo po enačbi:

$$VAR = \sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2$$

Rešitev za vsak vzorec posebej:

1. Vzorec:  $\sigma^2 = 1$
2. Vzorec:  $\sigma^2 = 0,25$
3. Vzorec:  $\sigma^2 = 0,25$
4. Vzorec:  $\sigma^2 = 0,25$
5. Vzorec:  $\sigma^2 = 2,25$
6. Vzorec:  $\sigma^2 = 1$

Aritmetična sredina iz vseh izračunanih vrednosti:  $\frac{5}{6} = 0,833$

Nepriustranske ocene vzorčnih varianc ( $s^2$ ) za vse možne vzorce izračunamo po enačbi:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{Y})^2$$

Rešitev za vsak vzorec posebej:

1. Vzorec:  $s^2 = 2$
2. Vzorec:  $s^2 = 0,5$
3. Vzorec:  $s^2 = 0,5$
4. Vzorec:  $s^2 = 0,5$
5. Vzorec:  $s^2 = 4,5$
6. Vzorec:  $s^2 = 2$

Aritmetična sredina iz nepriustranskih ocen variance vseh možnih vzorcev:  $\frac{10}{6} = 1,667$

Izračun  $\sigma^2$  ter  $s^2$  iz osnovnih štirih podatkov:

$$\sigma^2 = \frac{1}{4} [(3 - 3,5)^2 + (5 - 3,5)^2 + (4 - 3,5)^2 + (2 - 3,5)^2] = 1,25$$

$$s^2 = \frac{1}{3} \cdot 5 = 1,667$$

Aritmetična sredina izračunana iz nepristranskih ocen variance vseh možnih vzorcev je enaka nepristranski oceni variance, izračunane na osnovi podatkov iz celotne množice.

**Naloga 48 a)**

$N = 200$  zaposlenih

$n = 5$  zaposlenih

Vzorčni delež  $= \frac{n}{N} = \frac{5}{200} = \frac{1}{40}$ , kar pomeni, da v vzorec izberemo vsakega 40-tega zaposlenega iz statistične množice.

Prvo izbrano število je slučajno izbrano število od 1 do 40 (npr. izbrali smo št. 10):

Prvo izbrano št.: 10

Drugo izbrano št.:  $10 + 40 = 50$

Tretjo izbrano št.:  $50 + 40 = 90$

Četrto izbrano št.:  $90 + 40 = 130$

Peto izbrano št.:  $130 + 40 = 170$

**Naloga 48 b)**

$N = 200$  zaposlenih

$n = 10$  zaposlenih

Vzorčni delež  $= \frac{n}{N} = \frac{10}{200} = \frac{1}{20}$ , kar pomeni, da v vzorec vključimo vsakega 20-tega zaposlenega iz statistične množice.

Prvo izbrano število je slučajno izbrano število od 1 do 20 (npr. izbrali smo št. 8):

Prvo izbrano št.: 8, nato si sledijo 28 ( $8 + 20$ ), 48 ( $28 + 20$ ), 68, 88, 108, 128, 148, 168, 188.

**Naloga 49 a), b)**

Ponovitev korakov pri stratificiranem vzorčenju:

Korak 1	Opredelevanje populacije
Korak 2	Določitev značilnosti, na osnovi katere opredelimo stratum (k stratumov)
Korak 3	Razvrstitev vseh elementov po stratumih
Korak 4	Določitev zaporedne številke elementov v vsakem stratumu posebej
Korak 5	Določitev velikosti vzorca (n)
Korak 6	Izbira proporcionalne ali disproporcionalne stratifikacije



<i>Disproporcionalna stratifikacija</i>	<i>Proporcionalna stratifikacija</i>
Korak 7: Določitev števila elementov, ki jih izberemo iz vsakega stratuma in je enako za vsak stratum: velikost vzorca (n) / število stratumov (k)	Korak 7: Določitev deleža populacije, ki je zajeta v posameznem stratumu ( $p_i$ ), za $i$ -ti stratum ( $i=1,2,\dots,k$ ) in sicer: Število elementov v stratumu $i$ / velikost populacije (N)
Korak 8: Izbira izračunanega števila elementov iz vsakega stratuma s tehniko slučajnega vzorčenja	Korak 8: Določitev števila elementov, ki jih izberemo iz vsakega stratuma in je za $i$ -ti stratum enako: velikost vzorca (n) $\cdot p_i$
	Korak 9: Izbira izračunanega števila elementov iz vsakega stratuma s tehniko slučajnega vzorčenja

### 1. KORAK:

Opredelevanje populacije:  $N = 1.500$  zaposlenih oseb

Statistična enota ali element: 1 zaposleni

### 2. KORAK:

Opredelevanje stratumov:

$k = 3$  stratumi:

A: osebni dohodek do pod 800 €

B: osebni dohodek od 800 € do pod 1500 €

C: osebni dohodek 1500 € ali več

### 3. in 4. KORAK:

Razvrstitev elementov po stratumih:

A: zaporedne št. od 1 do 700 (podano v nalogi – 700 zaposlenih)

B: zaporedne št. od 1 do 500 (podano v nalogi – 500 zaposlenih)

C: zaporedne št. od 1 do 300 (podano v nalogi – 300 zaposlenih)

(Populacija:  $700 + 500 + 300 = 1500$  zaposlenih)



## 5. KORAK

Določitev velikosti vzorca:  $n = 200$  (podano v nalogi: skupno število statističnih enot v vzorcu naj bo  $n = 200$ )

**6. KORAK: Proporcionalno stratificirano vzorčenje (točka a)**

7. KORAK: Določitev deleža populacije, ki je zajeta v posameznem stratumu

$$p = \frac{n}{N}$$

$$p_A = \frac{700}{1500} = 0,467 \quad p_B = \frac{500}{1500} = 0,333 \quad p_C = \frac{300}{1500} = 0,200$$

8. KORAK: Določitev števila elementov, ki jih izberemo iz posameznega stratuma

$$n_A = 0,467 \cdot 200 = 93,4 = 93 \text{ (93 oseb)} \quad n_B = 0,333 \cdot 200 = 66,6 = 67 \quad n_C = 0,200 \cdot 200 = 40$$

9. KORAK: Izbira izračunanega števila elementov iz vsakega stratuma s tehniko slučajnega vzorčenja – npr. s sistematičnim vzorčenjem)

$$\text{Vzorčni delež} = \frac{n}{N}$$

A: vzorčni delež =  $\frac{93}{700} = \frac{1}{7,52} = \frac{1}{8}$  (izberemo vsakega 8. zaposlenega) → npr. prvo izbrano št. je 2., nato 10, 18, 26, 34 itd.

B: vzorčni delež =  $\frac{67}{500} = \frac{1}{7,46} = \frac{1}{7}$  (izberemo vsakega 7. zaposlenega) → npr. prvo izbrano št. je 4., nato 11, 18, 25, 32 itd.

C: vzorčni delež =  $\frac{40}{300} = \frac{1}{7,5} = \frac{1}{8}$  (izberemo vsakega 8. zaposlenega)

**6. KORAK: DISPROPORCIONALNA STRATIFIKACIJA (točka b)**

7. KORAK: Določitev deleža populacije ki je zajeta v stratum:

$$\text{Število enot iz vsakega stratuma} = \frac{\text{velikost vzorca } (n)}{\text{število stratumov } (k)} = \frac{200}{3} = 67$$

8. KORAK: Izbira elementov iz vsakega stratuma s tehniko sistematičnega vzorčenja:

$$A: \text{vzorčni delež} = \frac{67}{700} = \frac{1}{10,4} = \frac{1}{10} \text{ (izberemo vsakega 10. zaposlenega)}$$

$$B: \text{vzorčni delež} = \frac{67}{500} = \frac{1}{7,46} = \frac{1}{7} \text{ (izberemo vsakega 7. zaposlenega)}$$

$$C: \text{vzorčni delež} = \frac{67}{300} = \frac{1}{4,48} = \frac{1}{4} \text{ (izberemo vsakega 4. zaposlenega)}$$

### Naloga 50a)

Imamo velik vzorec, ker je  $n = 300$ , ter dvostransko intervalno ocenjevanje aritmetične sredine:

$$\gamma = 95 \%, \alpha = 5 \%, z = \pm 1,96$$

Izračunamo standardno napako ocene aritmetične sredine:

$$SE_{\bar{y}} = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$$

$$se_{\bar{y}} = \frac{8,6}{\sqrt{300}} = 0,497$$

$$P(\bar{Y} - z \cdot se_{\bar{y}} < \bar{y} < \bar{Y} + z \cdot se_{\bar{y}}) = \gamma$$

$$P(180 - 1,96 \cdot 0,497 < \bar{y} < 180 + 1,96 \cdot 0,497) = 95 \%$$

$$P(179,026 < \bar{y} < 180,974) = 95 \%$$

S 95-odstotno verjetnostjo ocenjujemo, da je povprečni čas pretečenega maratona v osnovni statistični množici med 179,026 in 180,974 minutami.

### Naloga 50b)

Enostransko intervalno ocenjevanje aritmetične sredine iz velikega vzorca s spodnjo mejo:

$$P(\bar{y} > \bar{Y} - z \cdot se_{\bar{y}}) = \gamma$$

$$\alpha = 1\%, z = -2,33$$

$$P(\bar{y} > 223,2 - 2,33 \cdot 12,819) = 99 \%$$

$$P(\bar{y} > 193,33) = 99 \%$$

Enostransko intervalno ocenjevanje aritmetične sredine iz malega vzorca z zgornjo mejo:

$$P(\bar{y} < \bar{Y} + z \cdot se_{\bar{y}}) = \gamma$$

$$\alpha = 1\%, z = 2,33$$

$$P(\bar{y} < 223,2 + 2,33 \cdot 12,819) = 99\%$$

$$P(\bar{y} < 253,07) = 99\%$$

### Naloga 50c)

Izračun 90-odstotnega intervala zaupanja za povprečni čas pretečenega maratona v osnovni statistični množici, če smo v slučajni vzorec zajeli  $n = 25$  tekačev:

Imamo mali vzorec, ker je  $n = 25$  tekačev, ter dvostransko intervalno ocenjevanje aritmetične sredine:

$$\alpha = 10\% \rightarrow t_{n-1; \alpha/2} = t_{24; 0,05} = \pm 1,711$$

Izračunamo standardno napako ocene aritmetične sredine:

$$se_{\bar{y}} = \frac{8,6}{\sqrt{25}} = 1,72$$

$$P(\bar{Y} - t_{n-1; \alpha/2} \cdot se_{\bar{y}} < \bar{y} < \bar{Y} + t_{n-1; \alpha/2} \cdot se_{\bar{y}}) = \gamma$$

$$P(180 - 1,711 \cdot 1,72 < \bar{y} < 180 + 1,711 \cdot 1,72) = 90\%$$

$$P(177,06 < \bar{y} < 182,94) = 90\%$$

Z 90-odstotno verjetnostjo ocenjujemo, da je povprečni čas pretečenega maratona v osnovni statistični množici med 177,06 in 182,94 minutami.

### Naloga 51

Imamo mali vzorec, ker je  $n = 6$ , ter dvostransko intervalno ocenjevanje aritmetične sredine:

$$\alpha = 5\%$$

$$\bar{Y} = 32,83$$

$$s^2 = 537,77$$

$$s = 23,19$$

$$se_{\bar{y}} = 9,467$$

$$t_{n-1; \alpha/2} = t_{5; 0,025} = 2,571 \text{ (gledamo tabelo kritična vrednosti za } t \text{ porazdelitev)}$$

$$\alpha/2 = 0,05/2 = 0,025$$

$$P(32,83 - 2,571 \cdot 9,467 < \bar{y} < 32,83 + 2,571 \cdot 9,467) = 95 \%$$

$$P(8,490 < \bar{y} < 57,169) = 95 \%$$

S 95-odstotno verjetnostjo ocenjujemo, da povprečno število knjig, izposojenih na dijaka v srednjih šolah, med 8,490 in 57,169 knjig.

## Naloga 52

$n = 700$  članov (velik vzorec)

$n_a = 235$  članov

$\alpha = 10 \%$   $\rightarrow z = \pm 1,645$

Izračun strukturnega odstotka iz vzorca:

$$p = 100 \frac{n_a}{n}$$

$$p = \frac{235}{700} \cdot 100 = 33,57 \%$$

Izračun standardne napake ocene strukturnega odstotka:

$$SE_{\pi} = \sqrt{\frac{p(100-p)}{n}}$$

$$se_{\pi} = \sqrt{\frac{33,57(100-33,57)}{700}} = 1,78$$

$$P(p - z \cdot se_{\pi} < \pi < p + z \cdot se_{\pi}) = \gamma$$

$$P(33,57 - 1,645 \cdot 1,78 < \pi < 33,57 + 1,645 \cdot 1,78) = 90 \%$$

$$P(30,64 < \pi < 36,50) = 90 \%$$

Z 90-odstotno verjetnostjo ocenjujemo, da je odstotek članov, ki aktivno izposoja knjige v mestni knjižnici, med 30,64 % in 36,50 %.

**Naloga 53**

Imamo velik vzorec, ker je  $n = 45$  startup podjetji, ter dvostransko intervalno ocenjevanje aritmetične sredine:

$$\gamma = 99 \%, \alpha = 1\%, z = \pm 2,58$$

$$\bar{Y} = 100.000 \text{ d.e.}$$

$$s = 890 \text{ d.e.}$$

$$se_{\bar{y}} = 132,67 \text{ d.e.}$$

$$P(100.000 - 2,58 \cdot 132,67 < \bar{y} < 100.000 + 2,58 \cdot 132,67) = 99 \%$$

$$P(99.657,71 < \bar{y} < 100.342,29) = 99 \%$$

Z 99-odstotno verjetnostjo ocenjujemo, da je povprečni letni dobiček startup podjetja v statistični množici med 99.657,71 d.e. in 100.342,29 d.e.

*Izračun 95-odstotnega intervala zaupanja za povprečni letni dobiček startup podjetja v osnovni statistični množici, če smo v slučajni vzorec zajeli  $n = 15$  startup podjetij:*

Imamo mali vzorec, ker je  $n = 15$  startup podjetji, ter dvostransko intervalno ocenjevanje aritmetične sredine:

$$se_{\bar{y}} = 229,80 \text{ d.e.}$$

$$\alpha = 5 \%$$

$$t_{n-1; \alpha/2} = t_{14; 0,025} = 2,145 \text{ (gledamo tabelo kritična vrednost za } t \text{ porazdelitev)}$$

$$P(100.000 - 2,145 \cdot 132,67 < \bar{y} < 100.000 + 2,145 \cdot 132,67) = 95 \%$$

$$P(99.715,422 < \bar{y} < 100.284,58) = 95 \%$$

Z 95-odstotno verjetnostjo ocenjujemo, da je povprečni letni dobiček startup podjetja v statistični množici med 99.715,422 d.e. in 100.284,58 d.e.

**Naloga 54**

$$H_0: \bar{y}_D = 5.000 \text{ kg}$$

$$H_1: \bar{y}_D \neq 5.000 \text{ kg}$$

$$\bar{y} = 5.200$$

$$s^2 = 800.000$$

$$s = 894,43$$

$$se_{\bar{y}} = 149,07$$

$\alpha = 10\%$  (oziroma 0,10), kar pomeni, da je kritična vrednost spremenljivke  $z = \pm 1,645$

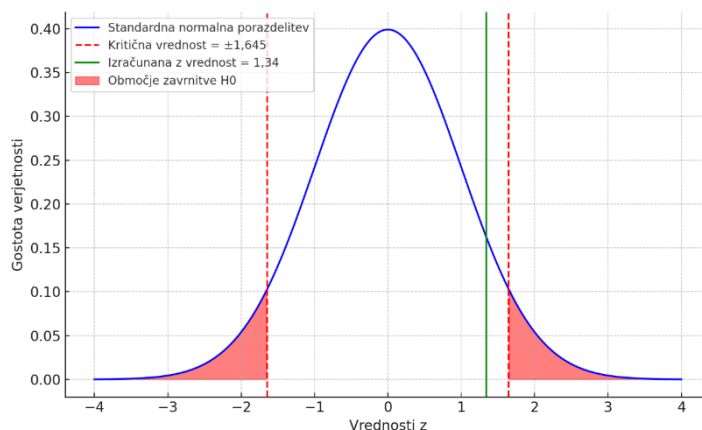
Izračunamo  $z$  po enačbi (izračun testne vrednosti pri preizkušanju domneve o aritmetični sredini – z-test):

$$z = \frac{\bar{Y} - \bar{y}_D}{SE_{\bar{y}}}$$

$$z = \frac{5200 - 5000}{149,07} = 1,34 \text{ (vidimo, da dobljeno število pade v interval } \pm 1,645\text{)}$$

Sprejmemo domnevo  $H_0$ .

Grafični prikaz dvostranskega preizkušanja domneve (z-test):



Na grafu so z rdečo črtkano črto označene kritične vrednosti  $z = \pm 1,645$  pri stopnji tveganja  $\alpha = 0,10$ . Te vrednosti določajo meje, znotraj katerih bi sprejeli ničelno hipotezo  $H_0$ , da je povprečna količina recikliranih materialov enaka 5.000 kg ( $H_0: \bar{y}_D = 5.000 \text{ kg}$ )

Izračunana z vrednost  $z_i$ , ki znaša 1,34 in je prikazana z zeleno črto, se nahaja znotraj teh kritičnih meja, kar pomeni, da ne moremo zavrniti ničelne hipoteze na osnovi naših podatkov.

Območje zavrnitve ničelne hipoteze, ki bi nakazovalo, da je povprečna količina recikliranih materialov statistično značilno različna od 5.000 kg, je označeno z rdečo barvo in se nahaja izven kritičnih vrednosti na obeh koncih porazdelitve. Ker se naša izračunana z vrednost ne nahaja v tem območju, tega zaključka ne moremo narediti iz zavrnemo hipotezo  $H_1: \bar{y}_D \neq 5.000 \text{ kg}$ .

**Naloga 55**

$$n = 50$$

$$\bar{Y} = 102$$

$$s^2 = 6,25 \rightarrow s = 2,5$$

$$se_{\bar{y}} = \frac{2,5}{\sqrt{50}} = 0,35$$

$$\alpha = 5\% \rightarrow z = \pm 1,96$$

Zastavimo ničelno in raziskovalno domnevo:

$$H_0: \bar{y}_D = 100 \text{ EUR}$$

$$H_1: \bar{y}_D \neq 100 \text{ EUR}$$

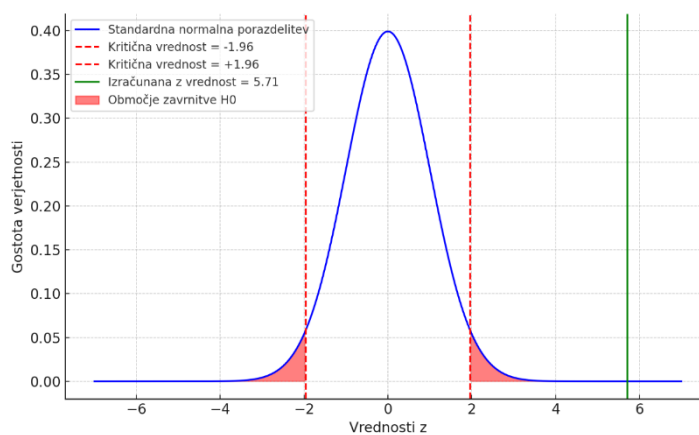
Izračunamo  $z$  po enačbi (izračun testne vrednosti pri preizkušanju domneve o aritmetični sredini –  $z$ -test):

$$z = \frac{\bar{Y} - \bar{y}_D}{SE_{\bar{y}}}$$

$$z_i = \frac{102 - 100}{0,35} = 5,71 \text{ (vidimo, da dobljeno število ne pade v interval } \pm 1,96\text{)}$$

Na osnovi rezultata vidimo, da delavci ne izdelajo povprečno 100 izdelkov v določeni časovni enoti, tako da zavrne domnevo  $H_0$  in sprejmemo raziskovalno domnevo  $H_1$ .

Grafični prikaz dvostranskega preizkušanja domneve ( $z$ -test):



Graf prikazuje standardno normalno porazdelitev, ki je osnova za dvostransko preizkušanje domneve. Os x predstavlja z vrednosti, ki segajo od negativnih do pozitivnih, s središčem pri vrednosti 0, kar predstavlja povprečje standardne normalne porazdelitve. Os y prikazuje gostoto verjetnosti za z-vrednosti.

Na grafu so z rdečo črtkano črto označene kritične vrednosti  $z = \pm 1,96$ , ki sovpadajo s stopnjo tveganja 5 % ( $\alpha = 0,05$ ). Te vrednosti določajo območje zavrnitve ničelne hipoteze ( $H_0$ ), ki predpostavlja, da je povprečni dnevni zaslužek vlagateljev točno 100 €. Izračunana z vrednost, ki znaša 5,71 in je prikazana z zeleno črto, se nahaja daleč desno od pozitivne kritične vrednosti. Območje zavrnitve ničelne domneve je poudarjeno z rdečo barvo in se nahaja izven vrednosti  $\pm 1,96$ .

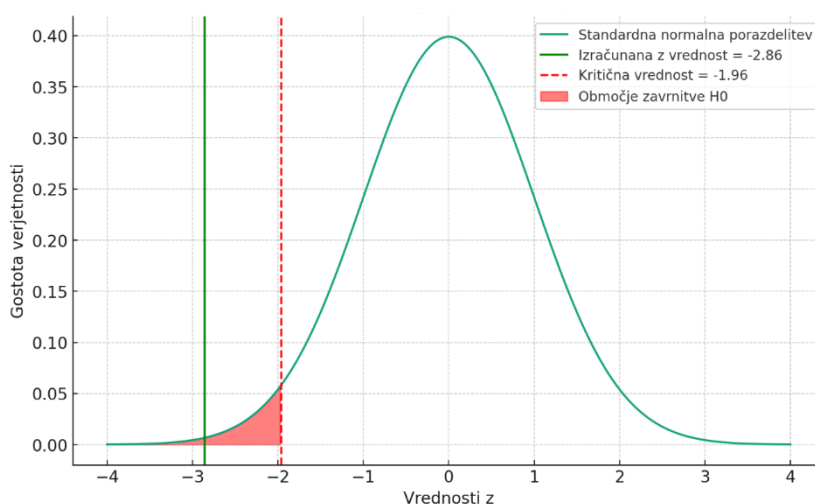
Preizkus domneve, da 50 naključno izbranih vlagateljev v povprečju zasluži manj kot 103 d.e. na dan:

Zastavimo ničelno in raziskovalno domnevo:

$H_0: \bar{y}_D \leq 103$  d.e.

$H_1: \bar{y}_D > 103$  d.e.

Grafični prikaz enostranskega preizkušanja domneve (z-test):



Izračunamo z po enačbi (izračun testne vrednosti pri preizkušanju domneve o aritmetični sredini – z-test):

$$z = \frac{\bar{Y} - \bar{y}_D}{SE_{\bar{y}}}$$



$$z_i = \frac{102 - 103}{0,35} = -2,86 \text{ (vidimo, da dobljeno število ne pade v interval } \pm 1,96)$$

Sprejmemo domnevo H1.

### Naloga 56a)

$n = 25$  študentov (mali vzorec)

$\bar{y} = 30$  knjig

$s = 5$  knjig

$$se_{\bar{y}} = \frac{s}{\sqrt{n}} = \frac{5}{\sqrt{25}} = 1$$

Zastavimo ničelno in raziskovalno domnevo:

H0:  $\bar{y}_D \geq 28$  knjig (povprečno število prebranih knjig na študenta je 28 ali več).

H1:  $\bar{y}_D < 28$  knjig (povprečno število prebranih knjig na študenta je manj kot 28).

Enostransko preizkušanje domneve za mali vzorec:

$\alpha = 10\% \rightarrow t_{n-1; \alpha} = t_{24; 0,1} = +1,318$  (kritična vrednost spremenljivke t.)

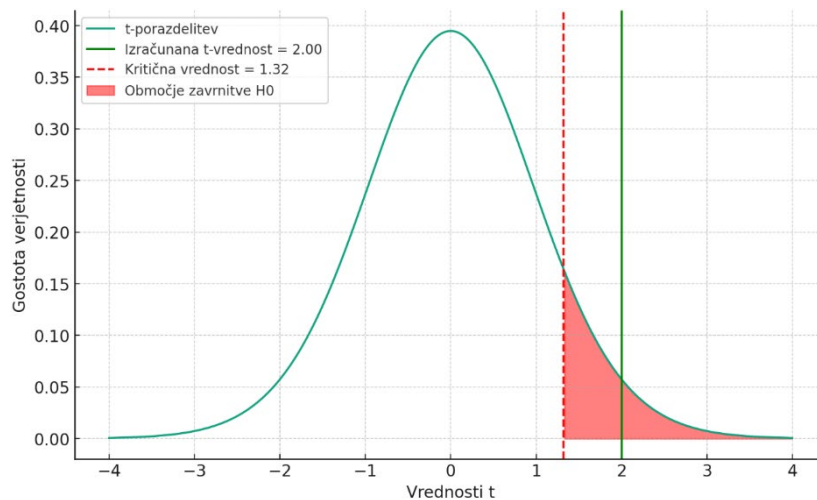
$$t = \frac{\bar{Y} - \bar{y}_D}{SE_{\bar{y}}}$$

$$t_i = \frac{30 - 28}{1} = 2$$

Ker je izračunana t-vrednost  $t_i=2$  večja od kritične vrednosti  $t_{24;0,1} = +1,318$ , to pomeni, da sprejmemo hipotezo H1.

Zelena črta na grafu prikazuje izračunano t-vrednost, ki je bila 2 in se nahaja desno od središča porazdelitve, kar kaže na pozitivno odstopanje od predpostavke ničelne hipoteze. Rdeča črtkana črta prikazuje kritično vrednost t-testa, ki je 1,318 za enostransko preizkušanje domneve pri stopnji tveganja  $\alpha = 10\%$ . Ta črta določa mejo, preko katere mora biti izračunana t-vrednost, da lahko zavrnemo ničelno hipotezo. Območje zavrnitve ničelne hipoteze je poudarjeno z rdečo barvo na desni strani kritične vrednosti. Ker izračunana t-vrednost presega to kritično mejo, graf kaže, da zavrnemo ničelno domnevo in sprejmemo alternativno domnevo, da študenti preberejo več kot 28 knjig na leto.

Grafični prikaz enostranskega preizkušanja domneve (t-test):



### Naloga 56b)

Zastavimo ničelno in raziskovalno domnevo:

$$H_0: \bar{y}_D = 22 \text{ knjig}$$

$$H_1: \bar{y}_D \neq 22 \text{ knjig}$$

Dvostransko preizkušanje domneve za mali vzorec:

$$\alpha = 1 \% \rightarrow t_{n-1; \alpha/2} = t_{24; 0,005} = \pm 2,797 \text{ (kritična vrednost spremenljivke } t.)$$

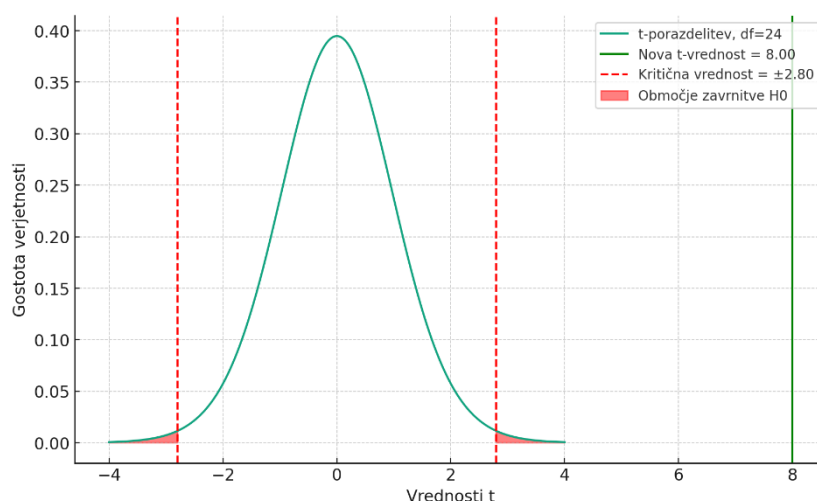
$$t = \frac{\bar{Y} - \bar{y}_D}{SE_{\bar{y}}}$$

$$t_i = \frac{30 - 22}{1} = 8$$

Ker je izračunana t-vrednost  $t_i=8$  večja od kritične vrednosti  $t_{24;0,005} = \pm 2,797$ , to pomeni, da sprejmemo hipotezo  $H_1$ .

Zelena črta na grafu predstavlja izračunano t-vrednost, ki znaša 8. Rdeče črtkane črte označujejo kritične vrednosti pri stopnji tveganja 1 % za dvostransko preizkušanje domneve s t-testom. To so meje, znotraj katerih bi obdržali ničelno domnevo, če bi naša t-vrednost padla mednje. Ker pa je naša izračunana t-vrednost precej zunaj teh meja, območje zavrnitve ničelne domneve, poudarjeno z rdečo barvo, nakazuje, da zavrnemo ničelno domnevo. S tem sprejmemo, da je povprečno število prebranih knjig na študenta statistično značilno različno (in v tem primeru večje) od 22 knjig na leto.

Grafični prikaz dvostranskega preizkušanja domneve (t-test):



### Naloga 56c)

Zastavimo ničelno in raziskovalno domnevo:

$H_0: \bar{y} \leq 20$  knjig (Povprečno število prebranih knjig na študenta je 20 knjig ali manj).

$H_1: \bar{y} > 20$  knjig (Povprečno število prebranih knjig na študenta je več kot 20 knjig).

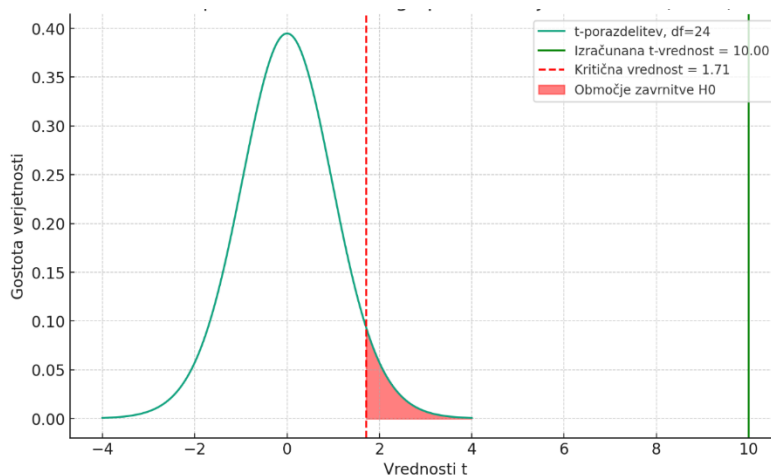
Enostransko preizkušanje domneve za mali vzorec:

$\alpha = 5\% \rightarrow t_{n-1; \alpha} = t_{24; 0,05} = 1,711$  (kritična vrednost spremenljivke t.)

$$t = \frac{\bar{Y} - \bar{y}_D}{SE_{\bar{y}}}$$

$$t_i = \frac{30 - 20}{1} = 10$$

Grafični prikaz enostranskega preizkušanja domneve (t-test):



Ker je izračunana t-vrednost večja od kritične vrednosti  $t_{24;0,05}=1,711$ , to pomeni, da sprejmemo hipotezo  $H_1$ .

### Naloga 57 a)

$n = 7$  (mali vzorec)

$\gamma = 80\% \rightarrow \alpha = 20\% \rightarrow t_{6;0,10} = 1,440$

$\bar{Y} = 27,286$

$s = 37,622$

$se_{\bar{y}} = 14,219$

$P(27,286 - 1,440 \cdot 14,219 < \bar{y} < 27,286 + 1,440 \cdot 14,219) = 80\%$

$P(6,811 < \bar{y} < 47,761) = 80\%$

### Naloga 57 b)

Enostransko preizkušanje domneve:

$\alpha = 5\% \rightarrow t_{n-1; \alpha} = t_{6;0,05} = +1,943$  (kritična vrednost spremenljivke t.)

$H_0: \bar{y}_D \leq 30$

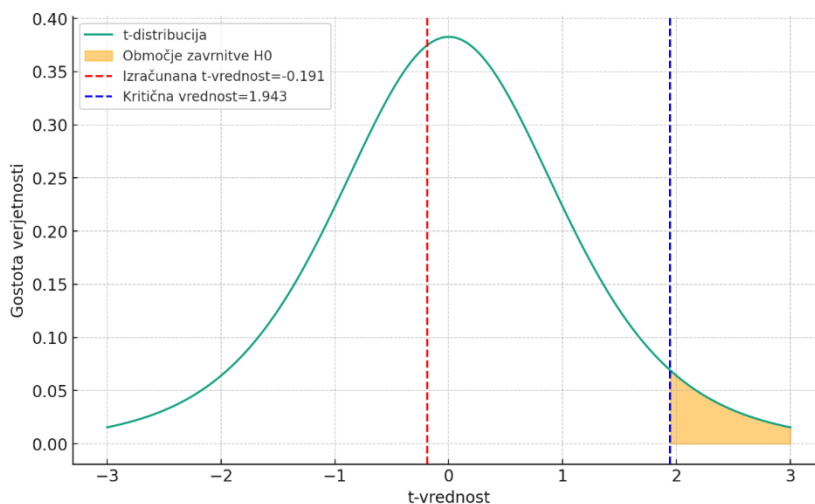
$H_1: \bar{y}_D > 30$

$se_{\bar{y}} = 14,219$

$t = -0,191$

Sprejmemo domnevo  $H_0$ .

Grafični prikaz enostranskega preizkušanja domneve (t-test):



Izračunana t-vrednost, ki znaša  $-0,191$ , je na grafu označena z rdečo črto. Na grafu modra črta prikazuje kritično vrednost  $1,943$ . Ta vrednost določa mejo za enostranski test pri stopnji tveganja  $5\%$ . Če bi izračunana t-vrednost preseгла to mejo na desni, bi zavrnili domnevo  $H_0$ . Območje v oranžni barvi desno od kritične vrednosti je območje zavrnitve domneve  $H_0$ . Ker izračunana t-vrednost ne pade v to območje, ampak ostaja na levi, sprejmemo domnevo  $H_0: \bar{y}_D \leq 30$ .

### Naloga 58a)

$$n = 5 \text{ (mali vzorec)}$$

$$\bar{Y} = 223,2$$

$$s = 28,665$$

$$se_{\bar{y}} = 12,819$$

$$z = \pm 1,96$$

$$P(223,2 - 1,96 \cdot 12,819 < \bar{y} < 223,2 + 1,96 \cdot 12,819) = 95\%$$

$$P(198,075 < \bar{y} < 248,325) = 95\%$$

### Naloga 58b)

Enostransko intervalno ocenjevanje aritmetične sredine iz malega vzorca s spodnjo mejo:

$$P(\bar{y} > \bar{Y} - t_{n-1;\alpha} \cdot se_{\bar{y}}) = \gamma$$

$$t_{n-1;\alpha} = t_{4;0,1} = 1,533$$

$$P(\bar{y} > 223,2 - 1,533 \cdot 12,819) = 90\%$$

$$P(\bar{y} > 203,55) = 90\%$$

Enostransko intervalno ocenjevanje aritmetične sredine iz malega vzorca z zgornjo mejo:

$$P(\bar{y} < \bar{Y} + t_{n-1;\alpha} \cdot se_{\bar{y}}) = \gamma$$

$$P(\bar{y} < 223,2 + 1,533 \cdot 12,819) = 90\%$$

$$P(\bar{y} < 242,85) = 90\%$$

### Naloga 58c)

$$H_0: \bar{y}_D \geq 220$$

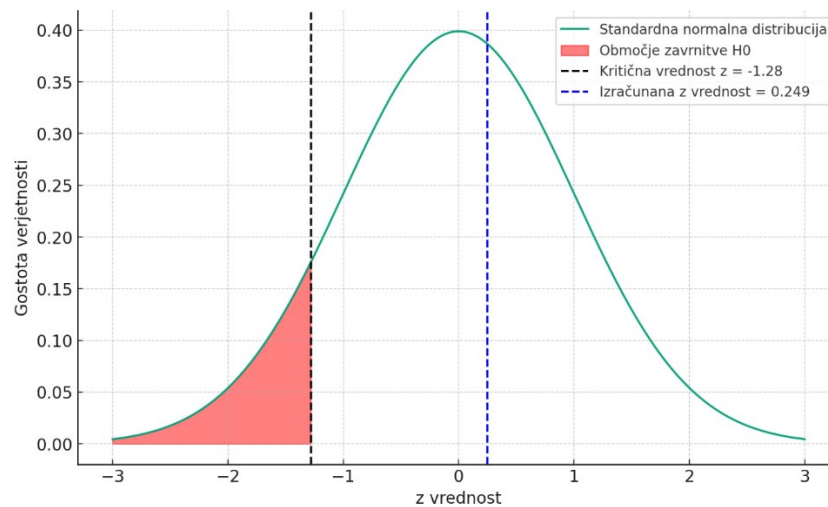
$$H_1: \bar{y}_D < 220$$

Kritična vrednost spremenljivke  $z = -1,28$  ( $\alpha = 10\%$ )

$$z = 0,249$$

Sprejmemo domnevo  $H_0$ .

Grafični prikaz enostranskega preizkušanja domneve (z-test):



### Naloga 59a)

Velik vzorec; dvostransko intervalno ocenjevanje aritmetične sredine

$$\bar{Y} = 60,515$$

$$s = 58,704$$

$$se_{\bar{y}} = 3,184$$

$$\gamma = 95 \% \rightarrow \alpha = 5 \% \rightarrow z = \pm 1,96$$

$$P(60,515 - 1,96 \cdot 3,184 < \bar{y} < 60,515 + 1,96 \cdot 3,184) = 95 \%$$

$$P(54,274 < \bar{y} < 66,756) = 95 \%$$

### Naloga 59b)

Velik vzorec; dvostransko preizkušanje domneve o aritmetični sredini

$$H_0: \bar{y}_D = 7.500$$

$$H_1: \bar{y}_D \neq 7.500$$

$$\bar{Y} = 60,515$$

$$SE_{\bar{y}} = 3,184$$

$$\text{Kritična vrednost spremenljivke } z = \pm 1,645 (\alpha = 0,10)$$

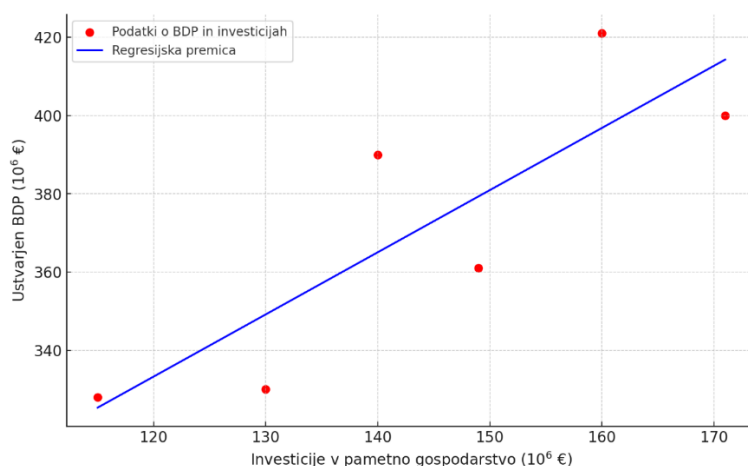
$$z = -4,549$$

Sprejmemo domnevo  $H_1$ .

## 4.6 Regresijska analiza

### Naloga 60 a)

Grafični prikaz:



Oblika: Linearna oblika

Smer med spremenljivkama: pozitivna smer, kar pomeni, da z naraščanjem investicij v pametno gospodarstvo ( $x$ ) v povprečju narašča ustvarjen BDP ( $y$ ).

Jakost povezanosti med spremenljivkama: obstaja močna povezanost med odvisno (ustvarjen BDP) in neodvisno spremenljivko (investicije v pametno gospodarstvo).

### Naloga 60 b)

Izračun parametrov regresijske premice:

Preden se lotimo izračuna regresijskih koeficientov  $b_0$  in  $b_1$ , izračunamo:

$$\sum x_i = 115 + 130 + 140 + 149 + 160 + 171 = 865$$

$$\sum x_i^2 = 115^2 + 130^2 + 140^2 + 149^2 + 160^2 + 171^2 = 126.767$$

$$\sum y_i = 328 + 330 + 390 + 361 + 421 + 400 = 2.230$$

$$\sum y_i^2 = 328^2 + 330^2 + 390^2 + 361^2 + 421^2 + 400^2 = 836.146$$

$$\sum x_i \cdot y_i = (115 \cdot 328) + (130 \cdot 330) + (140 \cdot 390) + \dots + (171 \cdot 400) = 324.769$$

$$\bar{x} = \frac{865}{6} = 144,1667$$

$$\bar{y} = \frac{2230}{6} = 371,6667$$

Oba regresijska koeficienta izračunamo po enačbi:

$$b_1 = \frac{\frac{1}{n-1} \sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\frac{1}{n-1} \sum x_i^2 - \bar{x}^2} = \frac{c_{xy}}{s_x^2} = \frac{(\sum x_i \cdot y_i) - n\bar{x}\bar{y}}{(n-1) \cdot s_x^2} = \frac{(\sum x_i \cdot y_i) - n\bar{x}\bar{y}}{(\sum x_i^2) - n\bar{x}^2}$$

$$b_0 = \bar{y} - b_1 \cdot \bar{x}$$

$$b_1 = \frac{(324769) - 6 \cdot 144,1667 \cdot 371,6667}{(126767) - 6 \cdot 144,1667^2} = 1,5887$$

$$b_0 = 371,6667 - 1,5887 \cdot 144,1667 = 142,6291$$

Izračunana regresijska koeficienta vstavimo v enačbo regresijske premice:

$$\hat{y}_i = b_0 + b_1 x_i$$

$$\hat{y} = 142,6291 + 1,5887 \cdot x_i$$

Pomen regresijskega koeficienta  $b_0$ : pri investicijah v pametno gospodarstvo  $x = 0$  lahko v povprečju pričakujemo, da bo ustvarjen BDP 142,6291 (v  $10^6$  EUR).

Pomen regresijskega koeficienta  $b_1$ : če se investicije v pametno gospodarstvo ( $x$ ) povečajo za eno enoto (v  $10^6$  EUR), se ustvarjen BDP ( $y$ ) v povprečju poveča za 1,5887 (v  $10^6$  EUR).

### Naloga 60 c)

$$x = 180 \text{ (v } 10^6 \text{ EUR)}$$

$$\hat{y} = 142,6291 + 1,5887 \cdot x_i$$

$$\hat{y}_{x=180} = 142,6291 + 1,5887 \cdot 180$$

$$\hat{y}_{x=180} = 428,649 \text{ (v } 10^6 \text{ EUR)}$$

### Naloga 60 d)

Parameter, na osnovi katerega določimo smer in jakost linearne korelacijske odvisnosti, je korelacijski koeficient.

Izračunamo ga po enačbi:

$$r_{xy} = \frac{(\sum x_i \cdot y_i) - n\bar{x}\bar{y}}{(n-1) \cdot s_x \cdot s_y}$$



Izračun  $s_x$  in  $s_y$ :

$$s_x^2 = \frac{1}{n-1} \cdot \sum (x_i - \bar{x})^2$$

$$s_x^2 = \frac{1}{5} \cdot [(115 - 144,1667)^2 + (130 - 144,1667)^2 + (140 - 144,1667)^2 + \dots + (171 - 144,1667)^2] = 412,5667$$

$$s_x = \sqrt{412,5667} = 20,3117$$

$$s_y^2 = \frac{1}{n-1} \cdot \sum (y_i - \bar{y})^2$$

$$s_y^2 = \frac{1}{5} \cdot [(328 - 371,6667)^2 + (330 - 371,6667)^2 + (390 - 371,6667)^2 + \dots + (400 - 371,6667)^2] = 1465,8667$$

$$s_y = \sqrt{1465,8667} = 38,2866$$

$$(\sum x_i \cdot y_i) - n \cdot \bar{x} \cdot \bar{y} = 324769 - 6 \cdot 144,1667 \cdot 371,6667 = 3277,2302$$

$$r_{xy} = \frac{3277,2302}{5 \cdot 20,3117 \cdot 38,2866} = 0,8428$$

Na osnovi rezultata ( $r_{xy} = 0,8428$ ) vidimo, da obstaja močna povezanost med odvisno (ustvarjen BDP) in neodvisno spremenljivko (investicije v pametno gospodarstvo). Smer povezanosti je pozitivna.

### Naloga 60 e)

Izračunati moramo delež pojasnjene variance v skupni varianci za odvisno spremenljivko (determinacijski koeficient):

$$r_{xy}^2 = 0,8428^2 = 0,7103 \text{ oz. } 71,03 \%$$

Delež pojasnjene variance v skupni varianci za odvisno spremenljivko znaša 71,03 %.

### Naloga 60 f)

Standardno napako ocene odvisne spremenljivke izračunamo po enačbi:

$$s_{y.x} = \sqrt{\frac{(\sum y_i^2) - b_0(\sum y_i) - b_1(\sum x_i \cdot y_i)}{n-2}}$$

$$s_{y.x} = \sqrt{\frac{836146 - (142,6291 \cdot 2230) - (1,5887 \cdot 324769)}{6-2}} = 23,036$$

Standardna napaka ocene odvisne spremenljivke je različna od 0, kar pomeni, da na ustvarjen BDP (odvisna spremenljivka) poleg investicij v pametno gospodarstvo (neodvisna spremenljivka) vplivajo še druge spremenljivke in slučajni vplivi.

### Naloga 60 g)

Intervalno oceno izračunamo po enačbi:

$$\hat{y}_{x_0} \pm t_{n-2; \alpha/2} \cdot s_{y.x} \cdot \sqrt{1 + h_1}$$

$$\hat{y}_x = 428,649 \text{ (izračunano pri nalogi c)}$$

$$s_{yx} = 23,036 \text{ (izračunano pri nalogi f)}$$

Upoštevamo, da je pri  $\gamma = 95\%$ ,  $\alpha = 5\%$ .

Izračun:  $t_{n-2; \alpha/2} = t_{4; 0,025} = 2,776$  (gledamo tabelo *kritične vrednosti za t porazdelitev*)

Upoštevamo, da je popravek  $h_1$  enak 0.

Izračunamo intervalno oceno:

$$P(428,649 - 2,776 \cdot 23,036 \cdot 1 < y_{x=180} < 428,649 + 2,776 \cdot 23,036 \cdot 1) = 95\%$$

$$P(364,701 < y_{x=180} < 492,597) = 95\%$$

Pri investicijah v pametno gospodarstvo  $x = 180$  (v  $10^6$  EUR) bo ustvarjen BDP med 364,701 EUR in 492,597 (v  $10^6$  EUR), kar trdimo s 95-odstotno verjetnostjo.

### Naloga 61 a), b)

$$\sum x_i = 31$$

$$\sum x_i^2 = 229$$

$$\sum y_i = 28,4$$

$$\sum y_i^2 = 171,34$$

$$\sum x_i \cdot y_i = 195,2$$

$$\bar{x} = \frac{31}{5} = 6,2$$

$$\bar{y} = \frac{28,4}{5} = 5,68$$

Izračun regresijskih koeficientov:

$$b_1 = \frac{\frac{1}{n-1} \sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\frac{1}{n-1} \sum x_i^2 - \bar{x}^2} = \frac{c_{xy}}{s_x^2} = \frac{(\sum x_i \cdot y_i) - n\bar{x}\bar{y}}{(n-1) \cdot s_x^2} = \frac{(\sum x_i \cdot y_i) - n\bar{x}\bar{y}}{(\sum x_i^2) - n\bar{x}^2}$$

$$b_0 = \bar{y} - b_1 \cdot \bar{x}$$

$$b_1 = \frac{(195,2) - 5 \cdot 6,2 \cdot 5,68}{(229) - 5 \cdot 6,2^2} = 0,5196$$

Pomen regresijskega koeficienta  $b_1$ : če se število usposabljanj za zaposlene ( $x$ ) poveča za eno enoto, se rast prihodka v podjetju ( $y$ ) v povprečju poveča za 0,5196 %.

$$b_0 = 5,68 - 0,5196 \cdot 6,2 = 2,4585$$

Pomen regresijskega koeficienta  $b_0$ : pri številu usposabljanj za zaposlene  $x = 0$  bi v povprečju rast prihodka v podjetju znašala 2,4585 %.

Izračunana regresijska koeficienta vstavimo v enačbo regresijske premice:

$$\hat{y}_i = b_0 + b_1 x_i$$

$$\hat{y} = 2,4585 + 0,5196 \cdot x_i$$

$$x = 15$$

$$\hat{y}_{x=15} = 2,4585 + 0,5196 \cdot 15$$

$$\hat{y}_{x=15} = 10,2525$$

Standardno napako ocene odvisne spremenljivke izračunamo po enačbi:

$$s_{y \cdot x} = \sqrt{\frac{(\sum y_i^2) - b_0(\sum y_i) - b_1(\sum x_i \cdot y_i)}{n-2}}$$

$$s_{yx} = \sqrt{\frac{171,34 - (2,4585 \cdot 28,4) - (0,5196 \cdot 195,2)}{5-2}} = \sqrt{\frac{0,09268}{3}} = 0,1758$$

Standardna napaka ocene odvisne spremenljivke je različna od 0, kar pomeni, da na rast prihodka v podjetju (odvisna spremenljivka) poleg števila usposabljanj za zaposlene (neodvisna spremenljivka) vplivajo še druge spremenljivke in slučajni vplivi.

Izračun intervalne ocene:

$$\hat{y}_{x=15} = 10,2525$$

$$s_{yx} = 0,1758$$

Pri  $\gamma = 95\%$  je  $\alpha = 5\%$ .

Izračun:  $t_{n-2; \alpha/2} = t_{3; 0,025} = 3,182$  (gledamo tabelo *kritične vrednosti za t porazdelitev*)

Vrednost za  $h_1$  izračunamo po enačbi:

$$h_1 = \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} = \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i)^2 - (\sum_i x_i)^2 / n}$$

$$h_1 = \frac{1}{5} + \frac{(15 - 6,2)^2}{229 - 31^2/5} = \frac{1}{5} + 2,1043 = 2,3043$$

Izračunane vrednosti vstavimo v enačbo:

$$\hat{y}_{x_0} \pm t_{n-2; \alpha/2} \cdot s_{y \cdot x} \cdot \sqrt{1 + h_1}$$

$$10,2525 \pm 3,182 \cdot 0,1758 \cdot \sqrt{3,3043}$$

Intervalna ocena:

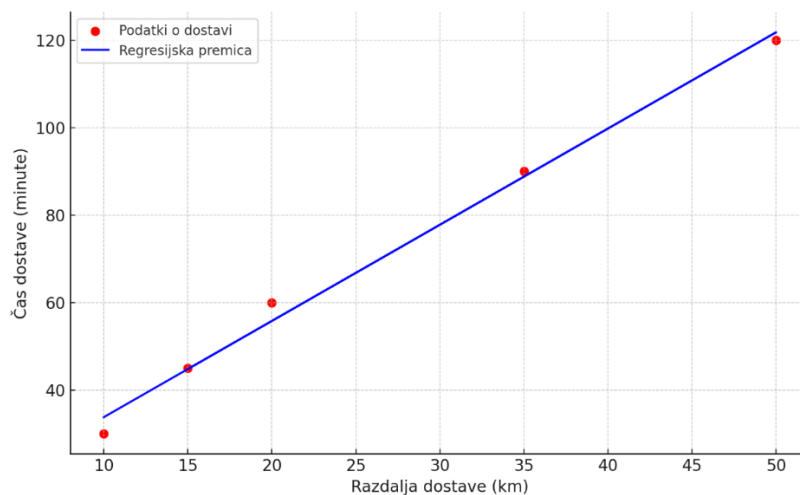
$$P(10,2525 - 3,182 \cdot 0,1758 \cdot 1,818 < y_{x=15} < 10,2525 + 3,182 \cdot 0,1758 \cdot 1,818) = 95\%$$

$$P(9,24 < y_{x=15} < 11,27) = 95\%$$

S 95-odstotno verjetnostjo ocenjujemo, da bo pri številu usposabljanj zaposlenih  $x = 15$  rast prihodkov v podjetju med 9,24 % in 11,27 %.

### Naloga 63a)

Razsevni grafikon kaže na pozitivno korelacijo med razdaljo in časom dostave: večja kot je razdalja, daljši je čas dostave. Regresijska premica kaže, kako čas dostave narašča z razdaljo. Smer povezanosti med odvisno (čas dostave) in neodvisno spremenljivko (razdalja dostave) je pozitivna. Oblika povezanosti med spremenljivkama je linearna.



### Naloga 63b)

$$\sum x_i = 130$$

$$\sum x_i^2 = 4450$$

$$\sum y_i = 345$$

$$\sum y_i^2 = 29025$$

$$\sum x_i \cdot y_i = 11325$$

$$\bar{x} = 26$$

$$\bar{y} = 69$$

Izračun regresijskih koeficientov:

$$b_1 = \frac{(11325) - 5 \cdot 26 \cdot 69}{(4450) - 5 \cdot 26^2} = 2,201$$

$$b_0 = 69 - 2,201 \cdot 26 = 11,774$$

Izračunana regresijska koeficienta vstavimo v enačbo regresijske premice:

$$\hat{y} = 11,774 + 2,201 \cdot x_i$$

Korelacijski koeficient izračunamo po enačbi:

$$r_{xy} = \frac{(\sum x_i \cdot y_i) - n\bar{x}\bar{y}}{(n-1) \cdot s_x \cdot s_y}$$

Izračun  $s_x$  in  $s_y$ :

$$s_x^2 = \frac{1}{n-1} \cdot \sum (x_i - \bar{x})^2$$

$$s_x^2 = \frac{1}{4} \cdot [(10 - 26)^2 + (20 - 26)^2 + (50 - 26)^2 + (35 - 26)^2 + (15 - 26)^2] = 267,5$$

$$s_x = \sqrt{267,5} = 16,355$$

$$s_y^2 = \frac{1}{n-1} \cdot \sum (y_i - \bar{y})^2$$

$$s_y^2 = \frac{1}{4} \cdot [(30 - 69)^2 + (60 - 69)^2 + (120 - 69)^2 + (90 - 69)^2 + (45 - 69)^2] = 1305$$

$$s_y = \sqrt{1305} = 36,125$$

$$r_{xy} = \frac{11325 - 5 \cdot 26 \cdot 69}{4 \cdot 16,355 \cdot 36,125} = 0,996$$

Determinacijski koeficient:

$$r_{xy}^2 = 0,996^2 = 0,992 \text{ oz. } 99,2 \%$$

Standardno napako ocene odvisne spremenljivke izračunamo po enačbi:

$$s_{y \cdot x} = \sqrt{\frac{(\sum y_i^2) - b_0(\sum y_i) - b_1(\sum x_i \cdot y_i)}{n-2}}$$

$$s_{yx} = \sqrt{\frac{29025 - (11,774 \cdot 345) - (2,201 \cdot 11325)}{5-2}} = 3,495$$

### Naloga 63 c)

$x = 25$  minut

$$\hat{y} = 11,774 + 2,201 \cdot x_i$$

$$\hat{y}_{x=25} = 11,774 + 2,201 \cdot 25$$

$$\hat{y}_{x=25} = 66,799$$

Izračun intervalne ocene:

$$\hat{y}_{x=25} = 66,799$$

$$s_{y \cdot x} = 3,495$$

$$\alpha = 5 \%$$

Izračun:  $t_{n-2; \alpha/2} = t_{3; 0,025} = 3,182$  (gledamo tabelo *kritične vrednosti za t porazdelitev*)

Upoštevamo, da je popravek  $h_1$  enak 0.

Intervalna ocena:

$$P(66,799 - 3,182 \cdot 3,495 \cdot 1 < y_{x=25} < 66,799 + 3,182 \cdot 3,495 \cdot 1) = 95 \%$$

$$P(55,678 < y_{x=25} < 77,920) = 95 \%$$

### Naloga 64

Povprečna stopnja vključenosti uporabnikov (%) je neodvisna spremenljivka  $x$ .  
Stroški oglaševalske kampanje (v d.e.) je odvisna spremenljivka  $y$ .

$$\Sigma x_i = 501,2$$

$$\Sigma x_i^2 = 36.105,98$$

$$\Sigma y_i = 689,91$$

$$\Sigma y_i^2 = 74.444,984$$

$$\Sigma x_i \cdot y_i = 50.210,032$$

$$\bar{x} = 71,6$$

$$\bar{y} = 98,559$$

Izračun regresijskih koeficientov:

$$b_1 = \frac{(50210,032) - 7 \cdot 71,6 \cdot 98,559}{(36105,98) - 7 \cdot 71,6^2} = 3,691$$

$$b_0 = 98,559 - 3,691 \cdot 71,6 = -165,717$$

Izračunana regresijska koeficienta vstavimo v enačbo regresijske premice:

$$\hat{y} = -165,717 + 3,691 \cdot x_i$$

$$\hat{y}_{x=90} = -165,717 + 3,691 \cdot 90$$

$$\hat{y}_{x=90} = 166,473$$

Standardna napaka ocene odvisne spremenljivke:

$$s_{yx} = \sqrt{\frac{74444,984 - (-165,717 \cdot 689,91) - (3,691 \cdot 50210,032)}{7 - 2}} = 26,266$$

Izračun intervalne ocene:

$$\hat{y}_{x=90} = 166,473$$

$$s_{yx} = 26,266$$

$$\alpha = 5 \%$$

Izračun:  $t_{n-2, \alpha/2} = t_{5, 0,025} = 2,571$  (gledamo tabelo *kritične vrednosti za t porazdelitev*)

Upoštevamo, da je popravek  $h_1$  enak 0.

$$166,473 \pm 2,571 \cdot 26,266 \cdot 1$$

Izračun intervalne ocene:

$$P(166,473 - 2,571 \cdot 26,266 \cdot 1 < y_{x=90} < 166,473 + 2,571 \cdot 26,266 \cdot 1) = 95 \%$$

$$P(98,943 < y_{x=90} < 234,003) = 95 \%$$

### Naloga 65

$$\Sigma x_i = 621$$

$$\Sigma x_i^2 = 50.001$$

$$\Sigma y_i = 707,2$$

$$\Sigma y_i^2 = 63.670,58$$

$$\Sigma x_i \cdot y_i = 56.233,9$$

$$\bar{x} = 77,625$$

$$\bar{y} = 88,4$$

Izračun regresijskih koeficientov:

$$b_1 = 0,745$$

$$b_0 = 30,569$$

Izračunana regresijska koeficienta vstavimo v enačbo regresijske premice:

$$\hat{y} = 30,569 + 0,745 \cdot x_i$$

Izračun  $s_x$  in  $s_y$ :

$$s_x^2 = 256,554$$



$$s_x = 16,017$$

$$s_y^2 = 164,871$$

$$s_y = 12,840$$

$$r_{xy} = \frac{1337,5}{7 \cdot 16,017 \cdot 12,840} = 0,929$$

Determinacijski koeficient:

$$r_{xy}^2 = 0,929^2 = 0,863 \text{ oz. } 86,3 \%$$

### Naloga 66 a), b)

Korelacijski koeficient:  $r_{xy} = 0,840$

Determinacijski koeficient:  $r_{xy}^2 = 0,706$

Standardna napaka ocene odvisne spremenljivke:  $s_{yx} = 11,334$

$$b_0 = -7,647$$

$$b_1 = 0,634$$

Izračunana regresijska koeficienta vstavimo v enačbo regresijske premice:

$$\hat{y} = -7,647 + 0,634 \cdot 35$$

$$\hat{y}_{x=35} = 14,543$$

## 4.7 Napovedovanje vrednosti v časovni vrsti z uporabo trenda in sezonske komponente

### Naloga 67a)

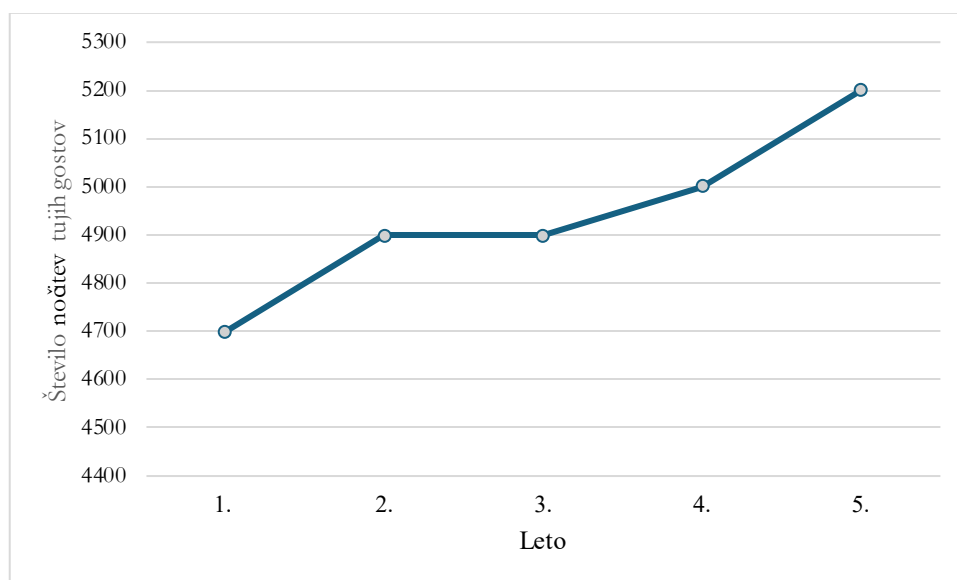
Sistem normalnih enačb za izračun funkcije trenda:

$$5a + 15b = 24.700$$

$$15a + 55b = 75.200$$

Funkcija trenda:  $\hat{Y} = 4610 + 110t$

Napoved za število nočitev za 6. leto:  $\hat{Y}_t = 4610 + 110 \cdot 6 = 5.270$



### Naloga 67 b)

Leto	I-IV	V-VIII	IX-XII	Skupaj
1	1.200	2.500	1.000	4.700
2	1.100	2.600	1.200	4.900
3	1.000	2.800	1.100	4.900
4	1.300	2.500	1.200	5.000
5	1.200	2.700	1.300	5.200
Skupaj	5.800	1.3100	5.800	24.700

Povprečno 4-mesečno število nočitev v obdobju petih let:  $\frac{24700}{3} = 8.233,3$  nočitev.

$$SI1 = \frac{5800}{8233,3} \cdot 100 = 70,4 \%$$

$$SI2 = \frac{13100}{8233,3} \cdot 100 = 159,2 \%$$

$$SI3 = \frac{5800}{8233,3} \cdot 100 = 70,4 \%$$

Povprečno 4-mesečno število nočitev v 6. letu:  $\frac{5270}{3} = 1.756,6$ .

Napoved po 4-mesečjih za 6. leto:

$$\hat{Y}_{I-IV;6.letu} = 0,704 \cdot 1.756,6 = 1.236,7$$

$$\hat{Y}_{V-VIII;6.letu} = 1,592 \cdot 1.756,6 = 2.796,6$$

$$\hat{Y}_{IX-XII;6.letu} = 0,704 \cdot 1.756,6 = 1.236,7$$

**Naloga 68a)**

Ocena stroškov za prihodnje leto, v d.e.

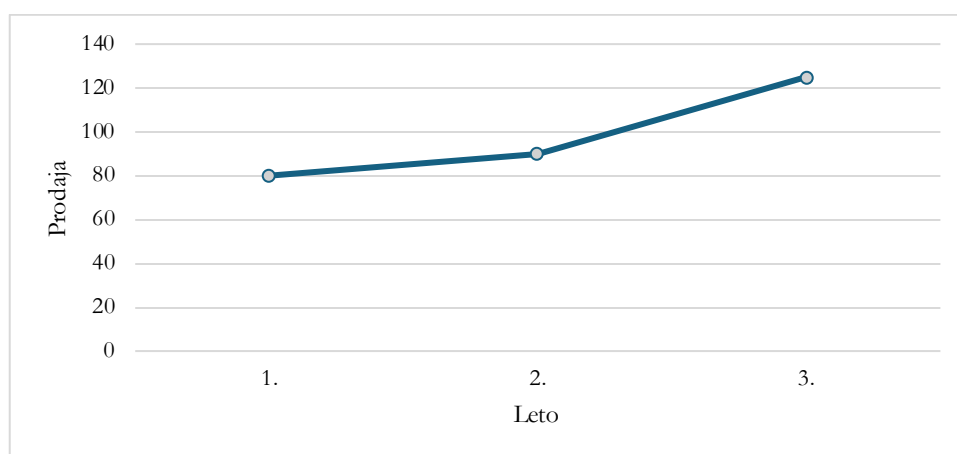
$$\hat{Y}_5 = 58,5 + 2,1 \cdot 5 = 69$$

**Naloga 68 c)**

Povprečni četrletni stroški za 5. leto:  $\frac{69}{4} = 17,25$  d.e.

Ocena stroškov za 4. četrletje, 5. leto, v d.e.:

$$\hat{Y}_{4. \text{ četrletje, 5. leto}} = 2,255 \cdot 17,25 = 38,90$$

**Naloga 69a)****Naloga 69b)**

Funkcija trenda:  $\hat{Y} = 53,33 + 22,5t$

**Naloga 69c)**

Povprečna 4-mesečna prodaja v obdobju treh let:  $\frac{295}{3} = 98,33$  enot

$$SI1 = \frac{60}{98,33} \cdot 100 = 61,02 \%$$

$$SI2 = \frac{160}{98,33} \cdot 100 = 162,71 \%$$

$$SI3 = \frac{75}{98,33} \cdot 100 = 76,27 \%$$

### Naloga 69d)

Napoved prodaje za 5. leto s funkcija trenda:

$$t = 5$$

$$\hat{Y} = 53,33 + 22,5 \cdot 5$$

$$\hat{Y} = 165,83$$

Povprečna 4-mesečna prodaja jagod za 5. leto:  $\frac{165,83}{3} = 55,28$  enot

Ocena prodaje jagod za 5. zaporedno leto, po 4-mesečjih:

$$\hat{Y}_{5. \text{ leto, prvo 4-mesečje}} = 0,6102 \cdot 55,28 = 33,73 \text{ enot}$$

$$\hat{Y}_{5. \text{ leto, drugo 4-mesečje}} = 1,6271 \cdot 55,28 = 89,95 \text{ enot}$$

$$\hat{Y}_{5. \text{ leto, tretje 4-mesečje}} = 0,7627 \cdot 55,28 = 42,16 \text{ enot}$$

### Naloga 70a)

Ocena prihodkov 1 za prihodnje leto, v d.e.

$$\hat{Y}_5 = 120 + 4,5 \cdot 5 = 142,5$$

### Naloga 70b)

Povprečni četrtni prihodki za 5. leto:  $\frac{142,5}{4} = 35,625$  d.e.

Ocena prihodkov za 4. četrtnje, 5. leto, v d.e.:

$$\hat{Y}_{4. \text{ četrtnje, 5. leto}} = 1,954 \cdot 35,625 = 69,611$$

### Naloga 71b)

*Linearna funkcija trenda:*

*Sistem normalnih enačb:*

$$Ta + \left(\sum_{t=1}^T t\right)b = \sum_{t=1}^T Y_t$$

$$\left(\sum_{t=1}^T t\right)a + \left(\sum_{t=1}^T t^2\right)b = \sum_{t=1}^T tY_t,$$

Leto	t	$Y_t$	$t \cdot Y_t$	$t^2$
2019	1	80	80	1
2020	2	120	240	4
2021	3	200	600	9
2022	4	300	1200	16
2023	5	450	2250	25
Skupaj	15	1150	4370	55

$$5a + 15b = 1150 \quad / : 5, \cdot (-15)$$

$$15a + 55b = 4370$$

$$\hline 0 + 10b = 920$$

$$b = 92$$

$$a = -46$$

Funkcija trenda:  $\hat{Y} = -46 + 92 \cdot t$

Napoved za število prodanih električnih avtomobilov podjetja X za leto 2027:

$$\hat{Y}_t = -46 + 92 \cdot 9 = 782 \text{ električnih avtomobilov}$$



## Ploščine $H(z)$ za standardizirano normalno porazdelitev

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4014
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4983	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987									
3,5	0,4997									
4,0	0,4999									

Primer: Za  $z = 1,96$  iz preglednice odčitamo površino 0,4750.

(Vir: Artenjak, 2003)

## Kritične vrednosti za $t$ porazdelitev

$$t = \frac{\bar{Y} - \bar{Y}_D}{SE_{\bar{y}}}$$

prostostne stopinje	0,10	0,05	0,025	0,01	0,005
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	2,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
60	1,296	1,671	2,000	2,390	2,660
120	1,289	1,658	1,980	2,358	2,617
	1,282	1,645	1,960	2,326	2,576

(Vir: Artenjak, 2003)



# Obrazci<sup>2</sup>

## 1. del

### Urejanje, prikazovanje in analiza podatkov

$N$	število enot v osnovni statistični množici
$y$	spremenljivka
$k$	razred
$r$	število razredov
$f_k$	število enot v $k$ -tem razredu
$y_{\min}$	najmanjša vrednost spremenljivke
$y_{\max}$	največja vrednost spremenljivke
$y_{k,\min}$	spodnja meja razreda $k$
$y_{k,\max}$	zgornja meja razreda $k$

$$i_k = y_{k,\max} - y_{k,\min} \quad \text{širina razreda } k$$

$$y_k = \frac{y_{k,\min} + y_{k,\max}}{2} \quad \text{sredina razreda } k$$

### Gostota frekvence

$$g_k = \frac{f_k}{i_k} \quad \text{za } k = 1, 2, \dots, r$$

### Kumulativni člani frekvenčne porazdelitve

$$F_1 = f_1, F_k = F_{k-1} + f_k \quad \text{za } k = 2, 3, \dots, r$$

*Strukturni odstotek:*

$$f\%(y_k) = 100 \frac{f(y_k)}{N} = 100 \cdot f^\circ(y_k) \quad \text{za } k = 1, 2, \dots, r$$

---

<sup>2</sup> Obrazci so povzeti po prilogi obrazci iz učbenika Artenjak (2003).

Ločne stopinje:

krog  $f^t(y_k) = 3,6 \cdot f\%(y_k)$  in  $f^t(y_k, x_j) = 3,6 \cdot f\%(y_k, x_j)$

polkrog  $f^t(y_k) = 1,8 \cdot f\%(y_k)$  in  $f^t(y_k, x_j) = 1,8 \cdot f\%(y_k, x_j)$

### Indeksi s stalno osnovo

$$I_{i/o} = 100 \times \frac{Y_i}{Y_o} \quad \text{za } i = 1, 2, \dots, N$$

### Preračunavanje indeksov

$$I_{i/j} = 100 \times \frac{I_{i/o}}{I_{j/o}}$$

### Indeksi s stalno osnovo iz časovnih vrst

$$I_{t/o} = 100 \times \frac{Y_t}{Y_o} \quad \text{za } t = 1, 2, \dots, T$$

### Verižni indeksi

$$V_t = \text{in } V_t = 100 \times \frac{Y_t}{Y_{t-1}} \quad \text{za } t = 2, 3, \dots, T$$

### Odnosi med členi v časovni vrsti

$$\frac{Y_t}{Y_{t-1}} = \frac{I_{t/o}}{I_{t-1/o}} = \frac{V_t}{100}$$

### Koeficienti dinamike

$$K_t = \text{in } K_t = \frac{Y_t}{Y_{t-1}} \quad \text{za } t = 2, 3, \dots, T,$$

### Stopnje rasti

$$S_t = \text{in } S_t = V_t - 100 \quad \text{za } t = 2, 3, \dots, T$$

$$S_t = -S_t = 100(K_t - 1) \quad \text{za } t = 2, 3, \dots, T$$

### Povezava med koeficientom dinamike in verižnim indeksom

$$K_t = (V_t/100); V_t = 100K_t$$

### Statistični koeficient

$$K = \frac{X}{Y} \times E$$

### Koeficient obračanja zalog

$$K = \frac{X \cdot E}{Y \cdot i},$$

### Povprečna vrednost zalog

$$\bar{Y} = \frac{1}{T}(Y_1 + Y_2 + \dots + Y_T),$$

$$\bar{Y} = \frac{1}{T}\left(\frac{1}{2}Y_0 + Y_1 + \dots + Y_{T-1} + \frac{1}{2}Y_T\right),$$

### Kvantili iz nerazvrščenih vrednosti

$$R_i = N \times P_i + 0,5$$

$$R_0 \leq R_i < R_1$$

$$y_i = y_0 + \frac{R_i - R_0}{R_1 - R_0} \times (y_1 - y_0)$$

### Kvantilni rangi iz nerazvrščenih vrednosti

$$y_0 \leq y_i < y_1$$

$$R_i = R_0 + \frac{y_i - y_0}{y_1 - y_0} \times (R_1 - R_0)$$

$$P_i = \frac{R_i - 0,5}{N},$$

## Aritmetična sredina iz nerazvrščenih vrednosti

$$\bar{y} = \frac{1}{N}(y_1 + y_2 + \dots + y_N) = \frac{1}{N} \sum_{i=1}^N y_i$$

## Aritmetična sredina iz razvrščenih vrednosti

$$\bar{y} = \frac{1}{N} \sum_{k=1}^r f_k y_k,$$

## Geometrijska sredina

*Povprečni koeficient dinamike*

$$K = \sqrt[r]{\frac{Y_T}{Y_1}},$$

$$K = \frac{V}{100} = \frac{\sqrt[r]{V_2 \times V_3 \times \dots \times V_T}}{100}$$

$$K = \sqrt[r]{\frac{I_{T/o}}{I_{1/o}}}$$

$$K = \sqrt[r]{K_2 \times K_3 \times \dots \times K_T}$$

## Povprečna stopnja rasti

$$S = (K - 1)100 = 100K - 100 = V - 100$$

## Variacijski razmik

$$VR = y_{max} - y_{min}$$

## Kvartilni razmik

$$Q = Q_3 - Q_1$$

## Decilni razmik

$$D = D_9 - D_1$$

**Varianca iz nerazvrščenih vrednosti**

$$VAR = \sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2$$

ali

$$VAR = \sigma^2 = \frac{1}{N} \sum_{i=1}^N y_i^2 - \bar{y}^2,$$

**Varianca iz razvrščenih vrednosti**

$$\sigma^2 = \frac{1}{N} \sum_{k=1}^r f_k (y_k - \bar{y})^2$$

ali

$$\sigma^2 = \frac{1}{N} \sum_{k=1}^r f_k y_k^2 - \bar{y}^2$$

**Standardni odklon**

$$SD = \sigma = \sqrt{VAR} = \sqrt{\sigma^2},$$

**Koeficient variabilnosti v odstotku**

$$KV\% = \frac{\sigma}{\bar{y}} \times 100$$

**Standardizirana spremenljivka**

$$z_i = \frac{y_i - \bar{y}}{\sigma} \quad \text{za } i = 1, 2, \dots, N$$

**Koeficient asimetrije na podlagi modusa**

$$KA_{Mo} = \frac{\bar{y} - Mo}{\sigma}$$

## Koeficient asimetrije na podlagi mediane

$$KA_{Me} = \frac{3(\bar{y} - Me)}{\sigma}$$

## 2. del

### Faktorska analiza

#### Določitev faktorjev – metoda glavnih component

$$z_1 = a_{11}F_1 + a_{12}F_2 + \dots + a_{1k}F_k$$

$$z_2 = a_{21}F_1 + a_{22}F_2 + \dots + a_{2k}F_k$$

...

$$z_k = a_{k1}F_1 + a_{k2}F_2 + \dots + a_{kk}F_k$$

$z_i$  – standardizirana vrednost  $i$ -te opazovane spremenljivke,  $i = 1, \dots, k$

$F_j$  –  $j$ -ta glavna komponenta oziroma faktor,  $j = 1, \dots, k$

$a_{ij}$  – faktorska utež pri  $i$ -ti spremenljivki in  $j$ -tem faktorju (korelacijski koeficienti, pri neodvisnih faktorjih)

#### Komunalitet $i$ -te merjene spremenljivke – $h_i^2$

$$h_i^2 = a_{i1}^2 + a_{i2}^2 + \dots + a_{im}^2$$

$a_{ij}^2$  – kvadrat faktorske uteži pri  $i$ -ti ( $i = 1, 2, \dots, k$ ) spremenljivki in  $j$ -tem faktorju ( $j = 1, 2, \dots, m$ ;  $m < k$ ), ki predstavlja delež celotne variance  $i$ -te spremenljivke, ki jo pojasni  $j$ -ti faktor.

#### Lastna vrednost $j$ -tega faktorja

$$a_{1j}^2 + a_{2j}^2 + \dots + a_{kj}^2 = \lambda_j$$

## 3. del

### Korelacija in regresija

#### *Sistem normalnih enačb za linearno funkcijo*

$$b_0 \cdot n + b_1 \sum x_i = \sum y_i$$

$$b_0 \cdot \sum x_i + b_1 \sum x_i^2 = \sum x_i y_i$$

$$b_1 = \frac{\frac{1}{n-1} \sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\frac{1}{n-1} \sum x_i^2 - \bar{x}^2} = \frac{c_{xy}}{s_x^2} = \frac{(\sum x_i \cdot y_i) - n\bar{x}\bar{y}}{(n-1) \cdot s_x^2} = \frac{(\sum x_i \cdot y_i) - n\bar{x}\bar{y}}{(\sum x_i^2) - n\bar{x}^2}$$

$$b_0 = \bar{y} - b_1 \cdot \bar{x}$$

## Kovarianca

$$c_{xy} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

ali

$$c_{xy} = \frac{1}{N} \sum_{i=1}^N x_i y_i - \bar{x} \cdot \bar{y},$$

## Determinacijski koeficient

$$r_{xy}^2 = \frac{\text{pojasnjena vsota kvadratov}}{\text{celotna vsota kvadratov}} = \left( \frac{c_{xy}}{s_x \cdot s_y} \right)^2$$

## Skupna variance

$$\sigma_y^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2$$

## Korelacijski koeficient

$$r_{xy} = \frac{c_{xy}}{s_x \cdot s_y}$$

$$r_{xy} = \frac{(\sum x_i \cdot y_i) - n\bar{x}\bar{y}}{(n-1) \cdot s_x \cdot s_y}$$

## Standardna napaka ocene odvisne spremenljivke

$$s_{y \cdot x} = \sqrt{\frac{(\sum y_i^2) - b_0(\sum y_i) - b_1(\sum x_i \cdot y_i)}{n-2}}$$

## Točkovna ocena

$$\hat{y}_{x_0} = b_0 + b_1 x_0$$

## Intervalna ocena

$$\hat{y}_{x_0} \pm t_{n-2; \alpha/2} \cdot s_{y \cdot x} \cdot \sqrt{1 + h_1}$$

$$h_1 = \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} = \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i)^2 - (\sum_i x_i)^2 / n}$$

## 4. del

### Časovne vrste

### Trendne funkcije

Enačba premice  $\hat{Y} = a + bt$

Parabola druge stopnje  $\hat{Y} = a + bt + ct^2$

Parabola tretje stopnje  $\hat{Y} = a + bt + ct^2 + dt^3$

Eksponentna funkcija  $\hat{Y} = ab^t$

Gompertzova krivulja  $\hat{Y} = Ka^{b^t}$

Pearl-Reedova logistična krivulja  $\hat{Y} = \frac{\hat{Y}_\infty}{1 + e^{a+bt}}$

## Sistem normalnih enačb za linearni trend

$$Ta + \left( \sum_{t=1}^T t \right) b = \sum_{t=1}^T Y_t$$

$$\left( \sum_{t=1}^T t \right) a + \left( \sum_{t=1}^T t^2 \right) b = \sum_{t=1}^T t Y_t,$$

$$b = \frac{\frac{1}{T} \sum_{t=1}^T (t - \bar{t})(Y_t - \bar{Y})}{\frac{1}{T} \sum_{t=1}^T (t - \bar{t})^2} =$$

$$b = \frac{\frac{1}{T} \sum_{t=1}^T t Y_t - \bar{t} \cdot \bar{Y}}{\frac{1}{T} \sum_{t=1}^T t^2 - \bar{t}^2}$$

$$a = \bar{Y} - b\bar{t}$$



## Sezonski indeksi

$$SI_p = \frac{100 \cdot r}{\sum_{i=1}^r \sum_{p=1}^r Y_{ip}} \sum_{i=1}^r Y_{ip} \quad p = 1, 2, \dots, r$$

## 5. del

### Osnovni pojmi statističnega sklepanja Dvostransko ocenjevanje aritmetične sredine

$$\bar{Y} - z \cdot SE_{\bar{y}} < \bar{y} < \bar{Y} + z \cdot SE_{\bar{y}}$$

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n y_i \quad \text{vzorčna aritmetična sredina iz nerazvrščenih vrednosti}$$

$$\bar{Y} = \frac{1}{n} \sum_{k=1}^r f_k y_k \quad \text{vzorčna aritmetična sredina iz razvrščenih vrednosti}$$

$$SE_{\bar{y}} = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}} \quad \text{standardna napaka ocene aritmetične sredine}$$

$$z = 1,645 \Rightarrow P(\bar{Y} - 1,645 \cdot SE_{\bar{y}} < \bar{y} < \bar{Y} + 1,645 \cdot SE_{\bar{y}}) = 0,90, \alpha = 0,10$$

$$z = 1,96 \Rightarrow P(\bar{Y} - 1,96 \cdot SE_{\bar{y}} < \bar{y} < \bar{Y} + 1,96 \cdot SE_{\bar{y}}) = 0,95, \alpha = 0,05$$

$$z = 2,58 \Rightarrow P(\bar{Y} - 2,58 \cdot SE_{\bar{y}} < \bar{y} < \bar{Y} + 2,58 \cdot SE_{\bar{y}}) = 0,99, \alpha = 0,01$$

## Ocenjena variance

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{Y})^2$$

$$s^2 = \frac{1}{n-1} \sum_{k=1}^r f_k (y_k - \bar{Y})^2,$$

## Dvostransko ocenjevanje totala

$$\hat{Y} - z \cdot SE_Y < Y < \hat{Y} + z \cdot SE_Y$$

$$N \cdot \bar{Y} - z \cdot N \cdot SE_{\bar{y}} < Y < N \cdot \bar{Y} + z \cdot N \cdot SE_{\bar{y}}$$

## Dvostransko ocenjevanje strukturnega odstotka

$$p - z \cdot SE_{\pi} < \pi < p + z \cdot SE_{\pi},$$

$$p = 100 \frac{n_a}{n} \quad \text{strukturni odstotek iz vzorca}$$

$$SE_{\pi} = \sqrt{\frac{p(100-p)}{n}} \quad \text{standardna napaka ocene strukturnega odstotka}$$

### Kritične vrednosti za spremenljivko $z$

Stopnja tveganja $\alpha$	Enostransko ocenjevanje	Dvostransko ocenjevanje
10 %	1,28 (ali -1,28)	$\pm 1,645$
5 %	1,645 (ali -1,645)	$\pm 1,96$
1 %	2,33 (ali -2,33)	$\pm 2,58$

### Enostransko ocenjevanje statističnih parametrov

$$\bar{y} > \bar{Y} - z \frac{\sigma}{\sqrt{n}}$$

$$Y > \hat{Y} - z \cdot N \cdot \frac{\sigma}{\sqrt{n}}$$

$$\pi > p - z \cdot \sqrt{\frac{p(100-p)}{n}}$$

$$\bar{y} < \bar{Y} + z \frac{\sigma}{\sqrt{n}}$$

$$Y < \hat{Y} + z \cdot N \cdot \frac{\sigma}{\sqrt{n}}$$

$$\pi < p + z \cdot \sqrt{\frac{p(100-p)}{n}}$$

### Dvostransko ocenjevanje aritmetične sredine iz malih vzorcev

$$\bar{Y} - t_{\alpha/2} \frac{s}{\sqrt{n}} < \bar{y} < \bar{Y} + t_{\alpha/2} \frac{s}{\sqrt{n}},$$

### Preizkušanje domneve o aritmetični sredini

$$z = \frac{\bar{Y} - \bar{y}_D}{SE_{\bar{y}}}$$

$$SE_{\bar{y}} = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$$

$$t = \frac{\bar{Y} - \bar{y}_D}{SE_{\bar{y}}}$$

### Kritične vrednosti standardizirane spremenljivke

$$z_s = -1,645 < z < z_z = +1,645 \Rightarrow \alpha = 0,10$$

$$z_s = -1,96 < z < z_z = +1,96 \Rightarrow \alpha = 0,05$$

$$z_s = -2,58 < z < z_z = +2,58 \Rightarrow \alpha = 0,01$$

### Kritične vrednosti spremenljivke

$$\bar{y} = \bar{y}_D \pm z \cdot SE_{\bar{y}}$$



## Literatura in viri

- Aickin, M. (2010). *Variance and Covariance, Reliability and Regression: A Brief Companion of Formulas and Methods for Data Analysis*. ZDA: CreateSpace Independent Publishing Platform.
- Altschuld, J. (2009). *Needs Assessment Phase II: Collecting Data*. United Kingdom: SAGE Publications.
- Artenjak, J. (2003). *Poslovna statistika. Prenovljena in dopolnjena izdaja*. Maribor: Ekonomsko-poslovna fakulteta.
- Barde M.P., Barde P.J. (2012). What to use to express the variability of data: Standard deviation or standard error of mean? *Perspectives in Clinical Research*, 3(3), 113–6.
- Boncz, I. (2015). *Introduction to research methodology*. Institute of Health Insurance: Faculty of Health Sciences of the University of Pécs.
- Boudah, D. J. (2019). *Conducting Educational Research*. USA: East Carolina University.
- Brockwell, P. J., Davis, R. A. (2016). *Introduction to Time Series and Forecasting*. Springer Nature.
- Campbell M.J., Machin D., Walters S.J. (2007). *Medical Statistics: A text book for the health sciences*. Chichester: John Wiley & Sons.
- Chaudhuri, A. (2014). *Modern Survey Sampling*. USA: Taylor & Francis Group.
- Corder, G. W., Foreman, D. I. (2014). *Nonparametric statistics for non-statisticians: A step-by-step approach*. CA: Wiley.
- Creswel, J. W. (2014). *Research design: qualitative, quantitative, and mixed methods approaches*. London: Sage.
- Denzin, N. K., Lincoln, Y. S. (2018). *Handbook of Qualitative Research*. United Kingdom: SAGE Publications.
- Evans, M., Hastings, N., Peacock, B., Forbes, C. (2010). *Statistical Distributions*. ZDA: Wiley.
- Fabrigar, L. R., Duane, T. (2012). *Wegener Exploratory Factor Analysis*. Oxford University Press.
- Flick, U. (2018). *Handbook of Qualitative Data Collection: United Kingdom*. SAGE Publications.
- Freedman, D., Pisani, R., Purves, R. (2007). *Statistics*. New York: W.W. Norton & Company.
- Frost, J. (2020). *Introduction to Statistics: An Intuitive Guide for Analyzing Data and Unlocking Discoveries*. USA: Statistics by Jim Publishing.
- Ghauri, P., Grønhaug, K., Strange, R. (2020). *Research Methods in Business Studies*. United Kingdom: University Printing House.
- Given, L. M. (2008). *The SAGE Encyclopedia of Qualitative Research Methods*. United Kingdom: SAGE Publications.
- Golmajer, M. (2013). *Desezoniranje časovnih vrst*. Ljubljana: Statistični urad Republike Slovenije
- Guest, G., Namey, E. E., Mitchell, M. L. (2012). *Collecting Qualitative Data: A Field Manual for Applied Research*. United Kingdom: SAGE Publications.
- Hayton, J. C., Allen, D. G., Scarpello, V. (2004). Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis. *Organizational Research Methods*, 7(2), 191–205.
- Heeringa, S. G., Brady T. W., Berglund, P. A. (2010). *Applied Survey Data Analysis*. USA: Taylor & Francis Group
- Heumann, C., Schomaker, M., Shalabh, S. (2016). *Introduction to Statistics and Data Analysis*. Singapore: Springer.
- Holmes, A., Illowsky, B., Dean, S. (2018). *Introductory Business Statistics*. Houston: Rice University.
- Iacono, J., Palmer Brown, A., Holtham, C. (2009). Research Methods—a Case Example of Participant Observation. [https://www.researchgate.net/publication/228365089\\_Research\\_Methods-a\\_Case\\_Example\\_of\\_Participant\\_Observation](https://www.researchgate.net/publication/228365089_Research_Methods-a_Case_Example_of_Participant_Observation)
- Jenn N. C. (2006). Designing A Questionnaire. *Malaysian Family Physician* 30(1), 32–5.
- Johnnie, D. (2012). *Sampling Essentials: Practical Guidelines for Making Sampling Choices*. United Kingdom: SAGE Publications.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141–151.
- Kaliyadan F., Kulkarni V. (2019). Types of variables, descriptive statistics, and sample size. *Indian Dermatol Online Journal*, 10, 82–6.
- Levy, P. S., Lemeshow, S. (2008). *Sampling of Populations: Methods and Applications*. ZDA: Wiley.
- Lind, D. A., Marchal, W. G., Wathen, S. A. (2021). *Basic Statistics in Business and Economics*. New York: McGraw-Hill Education.
- Martin, W.E., Bridgmon, K. D. (2012). *Quantitative and Statistical Research Methods: From Hypothesis to Results*. CA: Jossey-Bass.
- Mishra P., Pandey C. M., Singh U., Gupta A., Sahu C., Keshri A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of Cardiac Anaesthesia*, 22(1), 67–72.
- Mišić, E. (2022). Splošno metodološko pojasnilo: indeksna števila in deflacioniranje. RS: Statistični urad.
- Montgomery, D. C., Peck, E. A., Vining, G. G. (2021). *Introduction to Linear Regression Analysis*. ZDA: Wiley.
- Moore, D. S., McCabe, G. P., Craig, B. A. (2016). *Introduction to the Practice of Statistics*. USA: W. H. Freeman.
- Pardede, P. (2018). Identifying and Formulating the Research Problem. [https://www.researchgate.net/publication/329179630\\_Identifying\\_and\\_Formulating\\_the\\_Research\\_Problem](https://www.researchgate.net/publication/329179630_Identifying_and_Formulating_the_Research_Problem)
- Patten, M. (2014). *Questionnaire Research: A Practical Guide*. USA: Taylor & Francis Group.
- Ralph, J., O'Neill, R., Winton, J. (2015). *A Practical Introduction to Index Numbers*. ZDA: Wiley.
- Roopa, S., Satya R. M. (2012). Questionnaire Designing for a Survey. *The Journal of Indian Orthodontic Society*, 46(4), 37–41.

- Russell Bernard, H. (2011). *Research Methods in Anthropology*. Maryland: Alta Mira
- Seber, G. A. F., Lee, A. J. (2003). *Linear Regression Analysis*. ZDA: Wiley
- Tabachnick, B. G., Fidell, L. S. (2013). *Using multivariate statistics*. Boston: Pearson.
- Thomopoulos, N. T. (2017). *Statistical Distributions*. USA: Springer.
- Thukral A.K., Bhardwaj R., Kumar V., Sharma A. (2019). New indices regarding the dominance and diversity of communities, derived from sample variance and standard deviation. *Heliyon*, 5(10), 1–16.
- Tominc, P. (2008). *Izbrana poglavja iz poslovne statistike*. Ekonomsko poslovna fakulteta Maribor, Univerza v Mariboru Maribor.
- Tominc, P. (2016). *Statistika (2. del predmeta)*. Učno gradivo pri predmetu Statistika (2. del predmeta), interno gradivo. Maribor: EPF.
- Tominc, P., Kramberger, T. (2007). *Statistične metode v logistiki*. Celje: UM Fakulteta za logistiko.
- Valliant, R., Dever, J. A., Kreuter, F. (2018). *Practical Tools for Designing and Weighting Survey Samples*. Singapore: Springer.
- Wolf, C., Joye, D., Smith, T. W., Fu, Y. (2016). *The SAGE Handbook of survey Methodology*. United Kingdom: SAGE Publications.

# STATISTIKA Z

## RAZISKOVALNIMI METODAMI

MAJA ROŽMAN, POLONA TOMINC

Univerza v Mariboru, Ekonomsko-poslovna fakulteta, Maribor, Slovenija  
maja.rozman@um.si, polona.tomic@um.si

V današnjem hitro spreminjajočem se poslovnem okolju je ključnega pomena sposobnost razumevanja in uporabe statistike za sprejemanje utemeljenih poslovnih odločitev. Napredovanje tehnologije in povečevanje obsega podatkovnih baz omogoča podrobno analizo in interpretacijo podatkov, ki so temelj za prepoznavanje ključnih prihodnjih trendov in oblikovanje strategij na tej osnovi. V pričujoči zbirki obravnavamo statistična metodološka orodja, ki pomagajo pri razumevanju vpliva različnih dejavnikov na posamezne ekonomske in poslovne kategorije in odločitve, analiziramo sezonske komponente v časovnih vrstah ter opredeljujemo napovedovanje vrednosti v prihodnjih časovnih enotah, analiziramo večdimenzionalne spremenljivke in pristopamo k statističnim konceptom preverjanja domnev o različnih lastnostih statistične množice na osnovi slučajnega vzorca. Z razumevanjem teh procesov lahko posamezniki oblikujejo učinkovite strategije, ki temeljijo na dejstvih in podatkih, ter se pripravijo na prihodnje izzive in izkoristijo dinamiko tržnih sprememb.

DOI

[https://doi.org/  
10.18690/um.epf.7.2024](https://doi.org/10.18690/um.epf.7.2024)

ISBN

978-961-286-929-8

**Ključne besede:**

statistične metode,  
deskriptivna statistika,  
inferenčna statistika,  
univariatna statistika,  
multivariatna statistika

**DOI**

[https://doi.org/  
10.18690/um.epf.7.2024](https://doi.org/10.18690/um.epf.7.2024)

**ISBN**

978-961-286-929-8

**Keywords:**

statistical methods,  
descriptive statistics,  
inferential statistics,  
univariate statistics,  
multivariate statistics

# STATISTICS WITH RESEARCH METHODS

MAJA ROŽMAN, POLONA TOMINC

University of Maribor, Faculty of Economics and Business, Maribor, Slovenia  
[maja.rozman@um.si](mailto:maja.rozman@um.si), [polona.tomic@um.si](mailto:polona.tomic@um.si)

In today's rapidly changing business environment, the ability to understand and use statistics for making informed business decisions is crucial. The advancement of technology and the increasing size of data repositories enable detailed analysis and interpretation of data, which are essential for identifying key future trends and forming strategies based on these insights. In this collection, we discuss statistical methodological tools that assist in understanding the impact of various factors on specific economic and business categories and decisions. We analyze seasonal components in time series, define forecasting of values in future time units, analyze multidimensional variables, and approach statistical concepts of hypothesis testing regarding different properties of a statistical set based on a random sample. By understanding these processes, individuals can design effective strategies based on facts and data, prepare for future challenges, and leverage the dynamics of market changes.









Univerza v Mariboru

---

Ekonomsko-poslovna fakulteta