



STATISTIČNA POSLOVNA ANALIZA

GRADIVA ZA SEMINARSKE IN LABORATORIJSKE VAJE

POLONA TOMINC

MAJA ROŽMAN



Univerzitetna založba
Univerze v Mariboru





Univerza v Mariboru

Ekonomsko-poslovna fakulteta

Statistična poslovna analiza

Gradiva za seminarske in laboratorijske vaje

Avtorici

Polona Tominc

Maja Rožman

November 2021

Naslov <i>Title</i>	Statistična poslovna analiza <i>Statistical Business Analysis</i>	
Podnaslov <i>Subtitle</i>	Gradiva za seminarske in laboratorijske vaje <i>Seminar and Laboratory Tutorials</i>	
Avtorici <i>Author</i>	Polona Tominc (Univerza v Mariboru, Ekonomsko-poslovna fakulteta)	
	Maja Rožman (Univerza v Mariboru, Ekonomsko-poslovna fakulteta)	
Recenzija <i>Review</i>	Vesna Čančer (Univerza v Mariboru, Ekonomsko-poslovna fakulteta)	
	Majda Bastič (Univerza v Mariboru, Ekonomsko-poslovna fakulteta)	
Jezikovni pregled <i>Language editing</i>	Alenka Plos	
Tehnični urednik <i>Technical editor</i>	Jan Perša (Univerza v Mariboru, Univerzitetna založba)	
Oblikovanje ovitka <i>Cover designer</i>	Jan Perša (Univerza v Mariboru, Univerzitetna založba)	
Grafične priloge <i>Graphic material</i>	Avtorici	
Založnik <i>Published by</i>	Univerza v Mariboru, Univerzitetna založba Slomškovo trg 15, 2000 Maribor, Slovenija https://press.um.si , zalozba@um.si	
Izdajatelj <i>Issued by</i>	Univerza v Mariboru, Ekonomsko-poslovna fakulteta Razlagova ulica 14, 2000 Maribor, Slovenija https://www.epf.um.si , epf@um.si	
Izdaja <i>Edition</i>	Prva izdaja	Izdano <i>Published at</i> Maribor, november 2021
Vrsta publikacije <i>Publication type</i>	E-knjiga	Dostopno na <i>Available at</i> https://press.um.si/index.php/ump/catalog/book/572

CIP - Kataložni zapis o publikaciji
Univerzitetna knjižnica Maribor

311(076)

TOMINC, Polona
Statistična poslovna analiza [Elektronski vir]
: gradiva za seminarske in laboratorijske vaje /
Polona Tominc, Maja Rožman. - 1. izd. - E-
publikacija. - Maribor : Univerza v Mariboru,
Univerzitetna založba, 2021

Način dostopa (URL) :
<https://press.um.si/index.php/ump/catalog/book/572>
ISBN 978-961-286-471-2
doi: 10.18690/978-961-286-471-2
COBISS.SI-ID 85932035



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

Besedilo / *Text* © Tominc in Rožman, 2021

To delo je objavljeno pod licenco Creative Commons Priznanje avtorstva-Nekomercialno-Deljenje pod enakimi pogoji 4.0 Mednarodna. / *This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International.*

Uporabnikom se dovoli 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/>

ISBN 978-961-286-471-2 (pdf)

DOI <https://doi.org/10.18690/978-961-286-471-2>

Cena
Price Brezplačni izvod

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

Citiranje
Attribution Tominc, P. in Rožman, M. (2021). *Statistična poslovna analiza: gradiva za seminarske in laboratorijske vaje*. Maribor: Univerzitetna založba. doi: 10.18690/978-961-286-471-2

Kazalo

1	Uvod	1
2	Opisna statistika in vzorčni pristop	3
2.1	Deskriptivna statistika	3
2.2	Slučajni vzorec, statistična množica in statistična enota.....	3
2.3	Merske lestvice	4
2.4	Osnovni statistični parametri in statistike.....	5
2.5	Vzorčenje in intervalno ocenjevanje parametrov	7
3	Naloge – opisna in inferenčna statistika.....	10
	Naloga 1	10
	Naloga 2	11
	Naloga 3	13
	Naloga 4	21
	Naloga 5	25
	Naloga 6	27
	Naloga 7	31
	Naloga 8	35
4	Preizkušanje domnev o statistično značilnih razlikah med povprečnimi vrednostmi spremenljivke v več vzorcih.....	41
	Naloga 9	42
	Naloga 10	44
	Naloga 11	46
	Naloga 12	48
	Naloga 13	50
	Naloga 14	52
	Naloga 14	54
5	Regresijska analiza	59
5.1	Enostavna linearna regresija	59
	Naloga 16	61
	Naloga 17	64
	Naloga 18	68
5.2	Multipla regresija	69
	Naloga 19	70
	Naloga 20	73
	Naloga 21	74
	Naloga 22	78
	Naloga 23	80

6	Študija primera - mobilno oglaševanje.....	85
6.1.	Vprašalnik za mobilno oglaševanje	85
6.2	Naloge.....	89
	Naloga 1	89
	Naloga 2	90
	Naloga 3	91
	Naloga 4	91
	Naloga 5	91
	Naloga 6	92
	Naloga 7	92
	Naloga 1	92
	Naloga 9	93
7	Naloge za seminarske vaje.....	95
	Naloga 1	95
	Naloga 2	95
	Naloga 3	96
	Naloga 4	96
	Naloga 5	96
	Naloga 6	96
	Naloga 7	97
	Naloga 8	97
	Naloga 9	97
	Naloga 10.....	97
	Naloga 11	97
	Naloga 12.....	98
	Naloga 13.....	98
	Naloga 14.....	98
	Naloga 15.....	98
	Naloga 16.....	99
	Naloga 17.....	99
	Naloga 18.....	100
	Naloga 19.....	100
	Naloga 20.....	102
	Naloga 21.....	103
	Naloga 22.....	104
	Naloga 23.....	104
	Naloga 24.....	105
8	Naloge za samostojno delo	109
	Naloga 1	109
	Naloga 2	109
	Naloga 3	109
	Naloga 4	110
	Naloga 5	110
	Naloga 6	110
	Naloga 7	111
	Naloga 8	112
	Naloga 9	112
	Naloga 10.....	113
	Naloga 11	114
	Naloga 12.....	115

Naloga 13.....	116
Naloga 14.....	116
Naloga 15.....	117
Naloga 16.....	118
Naloga 17.....	119
Naloga 18.....	120
Naloga 18.....	121
9 Rešitve računskih nalog	125
Naloga 5.....	125
Naloga 6.....	125
Naloga 7.....	128
Naloga 8.....	129
Naloga 15.....	131
Naloga 16.....	134
Naloga 18.....	138
Naloga 1.....	140
Naloga 2.....	140
Naloga 3.....	141
Naloga 4.....	142
Naloga 5.....	142
Naloga 6.....	143
Naloga 7.....	145
Naloga 18.....	147
Naloga 19.....	147
Tabele	149
Tabela naključnih števil.....	149
Ploščine $H(z)$ za standardizirano normalno porazdelitev.....	150
Kritične vrednosti za t porazdelitev.....	151
Literatura in viri.....	153



1 Uvod

Poslovne raziskave in analizo lahko opišemo kot sistematično in organizirano prizadevanje za raziskovanje določenega problema, ki se je pojavil v delovnem okolju in ga je treba rešiti. Obsega vrsto korakov, ki so zasnovani in izvedeni z namenom iskanja odgovorov na vprašanja, s katerimi se srečujejo odločevalci v delovnem okolju. To pomeni, da je prvi korak v poslovnem odločanju analiza in identifikacija tega, kje v organizaciji obstajajo problematična področja, ter čim bolj jasno in natančno prepoznati probleme, ki jih je treba preučiti in rešiti. Ko je težava jasno opredeljena, se lahko sprejmejo koraki za določitev dejavnikov, povezanih s poslovnim problemom, zbiranje informacij, analiziranje podatkov, razvijanje obrazložitve obravnavane težave in nato njeno reševanje s potrebnimi korektivnimi ukrepi (Sekaran in Bougie, 2014).

Gradivo Statistična poslovna analiza vsebuje naslednje vsebinske sklope:

- a) opisna statistika in vzorčni pristop,
- b) preizkušanje domnev o statistično značilnih razlikah med povprečnimi vrednostmi spremenljivke v več vzorcih,
- c) regresijska analiza.

Študenti v okviru predmeta spoznajo uporabnost statističnih metod pri reševanju poslovnih problemov ter utrdijo in nadgradijo teoretično znanje na področju statističnih tehnik in metod, ki omogočajo spremeniti različne podatke v uporabne informacije za poslovno odločanje. Študenti usvojijo analitičen matematično statističen pristop k preučevanju poslovnih problemov, ki se sestoji iz naslednjih korakov:

- formulacija problema na statističen način,
- izbira ustrezne statistične metode,
- reševanje problema,
- interpretacija rezultatov v smislu možnih rešitev problema.

V tej zbirki nalog¹ opisujemo nekatere statistične kvantitativne metode, ki predstavljajo orodje v poslovnih raziskavah, pri čemer uporabljamo programsko orodje IBM SPSS.

¹ Nekatere naloge so prirejene po zbirkah gradiv (interna gradiva), ki so se uporabljale v preteklosti (Leskovar Špacapan in Tominc, 2008/2009; Tominc 2008/2009).

2 Opisna statistika in vzorčni pristop

2.1 Deskriptivna statistika

Opisna ali deskriptivna statistika obsega metode in postopke opisa značilnosti danega podatkovnega niza. O opisnih statističnih analizah govorimo takrat, kadar predstavljamo npr. število podjetij, ki imajo sedež v neki regiji (število statističnih enot), kadar predstavljamo povprečno število zaposlenih v podjetjih neke gospodarske panoge (aritmetična sredina), kadar analiziramo, kako se podjetja v neki gospodarski panogi razlikujejo po ustvarjenem dobičku (variabilnost, ki jo analiziramo z varianco, standardnim odklonom) itd.

2.2 Slučajni vzorec, statistična množica in statistična enota

Statistična množica ali *populacija* (N) je množica statističnih enot, ki izpolnjujejo določene opredeljujoče kriterije oziroma lastnosti (Heumann idr., 2016). Primer statistične množice: državljani Republike Slovenije na dan 31.12. 2021, študenti Ekonomsko-poslovne fakultete v študijskem letu 2020/21.

Statistična enota je vsak posamezen element statistične množice, ki je predmet opazovanja v danem časovnem trenutku (npr. državljan Republike Slovenije na dan 31. 12. 2021, študent Ekonomsko-poslovne fakultete v študijskem letu 2020/21, ipd.). Statistična enota je pojav, ki je opredeljen s stvarnega (kateri pojav?), krajevnega (kje?) in časovnega vidika (kdaj?) (Heumann idr., 2016).

Slučajni ali verjetnostni vzorec (n) je del celotne statistične množice, sestavljajo pa ga statistične enote iz statistične množice. Za tak vzorec je značilno, da ima vsak element v populaciji znano in neničelno verjetnost, da ga vključimo oziroma izberemo v slučajni vzorec in ta verjetnost je vnaprej znana (Tominc in Kramberger, 2007). Pogoj za izvedbo verjetnostnega vzorčenja je obstoj vzorčnega okvira.

2.3 Merske lestvice

Obstajajo štiri osnovne vrste merskih lestvic oziroma spremenljivk, ki jih merimo: nominalna, ordinalna, intervalna in razmernostna (Sekaran in Bougie, 2014):

- Nominalna lestvica je tista, ki raziskovalcu omogoča, da predmete razvrsti v določene kategorije ali skupine. Na primer, glede spremenljivke spol lahko anketirance razvrstimo v dve kategoriji - moški in ženske. Tema dvema skupinama je mogoče dodeliti kodni številki 1 in 2. Ti številki sta enostavni in priročni oznaki kategorij brez lastne vrednosti, njun pomen je samo razdelitev anketirancev v eno od dveh kategorij, ki se ne prekrivata ali se medsebojno izključujeta.
- Z ordinalno lestvico spremenljivke ne le kategoriziramo (kot nominalna lestvica, ki označuje razlike med različnimi kategorijami), ampak kategorije tudi smiselno razvrstimo v nekem vrstnem redu, na primer od »najboljše do najslabše«, »prva do zadnja« ipd, in jih lahko oštevilčimo 1, 2 itd. Na primer, od anketirancev se lahko zahteva, da navedejo po pomenu od najmanj pomembne do najbolj pomembne, pet značilnosti na delovnem mestu, ki so pomembne zanjihovo zadovoljstvo z delovnim mestom.
- Na intervalni lestvici ali enaki intervalni lestvici predstavljajo numerično enake razdalje na skali enake vrednosti v merljivih lastnostih. Medtem ko uporabljamo nominalno lestvico za kvalitativno razlikovanje skupin, in ordinalno lestvico za razvrščanje kategorij po nekem vrstnem redu, intervalna lestvica omogoča tudi primerjavo razlik med vrednostmi. Razlika med katerimakoli dvema sosednjima vrednostma na lestvici je enaka razliki med katerimakoli dvema drugima sosednjima vrednostima lestvice. Termometer je dober primer takšne merske lestvice in velikost razlike med 37 stopinjami celzija (kar naj bi bila normalna telesna temperatura) in 38 stopinje je enaka velikosti razlike med 50 in 51 stopinjami.
- Razmernostna merska lestvica ne meri le velikosti razlik med točkami na lestvici, temveč tudi izmeri razmerja med razlikami. Tehnica mase je dober primer takega merila razmerja. Na njej je umerjen absolutni (in ne samovoljni) ničelni začetek, kar

nam omogoča izračun razmerja med težami dveh enot. Na primer, oseba, ki tehta 90 kilogramov, je dvakrat tako težka kot tista, ki tehta 45 kilogramov.

2.4 Osnovni statistični parametri in statistike

Mere osrednje tendence

Obstajajo tri osnovne mere osrednje tendence (Sekaran in Bougie, 2014): povprečje, mediana in modus. *Povprečje (aritmetična sredina)* je merilo osrednje tendence, ki ponuja splošno sliko o podatkovnem Na primer, proizvodni oddelek lahko vodi podrobne evidence o tem, koliko enot izdelka se proizvede vsak dan. Za oceno zalog surovin je vse, kar bi odločevalec morda želel vedeti, koliko enot na mesec je v povprečju proizvedel oddelek v zadnjih šestih mesecih. Ta mera - torej povprečje - lahko odločevalcu ponudi dobro predstavo o količini materialov, ki jih je treba nabaviti..

Mediana je središčnica v skupini opazovanj, kadar so statistične enote po vrednosti spremenljivke razporejena v naraščajočem ali padajočem vrstnem redu. V nekaterih primerih pa nabor opazovanj ni smiselno predstavljen niti z aritmetično sredino, niti z mediano, lahko pa podamo najpogostejšo vrednost spremenljivke – *modus*.

Mere razpršenosti – variabilnosti

Mere razpršenosti vključujejo *variacijski razmik*, *standardni odklon*, *varianco* (ki je izračunana na osnovi aritmetične sredine) in *kvartilni razmik* (ki je izračunan na osnovi mediane). Tako kot mere osrednje tendence so tudi te mere razpršenosti primerne za nominalne in intervalne podatke (Sekaran in Bougie, 2014).

Dva podatkovna niza imata lahko isto povprečje, vendar je razpršenost vrednosti lahko različna. Če je na primer podjetje A prodalo 30, 40 oziroma 50 enot izdelka v aprilu, maju oziroma juniju, podjetje B pa je v istem obdobju prodalo 10, 40 in 70 enot, je povprečna prodaja na mesec za obe podjetji enaka - 40 enot - vendar je variabilnost ali razpršenost v podjetju B večja. *Variacijski razmik* se nanaša na skrajne vrednosti v nizu opazovanj. Variacijski razmik je med 30 in 50 enot za podjetje A (razpršenost 20 enot), medtem ko je razmik med 10 in 70 enotami (razpršenost 60 enot) za podjetje B.

Varianca je povprečni kvadratni odklon (razlika) vednosti od aritmetične sredine, *standardni odklon* pa je preprosto kvadratni koren iz variance.

Aritmetična sredina in standardni odklon sta najpogostejša kazalnika opisne statistike za spremenljivke, merjene na intervalni ali razmernostni lestvici. Standardni odklon je v povezavi s srednjo vrednostjo zelo uporabno orodje zaradi naslednjih statističnih pravil, ki veljajo za normalno porazdelitev (Sekaran in Bougie, 2014):

1. Skoraj vsa opazovanja se nahajajo v simetričnem intervalu treh standardnih odklonov okoli povprečja (99,7 %).
2. Več kot 90 % opazovanj je v simetričnem intervalu znotraj dveh standardnih odklonov okoli srednje vrednosti.
3. Več kot polovica opazovanj je v simetričnem intervalu znotraj enega standardnega odklona okoli srednje vrednosti.

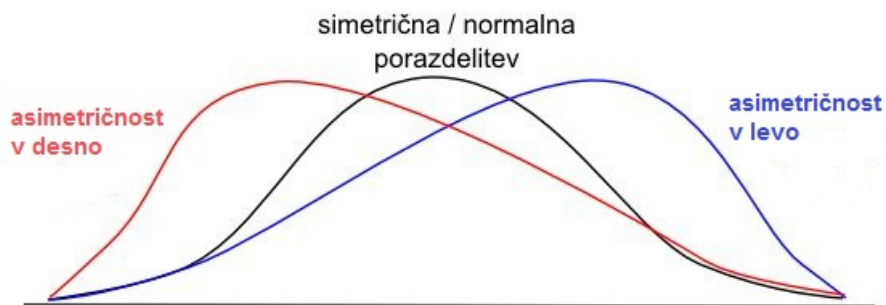
Asimetričnost in sploščenost

Asimetričnost in sploščenost merimo z merami asimetrije in merami sploščenosti (Landau in Everitt, 2004):

Asimetrične porazdelitve (ang. *skewness*) so lahko *asimetrične v desno* (pozitivna asimetrična porazdelitev), za katere je značilna večja gostitev pri manjših vrednostih spremenljivke ali *asimetrične v levo* (negativna asimetrična porazdelitev), za katere je značilna večja gostitev vrednosti pri večjih vrednostih spremenljivke. Koeficient asimetrije je manjši od nič, če je za porazdelitev spremenljivke značilna asimetrija v levo; pri asimetriji v desno je koeficient asimetrije večji od 0.

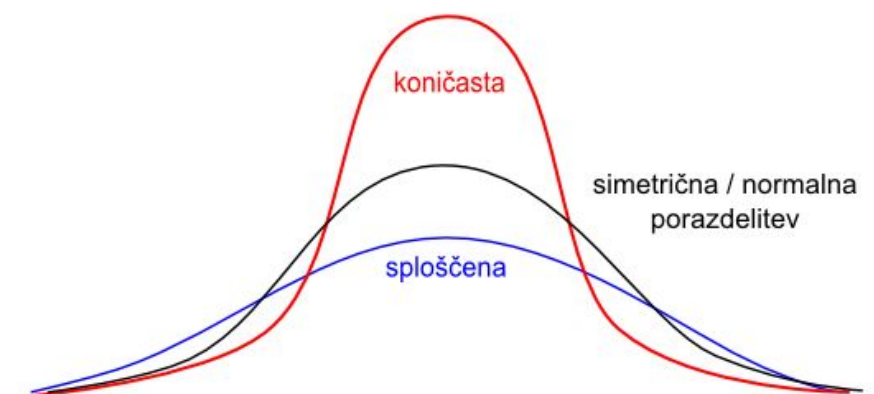
Sploščenost porazdelitve (ang. *kurtosis*) primerjamo z normalno porazdelitvijo, za katero rečemo, da je normalno sploščena. Če je porazdelitev bolj koničasta od normalne porazdelitve, rečemo, da je *porazdelitev koničasta*. Če je porazdelitev bolj sploščena od normalne, rečemo, da je *porazdelitev sploščena*. 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 normalni porazdelitvi sta koeficienta asimetričnosti in sploščenosti enaka 0.



Slika 1: Asimetrična porazdelitev

Vir: lasten



Slika 2: Sploščenost porazdelitve

Vir: lasten.

2.5 Vzorčenje in intervalno ocenjevanje parametrov

Vzorčenje

Za učinkovito načrtovanje slučajnih vzorcev se v praksi uporabljajo številne tehnike verjetnostnega vzorčenja. Pri enostavnem slučajnem vzorcu sta dve najpogosteje uporabljani metodi za slučajen izbor enot v vzorec, metoda z uporabo tablice slučajnih števil (v preteklosti) in z generatorjem slučajnih števil. Med najbolj razširjenimi so še sistematično vzorčenje, večstopenjsko vzorčenje in vzorčenje z verjetnostjo, ki je sorazmerno velikosti stratumov (podmnožic) statistične množice. Za slučajni vzorec je značilno, da je za vsako statistično enoto statistične množice zagotovljena enaka verjetnost, da je izbrana v vzorec. Verjetnost, da je posamezna enota izbrana v vzorec, je znana, oziroma jo je mogoče izračunati (Agresti in Finlay, 2009).

Na slučajnih vzorcih temelji statistična teorija vzorčenja, za katero je značilna možnost ocenjevanja kakovosti vzorčnih ocen statističnih parametrov z ustreznimi kazalci in opredelitev verjetnosti za nepravilni sklep, ki je sprejet na podlagi podatkov iz slučajnega vzorca. Statistična teorija torej temelji na slučajnih vzorcih, saj je z njimi zagotovljena objektivnost pri izbiri enot v opazovanje, reprezentativnost vzorca in možnost ugotavljanja kakovosti ocen za parametre, ki jih ocenjujemo iz vzorčnih podatkov (ibid).

Pri neslučajnem vzorčenju verjetnost, da je posamezna enota izbrana v vzorec ni znana. Izbor enote v vzorec poteka na osnovi subjektivne izbire. Metode neslučajnega vzorčenja so na primer: vzorčenje po principu kvot, priložnostno vzorčenje, vzorčenje po presoji, vzorčenje po principu kotaleče snežne kepe.

Poznamo naslednje metode slučajnega vzorčenja (slika 3):

Enostavno slučajno vzorčenje	Stratificirano slučajno vzorčenje	Sistematično vzorčenje
<ul style="list-style-type: none"> Vsak element v populaciji ima enako verjetnost za izbor v vzorec. 	<ul style="list-style-type: none"> Populacija je razdeljena v stratume tako, da je znotraj stratuma homogena glede na značilnost, na osnovi katere smo stratume oblikovali. Ločimo: proporcionalno stratificirano vzorčenje in disproporcionalno stratificirano vzorčenje. 	<ul style="list-style-type: none"> $N = 1000$ $n = 100$ * Izberemo vsak 10 element. - S pomočjo tablice slučajnih števil izberemo začetno število, ali pa izžrebamo začetno število izmed števil od 1 do 10. *Če je prvo izbrano število npr. 5, izbiramo nato 15, 25, 35,...-tega itd.

Slika 3: Metode slučajnega vzorčenja

Vir: lasten.

Celoten postopek izbire stratificiranega vzorca prikazuje spodnja preglednica (slika 4):

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
↓	
Disproporcionalna stratifikacija	Proporcionalna stratifikacija
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

Slika 4: Postopek izbire stratificiranega vzorca

Vir: Tominc, 2008/2009, str. 21

Intervalno ocenjevanje aritmetične sredine

Vrednost aritmetične sredine, ki jo izračunamo iz slučajnega vzorca, je točkovna ocena aritmetične sredine spremenljivke v statistični množici. S pomočjo točkovne ocene pa ocenjujemo vrednost statističnega parametra tudi z intervalno oceno, kar je način posplošitve rezultata iz vzorca na statistično množico. Intervalna ocena vključuje meje intervala, v katerem je z določeno stopnjo verjetnosti mogoče pričakovati, da se nahaja vrednost obravnavanega statističnega parametra. To je stopnja zaupanja (običajno 90 %, 95 % ali 99 %), interval pa imenujemo *interval zaupanja* (Rožman in Tominc, 2019).

Medtem ko pojem *natančnost* intervalne ocene označuje, kako natančno ocenjujemo parameter statistične množice na podlagi vzorčne ocene, *stopnja zaupanja* označuje, kako prepričani smo, da bodo naše ocene resnično veljale za statistično množico (Sekaran in Bougie, 2014).

3 Naloge – opisna in inferenčna statistika

Naloga 1

Raziskovalci so želeli analizirati delež žensk med zaposlenimi na dan 31. 12. 2021, za deset velikih podjetij v Sloveniji. V tabeli 1 so podani nekateri kazalniki opisne statistike. Pojasnite rezultate.

Tabela 1: Opisna statistika za spremenljivko delež žensk med zaposlenimi na dan 31. 12. 2021, desetih velikih podjetij (v tabelah so zapisi v angleščini – uporabljen je bil računalniški program SPSS).

N	Valid	10
	Missing	0
Mean		10,940
Median		11,250
Mode		13,1
Std. Deviation		2,1619
Variance		4,674
Range		6,6
Minimum		6,7
Maximum		13,3
Skewness		- 0,752
Kurtosis		- 0,091

V tabeli 1 so prikazani rezultati opisne statistike za spremenljivko delež žensk med zaposlenimi na dan 31. 12. 2021 desetih velikih podjetij. Vidimo, da je bilo v vzorec vključenih 10 velikih podjetij (N) in ni nobene manjkajoče vrednosti (Missing = 0).

Povprečni delež žensk med zaposlenimi na dan 31. 12. 2021 desetih velikih podjetij znaša 10,940 % vseh zaposlenih.

Mediana znaša 11,250 %, kar pomeni, da ima polovica (50 %) podjetij delež žensk med zaposlenimi na dan 31. 12. 2021 do 11,250 %, ter polovica (50 %) podjetij več kot 11,250 %.

Modus je 13,1 % in predstavlja tisto vrednost spremenljivke delež žensk med zaposlenimi na dan 31. 12. 2021v desetih velikih podjetij, ki se najpogosteje pojavlja.

Standardni odklon je kvadratni koren iz variance in v našem primeru znaša 2,162 %, ki opisuje razpršenost vrednosti spremenljivke okoli vrednosti aritmetične sredine.

Varianca znaša 4,674 (%)².

Variacijski razmik znaša 6,6 %-nih točk. Variacijski razmik je enak razliki med največjo (Maximum = 13,3 %) in najmanjšo (Minimum = 6,7 %) vrednostjo spremenljivke.

Naloga 2

Šest proizvodov podjetja je doseglo takšen obseg prodaje v zadnjem časovnem obdobju (v kosih): 841, 543, 654, 123, 543, 909.

a) Izračunajte in vsebinsko pojasnite aritmetično sredino ter standardni odklon (σ) in nepristransko oceno standardnega odklona (s).

b) Izračunajte standardno napako ocene aritmetične sredine in 80%-ni interval zaupanja za povprečno prodano število proizvodov in ga vsebinsko pojasnite. Opredelite pojem natančnosti ocene in ga vsebinsko pojasnite.

Potek za 80%-ni interval zaupanja za povprečno prodano število proizvodov: Kliknemo *Analyze*, nato *Descriptive statistics* in *Explore*. V desno okence *Dependent list* prenesemo spremenljivko *Prodaja* in kliknemo *Statistics*. Obkljukamo okence *Descriptives* in pod *Confidence interval for mean* napišemo 80 %. Kliknemo *Continue* in *OK*.

c) Izračunajte in vsebinsko pojasnite Q_1 , D_6 in C_{76} .

Potek: kliknemo *Analyze*, nato *Descriptive Statistics* in *Frequencies*. V okence *Dependent List* prenesemo spremenljivko *Prodaja* in kliknemo *Statistics* ter v okencu *Percentile Values* obkljukamo *Percentiles* ter vpišemo vrednosti, ki nas zanimajo, na primer 25 (za Q_1), 60 (za D_6), in 76 (za C_{76}). Kliknemo *Continue* in nato še *OK*.

d) Vse izračune preverite z SPSS-om.

Odgovori in izpisi rezultatov:

Tabela 2: Izpis delnih rezultatov z SPSS-om

		Statistic	Std. Error	
Prodaja	Mean	602,1667	114,08722	
	80% Confidence Interval for Mean	Lower Bound	433,7872	
		Upper Bound	770,5462	
	5% Trimmed Mean	611,7407		
	Median	598,5000		
	Variance	78095,367		
	Std. Deviation	279,45548		
	Minimum	123,00		
	Maximum	909,00		
	Range	786,00		
	Interquartile Range	420,00		
	Skewness	-,932	,845	
	Kurtosis	1,195	1,741	

V tabeli 2 vidimo, da spodnja meja (Lower Bound) intervala zaupanja znaša 433,7872 enot in zgornja meja (Upper Bound) znaša 770,5462 enot kar pomeni, da z 80 % verjetnostjo ocenjujemo, da se povprečni obseg prodaje šestih proizvodov podjetja giblje med 433,7872 in 770,5462 enot oziroma kosov.

Tabela 3: Izpis rezultatov za Q_1 , D_6 in C_{76}

N	Valid	6
	Missing	0
Percentiles	25	438,0000
	60	691,4000
	76	862,7600

Prvi kvartil ($Q_1 = 25\%$) znaša 438,00 enot, kar pomeni da je imelo 25 % proizvodov podjetja 438,00 enot prodaje ali manj.

Šesti decil ($D_6 = 60\%$) znaša 691,40 enot, kar pomeni da je imelo 60 % proizvodov podjetja 691,40 enot prodaje ali manj.

Šestinsedemdeseti centil ($C_{76} = 76\%$) znaša 862,76 enot, kar pomeni da je imelo 76 % proizvodov podjetja 862,76 enot prodaje ali manj.

Naloga 3

V tabeli 3 imamo podatke za $n = 15$ anketiranih oseb, ki so odgovarjale na 3 vprašanja:

1. Ali bi vas strah pred neuspehom odvrnil od tega, da bi ustanovili svoje podjetje?
2. Ali menite, da imate znanje, izkušnje in sposobnosti za ustanovitev novega podjetja?
3. Ali poznate nekoga, ki je v zadnjih dveh letih ustanovil podjetje?

Spol anketiranega je označen z 1 – moški in 2 – ženski.

Anketirani so odgovarjali na vprašanje z možnimi odgovori:

0 - NE

1 – DA

Tabela 4: Podatki anketiranih oseb

Spol	Starost	Strah pred neuspehom (1. vprašanje)	Znanje (2. vprašanje)	Poznanstvo (3. vprašanje)
1	45	0	1	1
1	45	0	1	1
1	58	1	1	1
1	62	1	1	1
1	64	0	1	0
1	55	0	1	1
1	57	0	0	0
2	20	1	1	1
2	23	1	0	1
2	20	0	1	0
2	21	1	0	0
2	33	1	1	1
2	25	1	0	1
2	33	0	1	1
2	30	1	0	1

a) Vnesite podatke v SPSS

b) Razdelite datoteko na 2 dela, glede na spol anketiranega in opravite ločeno analizo za moške in ženske o povprečnem odstotku tistih, ki bi jih strah pred neuspehom odvrnil od ustanovitve novega podjetja.

Potek: kliknemo na *Data* in nato *Split File* (ukaz *Split File* uporabljamo, kadar želimo rezultat obdelave pripraviti ločeno za skupine statističnih enot, na primer: ločeno za moški in ženski spol, ločeno po starostnih skupinah, ipd). Odpre se okence v katerem prenesemo spremenljivko spol in označimo »*organize output by groups*« - s tem potekom smo razdelili datoteko na dva dela glede na spol anketiranega. Opraviti še moramo ločeno analizo za moške in ženske o povprečnem odstotku tistih, ki bi jih strah pred neuspehom odvrnil od ustanovitve novega podjetja, kar naredimo v naslednjih korakih: kliknemo na *Analyze*, nato *Descriptive statistics* in *Frequencies* ali *Descriptives* ter prenesemo spremenljivko »*Strah pred neuspehom*«.

c) Opravite to analizo tako, da boste v izpisu dobili rezultat samo za ženske.

Potek: Kliknemo *Data* in nato *Select Cases* (ukaz *Select Cases* uporabimo, kadar želimo obdelavo izvesti samo za določeno skupino statističnih enot v datoteki podatkov, na primer: želimo obdelavo izvesti samo za ženski spol) označimo »*if condition is satisfied*« ter kliknemo na okence, ki piše »*if..*«. Prenesemo spremenljivko spol v desni okvirček ter napišemo: Spol=2 in kliknemo *Continue* in *OK*. Za ponovno analizo moramo izvesti korake za deskriptivno statistiko.

d) Oblikujte novo spremenljivko, ki bo imela negativno vrednost, če je oseba mlajša od 40 let in pozitivno vrednost, če je starejša od 40 let.

Potek: Kliknemo *Transform* in *Compute variable* (včasih potrebujemo novo spremenljivko, katere vrednosti oblikujemo iz obstoječih spremenljivk z ukazom *Compute*). V okence *Target Variable* napišemo novo ime za spremenljivko (na primer *starost1*), nato prenesemo spremenljivko *starost* (nahaja se v levem okencu) v desno okence, ki piše »*Numeric Expression*«, ter naredimo zapis: $Starost - 40$, nato kliknemo *OK*. V okencu *Data View* se prikaže negativna vrednost pri spremenljivki *Starost1* pri tistih osebah, ki so mlajše od 40 let in pozitivno vrednost imajo tiste osebe, ki so starejše od 40 let (primer prikaza smo prikazali v rešitvah pod točko d).

e) Dobljeno spremenljivko v nalogi d) pretvorite v novo spremenljivko, ki bo imela vrednost 0, če gre za osebo staro do 40 let in vrednost 1, če gre za osebo, ki je starejša od 40 let.

Potek: Kliknemo *Transform in Recode into different variables* (ta ukaz uporabimo, kadar želimo vrednosti neke spremenljivke preoblikovati oz. rekodirati). Novo spremenljivko, ki smo jo dobili pri nalogi d) prenesemo v desno okence, kjer piše »*Numeric Variable → Output Variable*«. V okencu *Output Variable* napišemo ime nove spremenljivke, ki jo bomo oblikovali tako, bo imela vrednost 0, če gre za osebo staro do 40 let in vrednost 1, če gre za osebo, ki je starejša od 40 let. Poimenovali jo bomo na primer *Starost2* ter kliknemo *Change*. Nato kliknemo okvirček *Old and new values*, v okence kjer piše »*Range, lowest trough value*« napišemo 0 in v okence kjer piše »*New value*« pod *Value* napišemo 0 ter kliknemo *add*. Nato kliknemo na »*All other values*« in pod »*New value*« v okence *Value* napišemo 1 ter kliknemo *add*. Obe spremembi se izpišeta v spodnjem okencu *Old → New*. Kliknemo *Continue*. V okencu *Data View* se prikaže nova spremenljivka, ki ima vrednost 0, če gre za osebo staro do 40 let in vrednost 1, če gre za osebo, ki je starejša od 40 let.

f) Izračunajte 80%-ni interval zaupanja za povprečno starost anketiranih oseb.

Potek: Kliknemo *Analyze*, nato *Descriptive statistics in Explore*. V desno okence *Dependent list* prenesemo spremenljivko *Starost* in kliknemo *Statistics*. Obkljukamo okence *Descriptives in* pod *Confidence interval for mean* napišemo 80 %. Kliknemo *Continue* in *OK*.

h) Razdelite datoteko v izpisu na tiste, ki menijo, da imajo znanje, izkušnje in sposobnosti in na tiste, ki menijo, da ga nimajo ter pojasnite povprečno starost v obeh skupinah.

Potek: Kliknemo *Data in Split file*, ki nam omogoča, da datoteko razdelimo na tiste, ki menijo, da imajo znanje, izkušnje in sposobnosti in na tiste, ki menijo, da ga nimajo. Nato kliknemo *Organize output by groups* in v desno okence prenesemo spremenljivko *Znanje*. Kliknemo *OK*. Moramo še analizirati povprečno starost v obeh skupinah, zato kliknemo *Analyze*, nato *Descriptive Statistics in Frequencies* (ali *Descriptives*). V desno okence prenesemo spremenljivko *Starost* in kliknemo okence *Statistics* ter označimo *Mean* (zanima nas povprečna starost).

i) Izračunajte vrednosti nove spremenljivke, ki naj bo seštevek vrednosti spremenljivk »znanje, izkušnje in sposobnosti« ter »poznavanje podjetnika«. Tako dobljeno spremenljivko prekodirajte tako, da bo imela vrednost 0, če posameznik meni, da nima znanja, izkušenj in sposobnosti za podjetništvo in da hkrati ne pozna nikogar, ki je v

zadnjih dveh letih ustanovil podjetje, ter vrednost 1, če posameznik meni, da ima znanja, izkušnje in sposobnosti za podjetništvo ali pozna nekoga, ki je v zadnjih dveh letih ustanovil podjetje.

Potek: Kliknemo *Transform* in nato *Compute variable*. V okvirček *Target variable* napišemo ime nove spremenljivke, na primer *Seštevek*. V okence *Numeric expression* napišemo: *znanje+poznanstvo*. Novo spremenljivko je treba prekodirati tako, da bo imela vrednost 0, če posameznik meni, da nima znanja, izkušenj in sposobnosti za podjetništvo in da hkrati ne pozna nikogar, ki je v zadnjih dveh letih ustanovil podjetje, ter vrednost 1, če posameznik meni, da ima znanja, izkušnje in sposobnosti za podjetništvo ali pozna nekoga, ki je v zadnjih dveh letih ustanovil podjetje. Navedeno naredimo tako, da kliknemo *Transform* in *Recode into different variable*. Novo spremenljivko *Seštevek* prenesemo v desno okence, kjer piše »*Numeric Variable* → *Output Variable*«. V okencu *Output Variable* napišemo ime nove spremenljivke *Seštevek2*, ki jo bomo oblikovali glede na nove vrednosti ter kliknemo *Change*. Nato kliknemo okvirček *Old and new values*, v okence kjer piše »*Range, lowest trough value*« napišemo 0 in v okence kjer piše »*New value*« pod *Value* napišemo 0 ter kliknemo *add*. Nato kliknemo na »*All other values*« in pod »*New value*« v okence *Value* napišemo 1 ter kliknemo *add*. Kliknemo *Continue*. V okencu *Data View* se prikaže nova spremenljivka.

Odgovori in izpisi rezultatov:

b) Prikaz izpisa za moške in ženske o povprečnem odstotku tistih, ki bi jih strah pred neuspehom odvrnil od ustanovitve novega podjetja (tabela 5, tabela 6):

Tabela 5: Deskriptivna statistika za moški spol

Statistics ^a		
Strah		
N	Valid	7
	Missing	0
Mean		,29
Std. Error of Mean		,184
Median		,00
Mode		0
Std. Deviation		,488
Variance		,238
Range		1
Minimum		0
Maximum		1

a. Spol = Moški

Strah pred neuspehom ^a					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Ne	5	71,4	71,4	71,4
	Da	2	28,6	28,6	100,0
	Total	7	100,0	100,0	

a. Spol = Moški

Tabela 6: Deskriptivna statistika za ženski spol

Statistics ^a		
Strah		
N	Valid	8
	Missing	0
Mean		,75
Std. Error of Mean		,164
Median		1,00
Mode		1
Std. Deviation		,463
Variance		,214
Range		1
Minimum		0
Maximum		1
a. Spol = Ženski		

Strah pred neuspehom ^a					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Ne	2	25,0	25,0	25,0
	Da	6	75,0	75,0	100,0
	Total	8	100,0	100,0	

a. Spol = Ženski

d) Prikaz postopka:

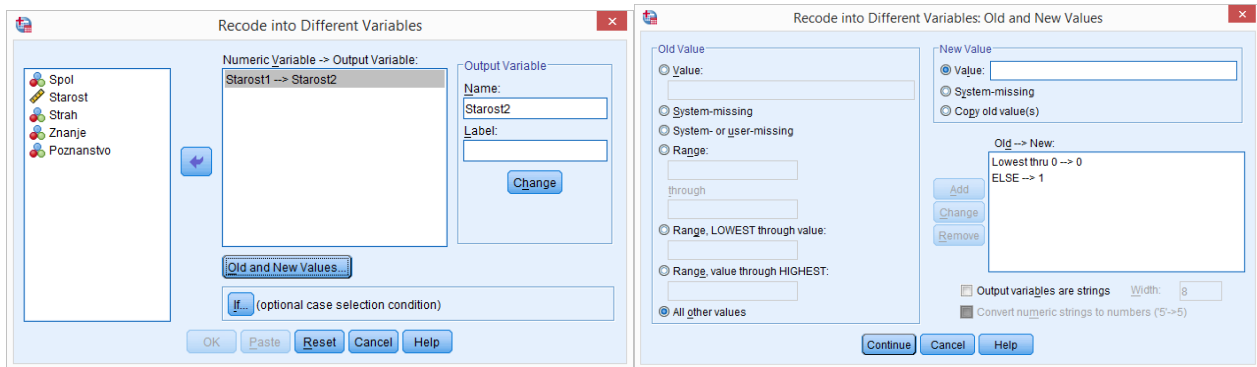
The screenshot shows the 'Compute Variable' dialog box in SPSS. The 'Target Variable' is 'Starost1' and the 'Numeric Expression' is 'Starost - 40'. The 'Function group' is set to 'All'. The 'Functions and Special Variables' list is empty. The 'if...' field is empty. The 'OK' button is highlighted.

Below the dialog box is the SPSS menu bar and toolbar. The data table below shows the results of the computation:

	Spol	Starost	Strah	Znanje	Poznanstvo	Starost1	var	var
1	1	45	0	1	1	5,00		
2	1	45	0	1	1	5,00		
3	1	58	1	1	1	18,00		
4	1	62	1	1	1	22,00		
5	1	64	0	1	0	24,00		
6	1	55	0	1	1	15,00		
7	1	57	0	0	0	17,00		
8	2	20	1	1	1	-20,00		
9	2	23	1	0	1	-17,00		
10	2	20	0	1	0	-20,00		
11	2	21	1	0	0	-19,00		
12	2	33	1	1	1	-7,00		
13	2	25	1	0	1	-15,00		
14	2	33	0	1	1	-7,00		
15	2	30	1	0	1	-10,00		
16								
17								

Slika 5: Prikaz poteka postopka

e) Prikaz postopka:



	Spol	Starost	Strah	Znanje	Poznanstvo	Starost1	Starost2	var
1	1	45	0	1	1	5,00	1,00	
2	1	45	0	1	1	5,00	1,00	
3	1	58	1	1	1	18,00	1,00	
4	1	62	1	1	1	22,00	1,00	
5	1	64	0	1	0	24,00	1,00	
6	1	55	0	1	1	15,00	1,00	
7	1	57	0	0	0	17,00	1,00	
8	2	20	1	1	1	-20,00	,00	
9	2	23	1	0	1	-17,00	,00	
10	2	20	0	1	0	-20,00	,00	
11	2	21	1	0	0	-19,00	,00	
12	2	33	1	1	1	-7,00	,00	
13	2	25	1	0	1	-15,00	,00	
14	2	33	0	1	1	-7,00	,00	
15	2	30	1	0	1	-10,00	,00	
16								
17								
18								
19								
20								
21								
22								
23								

Slika 6: Prikaz poteka postopka

f)

Tabela 7: 80 %-ni interval zaupanja za spremenjivko starost

		Statistic	Std. Error	
Starost	Mean	39,40	4,263	
	80% Confidence Interval for Mean	Lower Bound	33,67	
		Upper Bound	45,13	
	5% Trimmed Mean	39,11		
	Median	33,00		
	Variance	272,543		
	Std. Deviation	16,509		
	Minimum	20		
	Maximum	64		
	Range	44		
	Interquartile Range	34		
	Skewness	,236	,580	
Kurtosis	-1,656	1,121		

V tabeli 7 vidimo, da spodnja meja (Lower Bound) intervala zaupanja znaša 33,67 in zgornja meja (Upper Bound) znaša 45,13, kar pomeni, da z 80 % verjetnostjo ocenjujemo, da je povprečna starost anketiranih oseb med 33,67 let in 45,13 let.

h)

Tabela 8: Povprečna starost za osebe, ki menijo, da imajo znanje, izkušnje in sposobnosti

Statistics ^a		
Starost		
N	Valid	10
	Missing	0
Mean		43,50

a. Znanje = Da

V tabeli 8 vidimo, da povprečna starost za osebe, ki menijo, da imajo znanje, izkušnje in sposobnosti znaša 43,50 let.

Tabela 9: Povprečna starost za osebe, ki menijo, da nimajo znanje, izkušnje in sposobnosti

Statistics ^a		
Starost		
N	Valid	5
	Missing	0
Mean		31,20

a. Znanje = Ne

V tabeli 9 vidimo, da povprečna starost za osebe, ki menijo, da nimajo znanja, izkušnj in sposobnosti znaša 31,20 let.

Naloga 4

V tabeli 10 so podani podatki o številu okuženih prebivalcev s SARS-CoV-2 (COVID-19) za 11 držav na dan 20. 2. 2021.

Tabela 10: Število okuženih prebivalcev s SARS-CoV-2 po posameznih državah

Država	Število okuženih prebivalcev
Amerika	68.191
Velika Britanija	12.027
Brazilija	51.879
Španija	7.904
Turčija	7.419
Kolumbija	4.824
Indija	13.993
Rusija	12.953
Francija	23.170
Italija	15.470
Nemčija	9.164

Vir: WHO, 2021

- a) Opredelite statistično enoto in statistično spremenljivko.
- b) Podatke o okuženosti prebivalcev s SARS-CoV-2 na dan 20.2.2021 za enajst držav vnesite v program SPSS. Izračunajte in pojasnite rezultate opisne statistike za število okuženih prebivalcev s SARS-CoV-2 na dan 20.2.2021 za enajst držav.
- c) Na osnovi mere asimetričnosti in mere sploščenosti pojasnite ali so vrednosti navedene spremenljivke približno normalno porazdeljene.
- d) Izračunajte in 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*.

e) Izračunajte in pojasnite decilni in kvartilni razmik

Odgovori in izpisi rezultatov:

a)

Statistična enota: 1 država

Statistična spremenljivka: Število okuženih prebivalcev s SARS-CoV-2 na dan 20.2.2021 (numerična, nezvezna).

b in c)

Tabela 11: Opisna statistika za spremenljivko število okuženih prebivalcev s SARS-CoV-2 na dan 20.2.2021

N	Valid	11
	Missing	0
Mean		20635,82
Std. Error of Mean		6154,380
Median		12953,00
Mode		4824 ^a
Std. Deviation		20411,770
Variance		416640364,564
Skewness		1,821
Std. Error of Skewness		,661
Kurtosis		2,398
Std. Error of Kurtosis		1,279
Range		63367
Minimum		4824
Maximum		68191

a. Multiple modes exist. The smallest value is shown

V tabeli 11 so prikazani rezultati opisne statistike za spremenljivko *število okuženih prebivalcev s SARS-CoV-2 na dan 20.2.2021* za enajst držav.

Vidimo, da je bilo v vzorec vključenih 11 držav (n) in ni nobene manjkajoče vrednosti (Missing = 0).

Povprečna vrednost (ang. Mean) okuženih prebivalcev s SARS-CoV-2 na dan 20.2.2021 za enajst držav znaša 20.635,82 okuženih prebivalcev.

Standardna napaka ocene aritmetične sredine (ang. Standard error of mean) znaša 6.154,380 enot (okuženih prebivalcev) in pomeni, da manjša kot je njena vrednost, manjša je variabilnost med vzorčnimi povprečnimi vrednostmi in boljši predstavnik statistične množice je vzorec.

Mediana (ang. Median) znaša 12.953,00 enot, kar pomeni, da ima polovica (50 %) držav število okuženih prebivalcev s SARS-CoV-2 na dan 20.2.2021 manjšo ali enako od 12.953,00 okuženih prebivalcev ter polovica (50 %) držav, več kot 12.953,00 okuženih prebivalcev.

Modus (ang. Mode): zapisana vrednost je 4824 okuženih prebivalcev. Pri interpretaciji modusa moramo biti previdni. Modus namreč predstavlja tisto vrednost spremenljivke, ki je najpogosteje pojavlja. Pri spremenljivki *število okuženih prebivalcev s SARS-CoV-2 na dan 20.2.2021*, se vsaka vrednost spremenljivke pojavlja le enkrat (za vsako državo je vrednost različna). V tem primeru SPSS kot modus zapiše najmanjšo vrednost spremenljivke in pod tabelo dobimo označbo, da se več vrednosti spremenljivke pojavlja enako mnogokrat – v našem primeru je to sicer 1-krat, zato v tem primeru o pravem modusu ne moremo govoriti.

Standardni odklon (ang. Standard deviation) je kvadratni koren iz variance in v našem primeru znaša 20.411,770 okuženih prebivalcev ter opisuje razpršenost vrednosti spremenljivke okoli vrednosti aritmetične sredine. Varianca (ang. Variance) znaša 416.640.364,564 okuženih prebivalcev².

Koeficient asimetrije (ang. Skewness) znaša 1,821, kar pomeni, da je porazdelitev asimetrična v desno (pozitivna asimetrična porazdelitev). Koeficient sploščenosti (ang. Kurtosis) znaša 2,398, kar kaže na koničasto porazdelitev (pozitivna vrednost).

Variacijski razmik (ang. Range) znaša 63.367 enot in je enak razliki med največjo (Maximum = 68191 okuženih prebivalcev) in najmanjšo (Minimum = 4824 okuženih prebivalcev) vrednostjo spremenljivke.

d)

Tabela 12: Izračun decilov in kvartilov

N	Valid	11
	Missing	0
Percentiles	10	5343,00
	25	7904,00
	75	23170,00
	90	64928,60

Prvi decil ($D_1 = 10\%$) znaša 5.343, kar pomeni, da je imelo 10 % držav število okuženih prebivalcev s SARS-CoV-2 na dan 20.2.2021, 5.343 okuženih prebivalcev ali manj.

Deveti decil ($D_9 = 90\%$) znaša 64.928,60, kar pomeni, da je imelo 90 % držav število okuženih prebivalcev s SARS-CoV-2 na dan 20.2.2021, 64.928,60 okuženih prebivalcev ali manj.

Prvi kvartil ($Q_1 = 25\%$) znaša 7.904, kar pomeni da je imelo 25 % držav število okuženih prebivalcev s SARS-CoV-2 na dan 20.2.2021, 7.904 okuženih prebivalcev ali manj.

Tretji kvartil ($Q_3 = 75\%$) znaša 23.170, kar pomeni da je imelo 75 % držav število okuženih prebivalcev s SARS-CoV-2 na dan 20.2.2021, 23.170 okuženih prebivalcev ali manj.

e)

Kvartilni razmik: $Q = Q_3 - Q_1 = 23170 - 7904 = 15266$ okuženih prebivalcev

Odgovor: 50 % držav, ki glede na število okuženih prebivalcev s SARS-CoV-2 na dan 20.2.2021 ležijo na sredini ranžirne vrste, se razlikuje za največ 15.266 okuženih prebivalcev.

Decilni razmik: $D = D_9 - D_1 = 64928,60 - 5343 = 59585,6$ okuženih prebivalcev

Odgovor: 80 % držav, ki glede na število okuženih prebivalcev s SARS-CoV-2 na dan 20.2.2021 ležijo na sredini ranžirne vrste, se razlikuje za največ 59.585,6 okuženih prebivalcev.

Naloga 5

V tabeli 13 je prikazano doseženo število točk, ki so jih dosegli študenti ($n=10$) pri določenem predmetu na izpitu. Na izpitu je bilo možno doseči 40 točk.

Tabela 13: Doseženo število točk, ki so jih dosegli študenti ($n=10$) pri določenem predmetu na izpitu

Študent	Število točk
1	32
2	15
3	28
4	40
5	37
6	19
7	22
8	34
9	17
10	31

- Pojasnite povprečno vrednost, mediano in varianco za doseženo število točk, ki so jih dosegli študenti pri določenem predmetu na izpitu.
- Izračunajte 95 %-ni interval zaupanja za povprečno doseženo število točk, ki so jih dosegli študenti pri določenem predmetu na izpitu.
- Izračunajte delež standardnega odklona v aritmetični sredini (koeficient variabilnosti).

Odgovori in izpisi rezultatov:

a)

Tabela 14: Opisna statistika za spremenljivko doseženo število točk na izpitu za 10 študentov

N	Valid	10
	Missing	0
Mean		27,50
Std. Error of Mean		2,770
Median		29,50
Mode		15 ^a
Std. Deviation		8,759
Variance		76,722
Skewness		-,139
Std. Error of Skewness		,687
Kurtosis		-1,462
Std. Error of Kurtosis		1,334
Range		25
Minimum		15
Maximum		40

a. Multiple modes exist. The smallest value is shown

Povprečna vrednost: Povprečna vrednost doseženega števila točk za 10 študentov na izpitu znaša 27,50 točk.

Mediana: 50 % študentov ima doseženo število točk na izpitu manjšo ali enako od 29,50 točk ter 50 % študentov več kot 29,50 točk.

Varianca: Varianca znaša 76,722 (točk)².

b)

Tabela 15: 95 %-ni interval zaupanja za spremenljivko doseženo število točk na izpitu za 10 študentov

		Statistic	Std. Error	
Točke	Mean	27,50	2,770	
	95% Confidence Interval for Mean	Lower Bound	21,23	
		Upper Bound	33,77	
	5% Trimmed Mean	27,50		
	Median	29,50		
	Variance	76,722		
	Std. Deviation	8,759		
	Minimum	15		
	Maximum	40		
	Range	25		
	Interquartile Range	16		
	Skewness	-,139	,687	
	Kurtosis	-1,462	1,334	

c)

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

$$KV\% = \frac{8,759}{27,50} * 100 = 31,8509 \%$$

Odgovor: Delež standardnega odklona v aritmetični sredini znaša 31,8509 %.

Naloga 6

V času Covid-19 pandemije so v podjetju X želeli izmeriti zadovoljstvo svojih zaposlenih (n=12).

Zaposleni so na 7–stopenjski Likertovi lestvici označili svoje strinjanje od 1 do 7, kjer pomeni 1 – sploh se ne strinjam in 7 – popolnoma se strinjam.

Spol anketiranega je označen z 1 – moški in 2 – ženski.

Tabela 16 prikazuje odgovore zaposlenih o njihovem zadovoljstvu z delom.

Trditve so bile naslednje:

Trditev 1: Zadovoljen/na sem z delovnim časom in razporeditvijo delovnih obveznosti med Covid-19 pandemijo.

Trditev 2: Zadovoljen/na sem z ravnovesjem med zasebnim in poklicnim življenjem med Covid-19 pandemijo.

Trditev 3: Zadovoljen/na sem z vsemi prožnimi oblikami dela, ki mi jih omogoča podjetje med Covid-19 pandemijo.

Tabela 16: Odgovori zaposlenih o zadovoljstvu z delom

Starost	Spol	Trditev 1	Trditev 2	Trditev 3
25	1	6	6	6
33	1	7	7	6
46	1	5	6	4
52	2	6	4	5
29	2	4	5	5
43	2	5	3	4
57	2	4	5	7
61	1	3	5	2
55	1	5	7	5
36	1	7	7	6
27	2	6	4	7
58	2	5	5	4

a) Opravite ločeno analizo za moške in ženske in pojasnite povprečno stopnjo strinjanja z zadovoljstvom z ravnovesjem med zasebnim in poklicnim življenjem med Covid-19 pandemijo (trditev 2).

– Ukaz *Split file* uporabimo za ločeno analizo za moški in ženski spol.

b) Pojasnite mediano le za ženski spol za trditev 1: Zadovoljen/na sem z delovnim časom in razporeditvijo delovnih obveznosti med Covid-19 pandmeijo.

– Ukaz *Select cases* uporabimo, kadar želimo obdelavo podatkov izvesti le za določeno skupino statistični enot.

c) Spremenljivko starost pretvorite v novo spremenljivko (Starost2) tako, da bo imela vrednost 0, če gre za osebo staro do 50 let in vrednost 1, če gre za osebo, ki je starejša od 50 let. Za novo spremenljivko Starost2 prikažite frekvenčno preglednico in pojasnite rezultat.

– Ukaz *Recode* uporabimo takrat, kadar želimo vrednosti neke spremenljivke preoblikovati (rekodirati).

Ponovimo korake: Kliknemo *Transform* in *Recode into different variables*. Spremenljivko *Starost* prenesemo v desno okence, kjer piše »*Numeric Variable → Output Variable*«. V okencu *Output Variable* napišemo ime nove spremenljivke, ki jo bomo oblikovali tako, bo imela vrednost 0, če gre za osebo staro do 50 let in vrednost 1, če gre za osebo, ki je starejša od 50 let. Poimenovali jo bomo na primer *Starost2* ter kliknemo *Change*. Nato kliknemo okvirček *Old and new values*, v okence kjer piše »*Range, lowest trough value*« napišemo 50 in v okence kjer piše »*New value*« pod *Value* napišemo 0 ter kliknemo *add*. Nato kliknemo na »*All other values*« in pod »*New value*« v okence *Value* napišemo 1 ter kliknemo *add*. Kliknemo *Continue*. V okencu *Data View* se prikaže nova spremenljivka, ki ima vrednost 0, če gre za osebo staro do 50 let in vrednost 1, če gre za osebo, ki je starejša od 50 let.

Izračun rezultatov v frekvenčni preglednici: Kliknemo *Analyze* in nato *Frequencies*. Odpre se pogovorno okno, v katerem kliknemo novo spremenljivko *Starost2* in jo prenesemo v okence *Variable(s)* ter kliknemo *OK*.

d) Izračunajte variacijski razmik za stopnjo strinjanja zaposlenih s trditvijo 3: Zadovoljen/na sem z vsemi prožnimi oblikami dela, ki mi jih omogoča podjetje med Covid-19 pandemijo.

Odgovori in izpisi rezultatov:

a)

Tabela 17: Povprečna stopnja strinjanja s trditvijo 2 za moški spol

N	Valid	6
	Missing	0
Mean		6,33

Tabela 18: Povprečna stopnja strinjanja s trditvijo 2 za ženski spol

N	Valid	6
	Missing	0
Mean		4,33

b)

Tabela 19: Mediana za ženski spol za trditev 1

N	Valid	6
	Missing	0
Median		5,00

c)

Tabela 20: Frekvenčna preglednica za spremenljivko Starost2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	7	58,3	58,3	58,3
	1,00	5	41,7	41,7	100,0
	Total	12	100,0	100,0	

Tabela 20 kaže, da je vseh anketiranih oseb oziroma zaposlenih 12, kar predstavlja 100 %. Od vseh anketiranih zaposlenih v podjetju X ($n=12$) je 7 zaposlenih starih do 50 let (58,3 %) in 5 zaposlenih več kot 50 let (41,7 %).

d)

Tabela 21: Variacijski razmik za stopnjo strinjanja zaposlenih s trditvijo 3

N	Valid	12
	Missing	0
Range		5
Minimum		2
Maximum		7

Naloga 7

Anketirali smo 20 ljudi. Zanimala so nas njihova stališča o enakopravnosti položaja ženske v družbi, predvsem na delovnem mestu. Zanimalo nas je njihovo strinjanje s trditvami:

Trditev 1 – T1: Ženska naj bo zaposlena le, če ni poročena in če nima otrok.

Trditev 2 – T2: Če je ženska poročena, naj bo zaposlena le, če mož premalo zasluži ali če se z možem tako dogovorita.

Trditev 3 – T3: Ženske so na vodilnem položaju na delovnem mestu manj uspešne kot moški.

Trditev 4 – T4: Ženska na istem delovnem mestu kot moški, naj zasluži manj kot moški.

V tabeli 22 imamo rezultate anketiranja, kjer so anketirani na lestvici od 1 – v celoti se ne strinjam, do 7 – se v celoti strinjam, označili strinjanje s trditvami. Spol: 1 – moški, 2 – ženski

Tabela 22: Odgovori anketiranih oseb

Spol	Starost	T1	T2	T3	T4
1	21	5	4	3	3
1	24	3	4	2	2
1	26	2	2	3	3
1	32	1	2	2	1
1	34	1	2	3	1
1	32	5	5	4	3
1	25	3	3	2	2
1	44	2	3	3	2
2	41	1	1	1	1
2	39	2	2	2	2
2	44	2	2	2	2
2	35	1	1	1	1
2	21	1	2	1	1
2	24	2	3	1	3
2	36	1	2	1	2
2	35	4	5	5	1
2	42	3	5	5	2
2	35	3	5	5	3
1	41	1	2	3	5
1	44	2	1	2	1

- a) Izračunajte 90 %-ni interval zaupanja za povprečno starost anketiranih oseb in ga pojasnite.
- b) Analizirajte, kako se moški in ženske razlikujejo v povprečni stopnji strinjanja s posameznimi trditvami.
- c) Analizirajte zanesljivost merilnega inštrumenta.

Potek: Kliknemo *Analyze*, nato *Scale* in *Reliability Analysis*. V desno okence pod *Items* prenesemo vse trditve (T1, T2, T3 in T4) in nato *OK*.

Odgovori in izpisi rezultatov:

a)

Tabela 23: 90 %-ni interval zaupanja za povprečno starost anketiranih oseb

		Statistic	Std. Error	
Starost	Mean	33,75	1,765	
	90% Confidence Interval for Mean	Lower Bound	30,70	
		Upper Bound	36,80	
	5% Trimmed Mean	33,89		
	Median	35,00		
	Variance	62,303		
	Std. Deviation	7,893		
	Minimum	21		
	Maximum	44		
	Range	23		
	Interquartile Range	16		
	Skewness	-,289	,512	
Kurtosis	-1,223	,992		

b)

Tabela 24: Povprečna stopnja strinjanja s posameznimi trditvami za moški spol

		T1	T2	T3	T4
N	Valid	10	10	10	10
	Missing	0	0	0	0
Mean		2,50	2,80	2,70	2,30

Tabela 25: Povprečna stopnja strinjanja s posameznimi trditvami za ženski spol

		T1	T2	T3	T4
N	Valid	10	10	10	10
	Missing	0	0	0	0
Mean		2,00	2,80	2,40	1,80

c)

Tabela 26: Zanesljivost merilnega inštrumenta

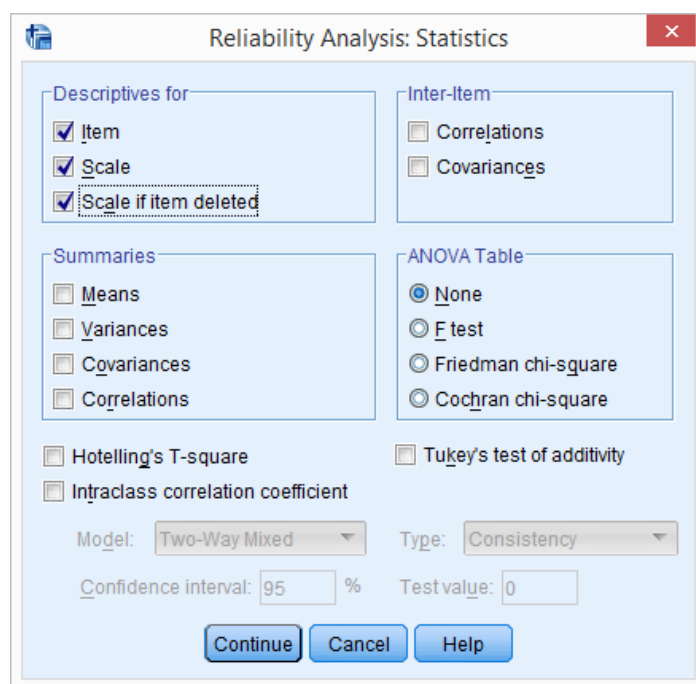
Reliability Statistics	
Cronbach's Alpha	N of Items
,817	4

Cronbachovo alfo uporabimo takrat, ko želimo preveriti, kako zanesljive so naše merske lestvice. Merimo jo z ustreznim kazalcem zanesljivost (reliability analysis), s pomočjo Cronbachovega kazalca ali koeficienta α .

Zanesljivost merjenja, ki ima koeficient $a \geq 0,80$, označimo kot zgledno, če je koeficient na intervalu $0,70 \leq a < 0,80$, kot zelo dobro, na intervalu $0,60 \leq a < 0,70$ kot zmerno, če je koeficient α manjši od 0,60, pa kot komaj sprejemljivo.

Tabela 26 kaže, da je vrednost Cronbachove alfe zgledna (0,817), zato lahko trdimo, da je merilni instrument oziroma vprašalnik zgledne zanesljivosti.

Lahko še dobimo podrobnejše informacije glede merilnega inštrumenta: Kliknemo *Analyze*, nato *Scale* in *Reliability Analysis*. V desno okence pod *Items* prenesemo vse trditve (T1, T2, T3 in T4) in kliknemo *Statistisc*. Odpre se pogovorno okno in pod *Descriptives for* obkljukamo *Item*, *Scale* in *Scale if item deleted* (pokaže nam, koliko bi se izboljšal oziroma spremenil Cronbachov alfa, če bi določeno trditev izločili).



Slika 7: Prikaz poteka postopka

Nato dobimo še dodaten izpis rezultata:

Tabela 27: Zanesljivost merilnega inštrumenta

Item-Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
T1: Ženska naj bo zaposlena le, če ni poročena in če nima otrok.	7,40	9,516	,726	,728
T2: Če je ženska poročena, naj bo zaposlena le, če mož premalo zasluži ali če se z možem tako dogovorita.	6,85	8,239	,844	,660
T3: Ženske so na vodilnem položaju na delovnem mestu manj uspešne kot moški.	7,10	9,253	,713	,733
T4: Ženska na istem delovnem mestu kot moški, naj zasluži manj kot moški.	7,60	13,516	,305	,893

V tabeli 27 vidimo, da se v primeru, če izločimo trditve T4, vrednost Cronbachove alfe poveča (z 0,817 na 0,893). V našem primeru ni potrebe po izločitvi trditve, saj je merilni instrument že zgledne zanesljivosti.

Naloga 8

Na nekem območju je 10 turistov sodelovalo v raziskavi glede pričakovanj turistov glede turistične ponudbe. V tabeli 28 so podatki o njihovi starosti, o maksimalnem znesku, ki so ga pripravljene odšteti za ceno polpenziona na turistični kmetiji ter o tem, ali želijo aktivno sodelovati na kulturnih dogodkih, festivalih ipd. področja, kjer so na počitnicah.

Tabela 28: Podatki anketirancev

Turist	Starost	Cena v EUR	Aktivno sodelovanje
1	45	35	DA
2	40	30	DA
3	23	25	NE
4	56	40	DA
5	54	42	DA
6	21	30	NE
7	67	50	DA
8	32	40	NE
9	43	35	NE
10	50	55	NE

- a) Ali si aktivnega sodelovanja na dogodkih in festivalih področja, kjer so na počitnicah, v povprečju bolj želijo mlajši ali starejši turisti? Na kakšne načine lahko analizirate to vprašanje?
- b) Pojasnite povprečen maksimalen znesek, ki so ga pripravljene odšteti za ceno polpenziona na turistični kmetiji tisti, ki si želijo aktivnega sodelovanja. Ali so tisti, ki si želijo aktivnega sodelovanja v teh dogodkih, v povprečju pripravljene odšteti tudi višjo ceno polpenziona na turistični kmetiji?
- c) Ali so obravnavane spremenljivke korelirane?

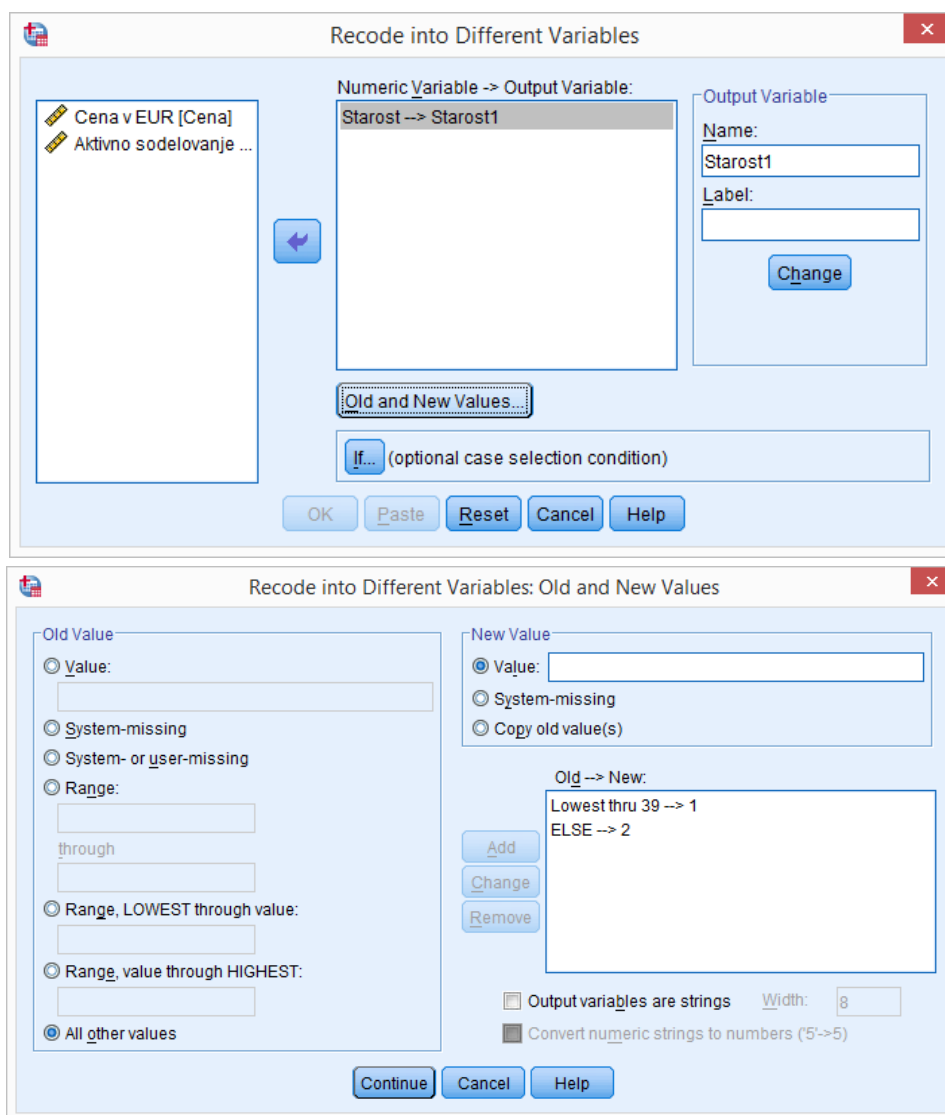
Postopek za izvedbo korelacije: Kliknemo *Analyze*, nato *Correlate* in nato *Bivariate*. V desno okence *Variables* prenesemo vse tri spremenljivke: *starost*, *cena*, *aktivno sodelovanje*. Pri *Correlation Coefficients* obkljukamo *Pearson* in *OK*.

Odgovori in izpisi rezultatv:

a)

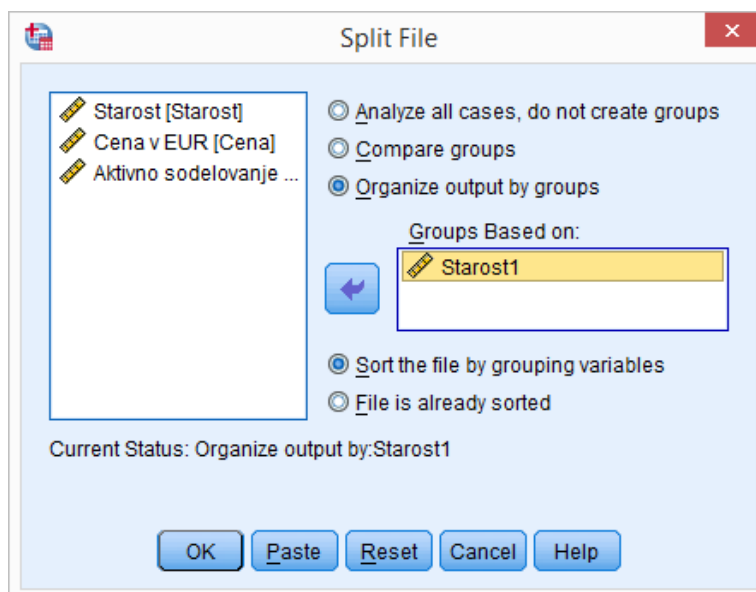
Prvo smo preoblikovali spremenljivko *Starost* v spremenljivko *Starost1*, ki ima vrednost 1, če je oseba mlajša od 40 let in vrednost 2, če je oseba starejša od 40 let (uporabili smo ukaz *recode into different variables*). Nato smo z ukazem *Split file* razdelili podatke na mlajše in starejše anketirance in izvedli analizo za spremenljivko aktivno sodelovanje.

Primer uporabe ukaza *recode into different variables*:



Slika 8: Prikaz poteka postopka

Primer uporabe ukaza *Split file*:



Slika 9: Prikaz poteka postopka

Tabela 29: Povprečna vrednost za osebe do 40 let

Statistics ^a		
Aktivno sodelovanje		
N	Valid	3
	Missing	0
Mean		1,33

a. Starost1 = 1

Tabela 30: Povprečna vrednost za osebe starejše od 40 let

Statistics ^a		
Aktivno sodelovanje		
N	Valid	7
	Missing	0
Mean		1,57

a. Starost1 = 2

b)

Uporabili smo ukaz *Split file*, da smo razdelili podatke na tiste, ki si želijo aktivnega sodelovanja v teh dogodkih in tiste, ki si ne želijo. Nato smo izvedli deskriptivno statistiko za povprečen maksimalen znesek, ki so ga pripravljene odšteti za ceno polpenziona na turistični kmetiji.

Tabela 31: Povprečen maksimalen znesek, ki so ga pripravljene odšteti tisti, ki si želijo aktivnega sodelovanja

Statistics ^a		
Cena		
N	Valid	5
	Missing	0
Mean		39,40

a. Aktivno sodelovanje = DA

Tabela 32: Povprečen maksimalen znesek, ki so ga pripravljene odšteti tisti, ki si ne želijo aktivnega sodelovanja

Statistics ^a		
Cena		
N	Valid	5
	Missing	0
Mean		37,00

a. Aktivno sodelovanje = NE

Preverimo še statistično značilnost povprečnih vrednosti:

One-Sample Test						
	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Aktivno sodelovanje	9,000	9	,000	1,500	1,12	1,88

Na osnovi rezultatov vidimo, da obstajajo statistično značilne razlike v povprečnem maksimalnem zneseku med tistimi, ki si želijo aktivnega sodelovanja v teh dogodkih in tistimi, ki si ga ne želijo.

c)

Tabela 33: Prikaz korelacije med spremenljivkami

Correlations				
		Starost	Cena	Aktivno sodelovanje
Starost	Pearson Correlation	1	,744*	-,669*
	Sig. (2-tailed)		,014	,034
	N	10	10	10
Cena	Pearson Correlation	,744*	1	-,137
	Sig. (2-tailed)	,014		,707
	N	10	10	10
Aktivno sodelovanje	Pearson Correlation	-,669*	-,137	1
	Sig. (2-tailed)	,034	,707	
	N	10	10	10

*. Correlation is significant at the 0.05 level (2-tailed).

Pearsonov koeficient meri linearno povezanost (korelacijo) med spremenljivkami.

Pearsonov koeficient med spremenljivko *starost* in *ceno* znaša 0,744, kar kaže na srednje močno povezanost med obema spremenljivkama. Stopnja značilnosti $p < 0,05$ (Sig. 0,014) kaže na to, da obstaja povezanost med spremenljivko *starost* in spremenljivko *ceno*.

Pearsonov koeficient med spremenljivko *starost* in *aktivno sodelovanje* znaša -0,669, kar kaže na srednje močno povezanost med obema spremenljivkama. Stopnja značilnosti $p < 0,05$ (Sig. 0,034) kaže na to, da obstaja povezanost med spremenljivkama.

4 Preizkušanje domnev o statistično značilnih razlikah med povprečnimi vrednostmi spremenljivke v več vzorcih

Za preizkušanje domnev o statistično značilnih razlikah med povprečnima vrednostma spremenljivke v dveh vzorcih uporabljamo parametrične in neparametrične teste.

Parametrični test za ugotavljanje statistično značilnih razlik med povprečnimi vrednostmi spremenljivk v dveh 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.

Za analizo značilnih razlik med povprečnima vrednostma spremenljivke v dveh odvisnih vzorcev bomo uporabili t-test za odvisna vzorca, med povprečnima vrednostma spremenljivke v dveh neodvisnih vzorcih pa t-test za neodvisna vzorca. Za analizo značilnih razlik med povprečnimi vrednostmi v več kot dveh neodvisnih vzorcev bomo uporabili analizo variance ANOVA.

Neparametrične teste 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). Za primerjavo dveh neodvisnih vzorcev se uporablja Mann-Whitneyev U test, za primerjavo dveh odvisnih vzorcev pa Wilcoxon signed rank test. V primeru, da je odvisna spremenljivka nominalna, pa lahko za primerjavo neodvisnih vzorcev uporabimo χ^2 – test.

V nadaljevanju se bomo ukvarjali s parametričnim testom. Za preizkušanje domnev o statistično značilnih razlikah med povprečnima vrednostma spremenljivke v dveh vzorcih bomo pri tem predmetu uporabili:

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

Naloga 9

Ugotoviti želimo, ali obstajajo statistično značilne razlike v povprečni porabi neke pijače na dan, med prebivalci toplejšega, primorskega dela nekega področja in prebivalci hladnejšega, gorskega dela. Izbrali smo 2 vzorca po 30 prebivalcev. Vzorčni podatki o porabi pijače na dan (v k.e.) so za vsakega od 30 prebivalcev zapisani v tabeli 34:

Tabela 34: Poraba pijače na dan v k.e.

Poraba prebivalcev gorskega dela			Poraba prebivalcev primorskega dela		
7	5	5	5	3	4
3	4	7	4	2	3
3	6	1	4	5	2
2	10	9	5	4	7
3	10	2	5	4	6
8	5	5	7	6	2
8	1	2	8	7	8
5	1	12	8	7	9
8	4	15	9	5	7
5	3	4	8	6	6

Za analizo podatkov bomo uporabili t-test za dva neodvisna vzorca.

Koraki: Kliknemo *Analyze* ter izberemo *Compare Means*, nato pa *Independent-Samples T Test*. Odvisno spremenljivko *povprečna poraba neke pijače na dan* prenesemo v desno okno *Test Variable(s)*. Nato spremenljivko *prebivalci* prenesemo v okno *Grouping Variable*. Kliknemo na *Define Groups* in definiramo vrednosti skupin: pri Group 1 vpišemo 1 (tj. prebivalci gorskega dela), pri Group 2 pa vpišemo 2 (tj. prebivalci primorskega dela) in kliknemo *Continue*. Kliknemo *OK*, da izvedemo t-test za dva neodvisna vzorca.

Zastavimo domnevi H0 in H1 o preverjanju statistično značilnih razlik v povprečni porabi neke pijače na dan med prebivalci toplejšega, primorskega dela in prebivalci hladnejšega, gorskega dela:

H0: V povprečni porabi neke pijače na dan ne obstajajo statistično značilne razlike med prebivalci toplejšega, primorskega dela in prebivalci hladnejšega, gorskega dela (H0: $\bar{y}_1 = \bar{y}_2$).

H1: V povprečni porabi neke pijače na dan obstajajo statistično značilne razlike med prebivalci toplejšega, primorskega dela in prebivalci hladnejšega, gorskega dela (H1: $\bar{y}_1 \neq \bar{y}_2$).

Odgovori in izpisi rezultatov - uporabili smo t-test za dva neodvisna vzorca:

Tabela 35: Rezultati t-testa za dva neodvisna vzorca.

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Povp. poraba neke pijače na dan	Equal variances assumed	4,994	,029	-,137	58	,891	-,100	,729	-1,560	1,360
	Equal variances not assumed			-,137	47,635	,892	-,100	,729	-1,567	1,367

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 stopnja značilnosti $p < 0,05$ (Sig. = 0,029), zato ničelno domnevo zavrnamo, torej enakih varianc ne predpostavimo (sprejmemo H1). V izpisu rezultatov t-testa za primerjavo povprečnih vrednosti neodvisnih vzorcev zato upoštevamo drugo vrstico izpisa. Odčitamo, da je stopnja značilnosti 0,892, kar nam pove, da je $p > 0,05$, torej ničelne domneve o enakosti povprečnih vrednosti ne zavrnamo: H0: v povprečni porabi neke pijače na dan ne obstajajo statistično značilne razlike med prebivalci toplejšega, primorskega dela in prebivalci hladnejšega, gorskega dela (sprejmemo domnevo H0: $\bar{y}_1 = \bar{y}_2$ in zavrnamo raziskovalno domnevo H1: $\bar{y}_1 \neq \bar{y}_2$).

Naloga 10

Raziskovalci so želeli ugotoviti ali obstajajo statistično značilne razlike v delovni zavzetosti med moškim in ženskim spolom v času Covid-19 epidemije. V vzorec je bilo vključenih 30 zaposlenih. Zaposleni so na 5–stopenjski Likertovi lestvici označili stopnjo strinjanja (kjer pomeni 1–sploh se ne strinjam in 5–popolnoma se strinjam) s trditvijo 1 (T1): Zaradi spremenjenih delovnih razmer v času Covid-19 epidemije je moja delovna zavzetost upadla.

Zaposleni so označili spol: 1-moški in 2-ženski.

Tabela 36 prikazuje stopnjo strinjanja zaposlenih po spolu:

Tabela 36: Odgovori anketirancev

Spol	Stopnja strinjanja s trditvijo T1
1	3
1	3
1	4
1	2
1	3
1	4
1	1
1	2
1	2
1	3
1	4
1	4
1	3
1	4
1	3
2	4
2	4
2	3
2	4
2	2
2	1
2	1
2	5
2	4
2	4
2	3
2	5
2	5
2	5
2	5

Odgovori in izpisi rezultatov:

Uporabili smo t-test za dva neodvisna vzorca:

Tabela 37: Povprečna stopnja strinjanja s trditvijo T1 po spolu

	Spol	N	Mean	Std. Deviation	Std. Error Mean
Zaradi spremenjenih delovnih razmer v času Covid-19 je moja delovna zavzetost upadla.	Moški spol	15	3,00	,926	,239
	Ženski spol	15	3,67	1,397	,361

Tabela 38: Rezultati t-testa za neodvisna vzorca – moški in ženski spol

		Levene's Test for Equality of Variances			t-test for Equality of Means					
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Zaradi spremenjenih delovnih razmer v času Covid-19 je moja delovna zavzetost upadla.	Equal variances assumed	2,932	,098	-1,540	28	,135	-,667	,433	-1,553	,220
	Equal variances not assumed			-1,540	24,306	,136	-,667	,433	-1,559	,226

Naloga 11

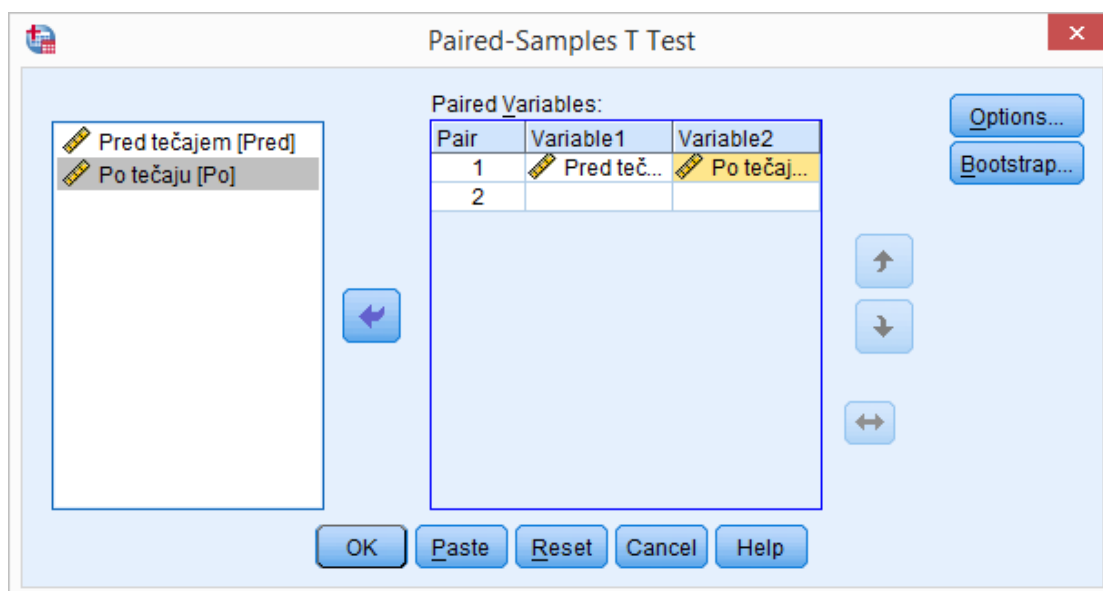
Želimo analizirati primernost izobraževalnega tečaja v nekem podjetju. Uspešnost merimo s številom opravljenih nalog v časovni enoti. V ta namen smo izbrali v vzorec 25 zaposlenih in merili število opravljenih nalog v časovni enoti in sicer preden so zaposleni obiskovali tečaj in po končanem tečaju. Torej imamo dva vzorca, v vsakem je 25 zaposlenih, vzorca pa sta odvisna, saj smo iste zaposlene anketirali pred in po tečaju. Vzorčni podatki o številu opravljenih nalog v časovni enoti so v tabeli 39.

Tabela 39: Vzorčni podatki o številu opravljenih nalog v časovni enoti

Pred	3	5	4	6	5	5	4	5	3	6	7	8	7	6	7	8	8	9	9	8	7	7	6	7	8
Po	7	8	6	7	8	9	6	6	7	8	8	7	9	10	9	9	8	8	4	4	5	6	9	8	12

Za analizo podatkov bomo uporabili t-test za dva odvisna vzorca.

Koraki: Kliknemo *Analyze* ter izberemo *Compare Means*, nato pa *Paired-Samples T Test*. Označimo spremenljivko *pred tečajem [pred]* in jo prenesemo v deno okence pod *Variable 1*, nato označimo spremenljivko *po tečaju [po]* in jo prenesemo v okence pod *Variable 2*. Kliknemo *OK*, da izvedemo t-test za dva odvisna vzorca:



Slika 10: Prikaz poteka postopka

Odgovori in izpisi rezultatov:

Tabela 40: t-test za dva odvisna vzorca

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Pred tečajem - Po tečaju	-1,200	2,449	,490	-2,211	-,189	-2,449	24	,022

Ugotoviti želimo uspešnost opravljenih nalog v časovni enoti preden so zaposleni obiskovali tečaj in po končanem tečaju. V ta namen bomo preverili naslednji domnevi:

H0: Povprečno število opravljenih nalog v časovni enoti pred obiskom tečaja je enako povprečnemu številu opravljenih nalog v časovni enoti po končanem tečaju (H0: $\bar{y}_1 = \bar{y}_2$)

H1: Povprečno število opravljenih nalog v časovni enoti pred obiskom tečaja ni enako povprečnemu številu opravljenih nalog v časovni enoti po končanem tečaju (H1: $\bar{y}_1 \neq \bar{y}_2$).

Iz rezultatov v zadnjem stolpcu je razvidno, da je stopnja značilnosti $p < 0,05$, torej ničelno domnevo zavrnamo in sprejememo raziskovalno domnevo (H1: $\bar{y}_1 \neq \bar{y}_2$).

V izpisu rezultatov še dobimo tabelo, ki se navezuje na korelacijo med spremenljivkama (tabela 41):

Tabela 41: Prikaz korelacije med spremenljivkama

		Paired Samples Correlations		
		N	Correlation	Sig.
Pair 1	Pred & Po	25	,051	,810

Zapišemo domnevi H0 in H1:

H0: korelacijski koeficient med spremenljivkama je enak 0.

H1: Korelacijski koeficient med spremenljivkama je različen od 0.

Iz tabele 41 je razvidno, da je stopnja značilnosti $p > 0,05$ (Sig. = 0,810), kar pomeni, da ničelno domnevo ne zavrnamo. Spremenljivki sta statistično značilno nekorelirani.

Naloga 12

V času Covid-19 epidemije se je pri nekaterih zaposlenih povečala stopnja stresa kar se odraža pri zmanjšanju njihove delovne učinkovitosti. Podjetje X je želelo ugotoviti ali so zaposleni ($n=20$) podvrženi stresu v času Covid-19 epidemije v primerjavi s časom pred Covid-19 epidemijo.

Zaposleni so na 5–stopenjski Likertovi lestvici označili stopnjo strinjanja (kjer pomeni 1–sploh se ne strinjam in 5–popolnoma se strinjam) z določeno trditvijo v dveh časovnih obdobjih (pred Covid-19 epidemijo in med Covid-19 epidemijo):

Trditev 1 (T1): Zaradi stresa na delovnem mestu opravi manj dela.

Analizirajte ali obstajajo statistično značilne razlike v stopnji strinjanja s trditvijo T1 pri zaposlenih podjetja X pred Covid-19 in med Covid-19 epidemijo.

Tabela 42 prikazuje stopnjo strinjanja zaposlenih v podjetju X v dneh časovnih obdobjih:

Tabela 42: Stopnja strinjanja zaposlenih v podjetju X v dneh časovnih obdobjih

Pred Covid-19 epidemijo	Med Covid-19 epidemijo
4	5
3	5
3	4
4	4
4	5
2	5
2	3
1	2
1	2
2	3
2	4
3	5
3	5
4	5
4	4
3	3
5	4
3	5
1	4
3	4

Odgovori in izpisi rezultatov:

Uporabili smo t-test za dva odvisna vzorca, saj so v vzorec zajeti zaposleni istega podjetja.

Tabela 43: Povprečna stopnja strinjanja zaposlenih s trditvijo pred Covid-19 in med Covid-19 epidemijo

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Pred Covid-19 epidemijo	2,85	20	1,137	,254
	Med Covid-19 epidemijo	4,05	20	,999	,223

Tabela 44: Rezultati t-testa za odvisna vzorca v podjetju X

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Pred Covid-19 epidemijo - Med Covid-19 epidemijo	-1,200	1,005	,225	-1,670	-,730	-5,339	19	,000

Naloga 13

Oblikovali smo tri različne oglaševalske akcije za nov mobilni telefon. Testirati želimo uspešnost teh treh oglaševalskih akcij in v ta namen smo izbrali 30 velikih trgovin. Razdelili smo jih v tri skupine po 10 trgovin in v vsaki skupini izvedli po eno oglaševalsko akcijo. Zabeležili smo prodajo po vsaki akciji. Podatki o prodaji v denarnih enotah v teh 30 trgovinah po oglaševalskih akcijah so v tabeli 45.

Tabela 45: Prodaja po oglaševalni akciji v d.e.

Prodaja po prvi akciji	Prodaja po drugi akciji	Prodaja po tretji akciji
81	88	89
85	86	91
72	88	96
50	85	87
77	79	89
98	82	90
77	84	89
66	78	96
65	85	97
67	93	94

Za analizo podatkov bomo uporabili ANOVO (analiza varince), saj imamo tri različne oglaševalske akcije.

Koraki: Kliknemo *Analyze*, nato *Compare Means*, ter *One-way ANOVA*. Odpre se okno *One-way ANOVA*, v katerem sta zapisani obe spremenljivki (*akcija*, *prodaja*). Spremenljivko *prodaja* prenesemo v okence *Dependent List*. Spremenljivko *akcija* prenesemo v okence *Factor*. Kliknemo na *Options*, nato na *Descriptive* in potem *Continue*.

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*. Nato kliknemo *OK*.

Odgovori in izpisi rezultatov:

Oblikovali smo domnevi:

H0: Povprečna prodaja novega mobilnega telefona je enaka v vseh treh skupinah.

H1: Povprečna prodaja novega mobilnega telefona se razlikuje vsaj v dveh skupinah.

Tabela 46: Opisna statistika – Prodaja novega mobilnega telefona posamezni oglaševalni akciji (skupini)

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Prva akcija	10	73,80	13,071	4,133	64,45	83,15	50	98
Druga akcija	10	84,80	4,442	1,405	81,62	87,98	78	93
Tretja akcija	10	91,80	3,615	1,143	89,21	94,39	87	97
Total	30	83,47	10,954	2,000	79,38	87,56	50	98

Tabela 47: Rezultati analize variance (ANOVA)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1646,667	2	823,333	12,129	,000
Within Groups	1832,800	27	67,881		
Total	3479,467	29			

Iz izpisa rezultatov analize variance v tabeli 47 je razvidno, da je stopnja značilnosti $p < 0,05$ (Sig. = 0,000), zato ničelno domnevo zavrnamo in sprejmemo raziskovalno domnevo: H1: Povprečna prodaja novega mobilnega telefona se razlikuje vsaj v dveh skupinah.

Naslednja tabela 48 prikazuje katere skupine se med seboj značilno razlikujejo.

Tabela 48: Rezultati post-hoc testa pri analizi variance

(I) Akcija	(J) Akcija	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Prva akcija	Druga akcija	-11,000*	3,685	,016	-20,14	-1,86
	Tretja akcija	-18,000*	3,685	,000	-27,14	-8,86
Druga akcija	Prva akcija	11,000*	3,685	,016	1,86	20,14
	Tretja akcija	-7,000	3,685	,158	-16,14	2,14
Tretja akcija	Prva akcija	18,000*	3,685	,000	8,86	27,14
	Druga akcija	7,000	3,685	,158	-2,14	16,14

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

Iz tabele 48 je razvidno, da obstajajo statistično značilne razlike med prodajo po prvi (Sig. = 0,016; $p < 0,05$) in drugi (Sig. = 0,000; $p < 0,05$) oglaševalni akciji ter med prodajo po prvi in tretji (Sig. = 0,000; $p < 0,05$) oglaševalni akciji.

Naloga 14

V tabeli 49 so podatki za 20 družinskih podjetij o tem, katera generacija družine ima trenutno v lasti podjetje in podatek o tem, kolikšno je število zaposlenih družinskih članov v družinskem podjetju.

Tabela 49: Generacija, ki ima v lasti podjetje in število zaposlenih družinskih članov

Generacija, ki ima v lasti podjetje	Število zaposlenih družinskih članov
1	3
2	4
3	4
2	5
3	4
1	3
1	2
1	3
3	4
1	4
2	3
1	2
2	2
1	1
2	1
1	1
2	3
1	1
1	2
1	1

- Vnesite podatke v SPSS
- Analizirajte ali obstajajo statistično značilne razlike v povprečnem številu zaposlenih družinskih članov v podjetjih ustanoviteljske (prve) generacije, in podjetji, ki jih imajo v lasti člani naslednjih generacij.

Izpisi rezultatov:

Tabela 50: Opisna statistika – Število zaposlenih družinskih članov po posamezni generaciji

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1 generacija	11	2,09	1,044	,315	1,39	2,79	1	4
2 generacija	6	3,00	1,414	,577	1,52	4,48	1	5
3 generacija	3	4,00	,000	,000	4,00	4,00	4	4
Total	20	2,65	1,268	,284	2,06	3,24	1	5

Tabela 51: Rezultati analize variance (ANOVA)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	9,641	2	4,820	3,919	,040
Within Groups	20,909	17	1,230		
Total	30,550	19			

Tabela 52: Rezultati post-hoc testa pri analizi variance

(I) Generacija	(J) Generacija	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1 generacija	2 generacija	-,909	,563	,266	-2,35	,53
	3 generacija	-1,909*	,722	,043	-3,76	-,06
2 generacija	1 generacija	,909	,563	,266	-,53	2,35
	3 generacija	-1,000	,784	,428	-3,01	1,01
3 generacija	1 generacija	1,909*	,722	,043	,06	3,76
	2 generacija	1,000	,784	,428	-1,01	3,01

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

Naloga 14

Družbo za upravljanje (DZU) zanima, ali se njihove stranke, ki vlagajo v njihove sklade različno visoke zneske, med seboj razlikujejo po starosti. Glede na višino zneskov, ki jih ljudje vlagajo v sklade, so vse vlagatelje razdelili v 3 skupine, ter primerjali vložene zneske v sklade glede na starostne skupine. Podatki so v tabeli 53.

- Izračunajte in pojasnite 95 %-ni interval za povprečni znesek vlagateljev v vsaki posamezni skupini.
- Ali se vlagatelji v teh treh skupinah med seboj statistično značilno razlikujejo po njihovi starosti?

Tabela 53: Podatki o strankah

Skupina	Znesek (v d.e.)	Starost
1	123	21
1	234	34
1	234	32
1	231	21
1	213	65
1	432	32
1	321	21
1	231	32
1	223	21
1	122	20
1	432	34
1	431	44
1	444	32
2	570	56
2	789	45
2	768	43
2	786	55
2	789	43
2	987	54
2	675	54
2	567	65
2	768	65
2	789	54
2	999	46
2	876	56
3	1234	65
3	1234	66
3	456	62
3	7654	63
3	4345	65

Izpis rezultatov:

a)

Uporabili smo ukaz *Split file*, da smo lahko rezultat obdelave pripravili ločeno za vse tri skupine.

Tabela 54: 95 %-ni interval zaupanja za povprečni znesek vlagateljev v prvi skupini

			Statistic	Std. Error
Znesek	Mean		282,38	32,413
	95% Confidence Interval for Mean		Lower Bound	211,76
			Upper Bound	353,01
	5% Trimmed Mean		282,32	
	Median		234,00	
	Variance		13658,090	
	Std. Deviation		116,868	
	Minimum		122	
	Maximum		444	
	Range		322	
	Interquartile Range		214	
	Skewness		,311	,616
	Kurtosis		-1,309	1,191

Tabela 55: 95%-ni interval zaupanja za povprečni znesek vlagateljev v drugi skupini

			Statistic	Std. Error
Znesek	Mean		780,25	39,031
	95% Confidence Interval for Mean		Lower Bound	694,34
			Upper Bound	866,16
	5% Trimmed Mean		779,94	
	Median		787,50	
	Variance		18280,568	
	Std. Deviation		135,206	
	Minimum		567	
	Maximum		999	
	Range		432	
	Interquartile Range		156	
	Skewness		,001	,637
	Kurtosis		-,098	1,232

Tabela 56: 95%-ni interval zaupanja za povprečni znesek vlagateljev v tretji skupini

			Statistic	Std. Error
Znesek	Mean		2984,60	1344,935
	95% Confidence Interval for Mean		Lower Bound	-749,54
			Upper Bound	6718,74
	5% Trimmed Mean		2865,67	
	Median		1234,00	
	Variance		9044250,800	
	Std. Deviation		3007,366	
	Minimum		456	
	Maximum		7654	
	Range		7198	
	Interquartile Range		5155	
	Skewness		1,186	,913
	Kurtosis		,229	2,000

b)

Tabela 57: Opisna statistika – Starost vlagateljev

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1 skupina	13	31,46	12,467	3,458	23,93	39,00	20	65
2 skupina	12	53,00	7,544	2,178	48,21	57,79	43	65
3 skupina	5	64,20	1,643	,735	62,16	66,24	62	66
Total	30	45,53	16,068	2,934	39,53	51,53	20	66

Tabela 58: Rezultati analize variance (ANOVA)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4985,436	2	2492,718	26,900	,000
Within Groups	2502,031	27	92,668		
Total	7487,467	29			

Tabela 59: Rezultati post-hoc testa pri analizi variance

(I) Skupina	(J) Skupina	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1 skupina	2 skupina	-21,538*	3,854	,000	-31,09	-11,98
	3 skupina	-32,738*	5,066	,000	-45,30	-20,18
2 skupina	1 skupina	21,538*	3,854	,000	11,98	31,09
	3 skupina	-11,200	5,124	,092	-23,90	1,50
3 skupina	1 skupina	32,738*	5,066	,000	20,18	45,30
	2 skupina	11,200	5,124	,092	-1,50	23,90

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

5 Regresijska analiza

5.1 Enostavna linearna regresija

S pomočjo enostavne linearne regresije analiziramo odvisnost med eno odvisno (y) in eno neodvisno ali pojasnjevalno spremenljivko (x_1) (če so izpolnjeni pogoji za izvedbo regresijske analize).

Kazalci enostavne linearne regresije so (Bastič, 2006):

- Ocenjeni vrednosti obeh regresijskih koeficientov (pri regresijski konstanti in pri pojasnjevalni spremenljivki).
- Determinacijski koeficient (r^2_{xy}): pove, kolikšen % celotne variance spremenljivke y (odvisna spremenljivka) je razložen z regresijsko funkcijo oz. s spremenljivko x_1 (neodvisna spremenljivka). Opredeljuje jakost linearne povezanosti med spremenljivkama. Vrednost determinacijskega koeficienta se giblje med 0 in 1 ($0 \leq r^2_{xy} \leq 1$).
- Korelacijski koeficient (r_{xy}): opredeljuje jakost in smer linearne povezanosti med odvisno in neodvisno spremenljivko. Vrednost korelacijskega koeficienta se giblje med -1 in 1 ($-1 \leq r_{xy} \leq 1$).

Moč linearne povezanosti med spremenljivkama glede na vrednost korelacijskega in determinacijskega koeficienta je prikazana v tabeli 60.

Tabela 60: Moč linearne povezanosti glede na vrednost korelacijskega in detereminacijskega koeficienta

Korelacijski koeficient (r_{xy})	Determinacijski koeficient (r^2_{xy})	Moč linearne povezave
0	0	Korelacije ni
0–0,5	0–0,25	Šibka 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

- Standardna napaka ocene odvisne spremenljivke (σ_e): pokaže, ali na variabilnost spremenljivke y , razen spremenljivke x_1 , vplivajo še druge spremenljivke in slučajni vplivi.
- Regresijski koeficient b_0 (konstanta) pove, povprečno vrednost odvisne spremenljivke, če je vrednost neodvisne spremenljivke enaka 0.
- Regresijski koeficient b_1 pove, za koliko enot se v povprečju spremeni vrednost odvisne spremenljivke, če se neodvisna spremenljivka spremeni (poveča ali zmanjša – gledamo predznak) za eno enoto.

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_1 testiramo s t-testom, pri čemer preizkušamo domnevo:

H0: Regresijski koeficient β_1 je enak 0 ($\beta_1 = 0$)

H1: Regresijski koeficient β_1 ni enak 0 ($\beta_1 \neq 0$).

Naloga 16

Zavod republike Slovenije za zaposlovanje v okviru programov spodbujanja zaposlovanja delodajalcem omogoča različne subvencije oziroma delno povračilo stroškov za zaposlitev novih zaposlenih (na primer, subvencija za skrajšani polni delovni čas, subvencija za zaposlitev brezposelnih, ki so starejši od 50 let, subvencija za zaposlitev brezposelnih mladih pred dopolnjenim 30. letom, davčna olajšava za zaposlovanje brezposelnih, davčna olajšava za zaposlovanje invalidov, ipd). Raziskovalci so želeli ugotoviti ali podjetja, ki se poslužujejo različnih subvencij zaposlovanja zaposlujejo več zaposlenih. Analizirajte ali je število novo zaposlenih (y) odvisno od števila različnih subvencij zaposlovanja (x), ki se jih poslužujejo podjetja. Izvedite enostavno linearno regresijo in pojasnite rezultate. Podatki so v tabeli 61.

Tabela 61: Podatki o številu različnih subvencij zaposlovanja in številu novo zaposlenih

Podjetje	Število uveljavljenih subvencij	Število novo zaposlenih
A	1	3
B	3	6
C	5	11
D	4	8
E	2	4
F	6	15
G	5	13

Postopek – izvedba enostavne linearne regresije: Kliknemo na *Analyze*, nato *Regression* in izberemo *Linear*. Odpre se pogovorno okno, v katerem spremenljivko *Število novo zaposlenih* (odvisna spremenljivka) prenesemo v desno okence, v polje *Dependent*, spremenljivko *število različnih subvencij zaposlovanja* (neodvisna spremenljivka) prenesemo v desno okence, v polje *Independent*. Kliknemo *Continue* in *OK*.

Odgovori in izpisi rezultatov:

Tabela 62: Rezultat – korelacijski in determinacijski koeficient

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,974 ^a	,949	,939	1,133

a. Predictors: (Constant), Subvencije

Iz tabele 62 je razvidno, da vrednost korelacijskega koeficienta (r_{xy}) znaša 0,974, kar kaže na to, da med spremenljivkama *število novo zaposlenih* in *število različnih subvencij zaposlovanja* obstaja močna linerna povezanost, smer povezanosti pa je pozitivna. Smer povezanosti med odvisno in neodvisno spremenljivko pri enostavni linearni regresiji razberemo iz predznaka pri izračunani oceni regresijskega koeficienta pri neodvisni spremenljivki, ki je podan v tabeli 64 in ima pozitiven predznak (ocenjena vrednost regresijskega koeficienta znaša 2,478 – tabela 64).

Vrednost determinacijskega koeficienta (r^2_{xy}) znaša 0,949 oz. 94,9 %. Determinacijski koeficient pojasnjuje, da je 94,9 % celotne variance števila novo zaposlenih pojasnjene z variabilnostjo neodvisne spremenljivke (število različnih subvencij zaposlovanja) (tabela 62).

Standardna napaka ocene odvisne spremenljivke znaša 1,133. Vrednost je različna od nič, kar pomeni, da na odvisno spremenljivko - število novo zaposlenih vplivajo še druge spremenljivke in slučajni vplivi, ne le neodvisna spremenljivka - število različnih subvencij zaposlovanja (tabela 62).

Tabela 63: Rezultat – F-test (ANOVA^a)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	119,295	1	119,295	92,922	,000 ^b
	Residual	6,419	5	1,284		
	Total	125,714	6			

a. Dependent Variable: Zaposleni

b. Predictors: (Constant), Subvencije

Kakovost regresijskega modela ugotavljamo s F-testom, pri čemer lahko glede na vrednost p ($p < 0,001$) trdimo, da je kakovost modela dobra. Obstaja linearna odvisnost med odvisno spremenljivko - število novo zaposlenih in neodvisno spremenljivko - število različnih subvencij zaposlovanja (tabela 63).

Z F-testom preizkušamo domnevo:

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

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

V tabeli 63 na osnovi rezultata (Sig. = 0,000, $p < 0,001$) ničelno domnevo: $H_0: r^2_{xy} = 0$ zavrnemo in sprejmemo raziskovalno domnevo: $H_1: r^2_{xy} > 0$.

Tabela 64: Rezultat – regresijski koeficienti in t-test

Model		Coefficients ^a				
		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-,632	1,046		-,604	,572
	Subvencije	2,478	,257	,974	9,640	,000

a. Dependent Variable: Zaposleni

Vrednost t-testa in stopnja značilnosti ($p < 0,001$) kažeta, da je regresijski koeficient β_1 različen od nič (zavržemo ničelno domnevo $H_0: \beta_1 = 0$ in potrdimo raziskovalno domnevo $H_1: \beta_1 \neq 0$), kar pomeni, da neodvisna spremenljivka *število različnih subvencij zaposlovanja* statistično značilno vpliva na odvisno spremenljivko *število novo zaposlenih* (tabela 64).

V tabeli 64 lahko razberemo ocenjeno regresijsko funkcijo:

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

\hat{y}_i – ocenjena vrednost spremenljivke y pri i-ti opazovani vrednosti spremenljivke x
 b_0 in b_1 – ocenjena regresijska koeficienta

$$\hat{y} = -0,632 + 2,478x_1 \text{ (kjer pomeni } x_1 \text{ število različnih subvencij zaposlovanja).}$$

Vrednost konstante (b_0) lahko vsebinsko pojasnimo takole: če je število subvencij za zaposlovanje enako 0, lahko v povprečju pričakujemo -0,632 novo zaposlenih.

Vrednost regresijskega koeficienta pri neodvisni spremenljivki (b_1) lahko vsebinsko pojasnimo takole: če zavod republike Slovenije za zaposlovanje poveča subvencije zaposlovanja za 1 enoto, se število novo zaposlenih v povprečju poveča za 2,478 novo zaposlenih.

Naloga 17

V podjetju, ki načrtuje uvajanje nove storitve na trg, želijo analizirati, koliko denarja bi bile za urejanje potrebne dokumentacije v zvezi s pridobitvijo lokacijske informacije za gradnjo, pripravljene odšteti njihove potencialne stranke. V tabeli 65 so odgovori za 10 anketiranih oseb na današnji dan.

Tabela 65: Podatki o anketiranih osebah

Spol	Znesek (v €)	Povprečna mesečna neto plača osebe (v €)
M	55	1050
Ž	90	1150
M	50	950
Ž	95	1200
M	45	1200
M	75	800
M	65	1000
Ž	80	1100
Ž	60	950
M	60	900

- a) Vnesite podatke v SPSS.
- b) Izračunajte povprečen znesek, ki so ga pripravljene odšteti anketirani. Izračunajte in pojasnite vse rezultate opisne statistike za to spremenljivko.
- c) Izračunajte povprečen znesek, ki so ga pripravljene odšteti moški in ženske posebej.
- d) Preračunajte znesek v USD (če veste, da je 1€ = 1,2142 USD)

Koraki: kliknemo *Transform* in nato *Compute variable*. V okence *Target Variable* napišemo ime nove spremenljivke *Znesek2*. V desno okence *Numeric Expression* prenesemo spremenljivko *Znesek* in napišemo: $Znesek * 1.2142$ in kliknemo *OK*. V okencu *Data View* se bo prikazala nova spremenljivka *Znesek2* s preračunano valuto € v USD.

Prikaz postopka:

The screenshot shows the 'Compute Variable' dialog box in IBM SPSS Statistics. The 'Target Variable' is 'Znesek2' and the 'Numeric Expression' is 'Znesek * 1.2142'. The 'Function group' is set to 'All'. The 'Functions and Special Variables' list is empty. The 'If...' field is empty. The 'OK' button is highlighted.

The resulting data table is as follows:

	Spol	Znesek	Plača	Znesek2	var	var	var	var	var
1	1	55	1050	66,78					
2	2	90	1150	109,28					
3	1	50	950	60,71					
4	2	95	1200	115,35					
5	1	45	1200	54,64					
6	1	75	800	91,07					
7	1	65	1000	78,92					
8	2	80	1100	97,14					
9	2	60	950	72,85					
10	1	60	900	72,85					
11									
12									
13									

Slika 11: Prikaz poteka postopka

- e) Izračunajte in pojasnite 95 %-ni interval zaupanja za povprečni znesek, ki so ga ljudje pripravljene odšteti.
- f) Ali je znesek, ki so ga pripravljene odšteti, odvisen od njihove povprečne mesečne neto plače?
- g) Ali lahko trdimo, da je povprečni znesek, ki so ga pripravljene odšteti enak 75€ ?

Odgovori in izpisi rezultatov:

b)

Tabela 66: Opisna statistika za spremenljivko Znesek (v €)

N	Valid	10
	Missing	0
Mean		67,50
Std. Error of Mean		5,336
Median		62,50
Mode		60
Std. Deviation		16,874
Variance		284,722
Skewness		,434
Std. Error of Skewness		,687
Kurtosis		-1,010
Std. Error of Kurtosis		1,334
Range		50
Minimum		45
Maximum		95

c)

Uporabimo ukaz *Split file*, ki nam omogoča analizo podatkov ločeno po spolu.

Tabela 67: Opisna statistika – povprečen znesek, ki so ga pripravljene odšteti moški

N	Valid	6
	Missing	0
Mean		58,33

Tabela 68: Opisna statistika – povprečen znesek, ki so ga pripravljene odšteti ženske

N	Valid	4
	Missing	0
Mean		81,25

e)

Tabela 69: 95 %-ni interval zaupanja za povprečni znesek, ki so ga ljudje pripravljene odšteti

		Statistic	Std. Error	
Znesek	Mean	67,50	5,336	
	95% Confidence Interval for Mean	Lower Bound	55,43	
		Upper Bound	79,57	
	5% Trimmed Mean		67,22	
	Median		62,50	
	Variance		284,722	
	Std. Deviation		16,874	
	Minimum		45	
	Maximum		95	
	Range		50	
	Interquartile Range		29	
	Skewness		,434	,687
Kurtosis		-1,010	1,334	

f)

Tabela 70: Rezultat - korelacijski in determinacijski koeficient

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,271 ^a	,073	-,043	17,229

a. Predictors: (Constant), Plača

Tabela 71: Rezultat – F-test (ANOVA)

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	187,888	1	187,888	,633	,449 ^b
	Residual	2374,612	8	296,826		
	Total	2562,500	9			

a. Dependent Variable: Znesek

b. Predictors: (Constant), Plača

Tabela 72: Rezultat – regresijski koeficienti in t-test

Model		Coefficients ^a			t	Sig.
		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta		
1	(Constant)	32,314	44,560		,725	,489
	Plača	,034	,043	,271	,796	,449

a. Dependent Variable: Znesek

g)

Tabela 73: Deskriptivna statistika za povprečni znesek

N	Valid	10
	Missing	0
Mean		67,50

Povprečni znesek, ki so ga pripravljene odšteti anketirane osebe ni enak 75€.

Naloga 18

Za osem podjetij so podatki o investicijah v novo tehnologijo (v 10^6 €) ter ustvarjenim dobičkom (v 10^3) v letu 2021 prikazani v tabeli 74. Analizirajte ali je ustvarjen dobiček odvisen od investicij v tehnologijo.

Tabela 74: Podatki o investicijah v novo tehnologijo (v 10^6 €) ter ustvarjenim dobičkom (v 10^3) v letu 2021

Podjetje	Investicije v tehnologijo	Ustvarjen dobiček
A	117	329
B	132	330
C	144	397
D	149	365
E	161	428
F	167	439
G	172	405
H	176	468

Izpis rezultata:**Tabela 75: Rezultat - korelacijski in determinacijski koeficient****Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,902 ^a	,813	,782	23,633

a. Predictors: (Constant), Investicije v tehnologijo

Tabela 76: Rezultat – F-test (ANOVA)**ANOVA^a**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14587,801	1	14587,801	26,119	,002 ^b
	Residual	3351,074	6	558,512		
	Total	17938,875	7			

a. Dependent Variable: Ustvarjen dobiček

b. Predictors: (Constant), Investicije v tehnologijo

Tabela 77: Rezultat – regresijski koeficienti in t-test**Coefficients^a**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	58,241	66,445		,877	,414
	Investicije v tehnologijo	2,213	,433	,902	5,111	,002

a. Dependent Variable: Ustvarjen dobiček

5.2 Multipla regresija

Multipla regresijsko uporabimo takrat, kadar na odvisno spremenljivko (y) vpliva več neodvisnih spremenljivk x_i ($i = 1, 2, \dots, k$).

V tem primeru sta analitična kazalca multipli korelacijski koeficient, R , ki kaže jakost povezanosti med odvisno in k neodvisnimi spremenljivkami in je prikazan v absolutni vrednosti. Prilagojeni determinacijski koeficient R^2 (*Adjusted R Square*) pa predstavlja delež variance odvisne spremenljivke, ki je pojasnjena z variabilnostjo v model vključenih neodvisnih spremenljivk (Tabachnick in Fidell, 2013).

Naloga 19

Podjetje X prodaja računalniško opremo. Za povečanje prodaje računalniške opreme so v podjetju na novo uvedli telefonsko trženje in možnost nakupa računalniške opreme preko interneta. Analizirajte ali je vrednost prodaje računalniške opreme (v d.e.) odvisna od telefonskega trženja (x_1) in možnosti internetne prodaje (x_2). V tabeli 78 so podatki o prodaji računalniške opreme.

Tabela 78: Podatki o prodaji računalniške opreme

Vrednost prodaje izdelka X (v d.e.)	Prodani izdelki preko telefonskega trženja	Prodani izdelki preko interneta
2700	6	3
3900	7	5
2800	6	3
4199	9	4
8500	12	6
3153	8	3
5651	11	7
5702	6	8
4261	13	4
3150	7	3
4036	8	6
2215	4	4
3436	9	4
6446	7	8
5204	8	7

- 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 in pojasnite enačbo regresijske funkcije.

Postopek za izvedbo multiple regresije: Kliknemo na *Analyze*, nato *Regression* in izberemo *Linear*. Odpre se pogovorno okno, v katerem spremenljivko *vrednost prodaje izdelka* (odvisna spremenljivka) prenesemo v desno okence, v polje *Dependent*, obe neodvisni spremenljivki *prodani izdelki preko telefonskega trženja* in *prodani izdelki preko interneta* prenesemo v desno okence, v polje *Independent*. Kliknemo *Continue* in *OK*.

Odgovori in izpisi rezultatov:

a)

Tabela 79: Rezultat – multipli korelacijski in prilagojeni determinacijski koeficient

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,877 ^a	,770	,732	871,084

a. Predictors: (Constant), Prodani izdelki preko interneta, Prodani izdelki preko telefonskega trženja

Vrednost multiplega korelacijskega koeficienta (R) znaša 0,877, kar kaže na to, da med odvisno spremenljivko *vrednost prodaje izdelka* in neodvisnima spremenljivkama *prodani izdelki preko telefonskega trženja* in *prodani izdelki preko interneta* obstaja močna povezanost (tabela 79).

Vrednost prilagojenega multiplega determinacijskega koeficienta (R^2) znaša 0,732 oz. 73,2 %. Prilagojen determinacijski koeficient pojasnjuje, da je 73,2 % celotne variance vrednosti *prodaje izdelka* (odvisne spremenljivke) pojasnjene z regresijskim modelom (z neodvisnima spremenljivkama *prodani izdelki preko telefonskega trženja* in *prodani izdelki preko interneta*) (tabela 79).

b)

Tabela 80: Rezultat – F-test (ANOVA)

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	30482625,748	2	15241312,874	20,086	,000 ^b
	Residual	9105451,985	12	758787,665		
	Total	39588077,733	14			

a. Dependent Variable: Vrednost prodaje računalniške opreme v d.e.

b. Predictors: (Constant), Prodani izdelki preko interneta, Prodani izdelki preko telefonskega trženja

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 (tabela 80). Navedeno pomeni, da obstaja odvisnost med odvisno spremenljivko (*vrednost prodaje*) in vsaj eno neodvisno spremenljivko.

Z F-testom preizkušamo domnevo:

H_0 : Prilagojeni determinacijski koeficient je enak 0 ($R^2 = 0$)

H1: Prilagojeni determinacijski koeficient je večji od 0 ($R^2 > 0$).

Na osnovi rezultatov ($p < 0,001$) lahko zavrnamo ničelno domnevo, da je $R^2 = 0$.

c)

Tabela 81: Rezultat – regresijski koeficienti

Model		Coefficients ^a			t	Sig.
		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta		
1	(Constant)	-1318,792	965,565		-1,366	,197
	Prodani izdelki preko telefonskega trženja	309,812	96,431	,448	3,213	,007
	Prodani izdelki preko interneta	635,302	126,753	,700	5,012	,000

a. Dependent Variable: Vrednost prodaje računalniške opreme v d.e.

Statistično značilen vpliv pojasnjevalne spremenljivke x_i ($i=1,2,\dots, k$) na odvisno spremenljivko preverjamo s t-testom, pri čemer preizkušamo domnevo:

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

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

Podobno lahko preverimo tudi statistično značilnost regresijske konstante.

T-test in stopnja značilnosti za oba regresijska koeficienta ($p < 0,05$) kažeta, da obe neodvisni spremenljivki oz. pojasnjevalni spremenljivki (*prodani izdelki preko telefonskega trženja* in *prodani izdelki preko interneta*) statistično značilno vplivata na odvisno spremenljivko (*vrednost prodaje izdelka*). V obeh primerih namreč lahko ničelno domnevo zavrnamo z manj kot 5 %-nim tveganjem.

d)

Dobljena enačba regresijske funkcije, z ocenjenimi vrednostmi regresijskih koeficientov na osnovi uporabljenih vzorčnih podatkov, se glasi (tabela 81):

$$\hat{y} = -1318,792 + 309,812x_1 + 635,302x_2$$

(kjer pomeni x_1 število prodanih izdelkov preko telefonskega trženja ter x_2 število prodanih izdelkov preko interneta).

Ocenjeni vrednosti regresijskih koeficientov povesta, za koliko enot se v povprečju spremeni vrednost odvisne spremenljivke, če se vrednost posamezne pojasnjevalne spremenljivke spremeni za 1 enoto, vrednost druge pojasnjevalne spremenljivke pa ostane pri tem nespremenjena (v nasprotnem primeru govorimo o multikolinearnosti). Tako v našem primeru ocenjeni regresijski koeficient pri x_1 pove, da se vrednost prodaje izdelka v povprečju poveča za 309,812 d.e., če se število prodanih izdelkov preko telefonskega trženja poveča za 1 enoto in pri tem vrednost spremenljivke število prodanih izdelkov preko interneta ostane nespremenjena.

Naloga 20

Raziskovalci so v vzorec zajeli 10 različnih hotelov. Želeli so analizirati ali je število gostov v hotelu odvisno od števila različnih organiziranih dogodkov in števila wellnes ponub, ki jih nudijo hoteli. Podatki so v tabeli 82.

Podatke vnesite v SPSS in pojasnite vse rezultate multiple regresije.

Tabela 82: Podatki o številu gostov, številu organiziranih dogodkov in številu wellnes ponub

Hotel	Število gostov	Število organiziranih dogodkov	Število wellnes ponub
A	60	4	1
B	100	7	2
C	12	2	3
D	17	5	3
E	43	8	5
F	100	4	2
G	66	10	4
H	90	15	6
I	7	1	3
J	26	4	2

Izpis rezultata:**Tabela 83: Rezultat – multipli korelacijski in prilagojeni determinacijski koeficient****Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,802 ^a	,644	,542	24,623

a. Predictors: (Constant), Število wellnes ponub, število dogodkov

Tabela 84: Rezultat – F-test (ANOVA)**ANOVA^a**

Model	Sum of Squares	df	Mean Square	F	Sig.	
1	Regression	7674,986	2	3837,493	6,330	,027 ^b
	Residual	4243,914	7	606,273		
	Total	11918,900	9			

a. Dependent Variable: Število gostov

b. Predictors: (Constant), Število wellnes ponub, število dogodkov

Tabela 85: Rezultat – regresijski koeficienti**Coefficients^a**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	50,929	18,435		2,763	,028
	število dogodkov	10,335	2,910	1,182	3,552	,009
	Število wellnes ponub	-19,626	7,950	-,822	-2,469	,043

a. Dependent Variable: Število gostov

Naloga 21

Natakar je analiziral povezanost med višino napitnine (odvisna spremenljivka) in zneskom računa ter številom večerij pri omizju (dve neodvisni spremenljivki). V ta namen je zbral podatke za 8 računov, ki so prikazani v tabeli 86.

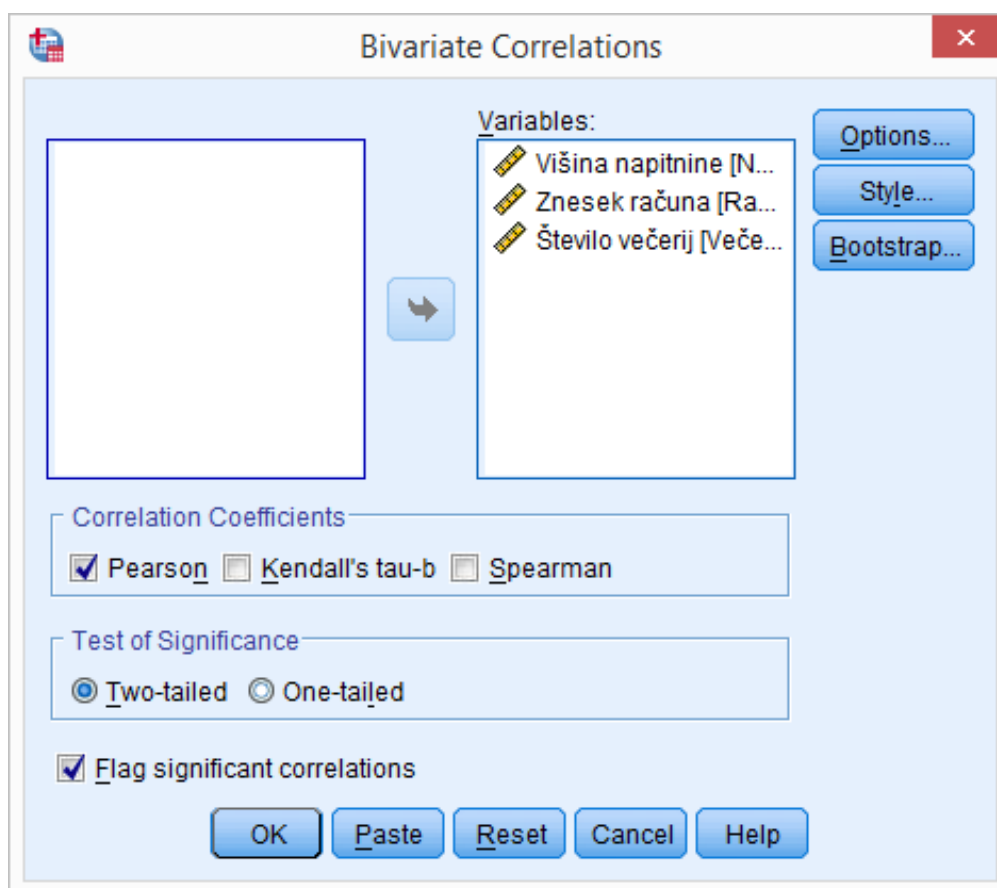
Tabela 86: Podatki o višini napitnine, zneseku računa in številu večerij

Zaporedna številka	Višina napitnine	Znesek računa	Število večerij
1	7,5	40	2
2	0,5	15	1
3	2	30	3
4	3,5	25	4
5	9,5	50	4
6	2,5	20	5
7	3,5	35	5
8	1,0	10	2

- Vnesite podatke v SPSS.
- Analizirajte rezultate multiple korelacije in regresije.

Postopek za izvedbo multiple korelacije: Kliknemo *Analyze*, nato *Correlate* in nato *Bivariate*. V desno okence *Variables* prenesemo vse tri spremenljivke: *višina napitnine*, *znesek računa*, *število večerij*. Pri *Correlation Coefficients* obkljukamo *Pearson* in *OK*.

Prikaz postopka:



Slika 12: Prikaz poteka postopka

Odgovori in zpski rezultatov:

b)

Rezultati multiple korelacije:

Tabela 87: Rezultat – multipla korelacija

		Correlations		
		Višina napitnine	Znesek računa	Število večerij
Višina napitnine	Pearson Correlation	1	,910**	,243
	Sig. (2-tailed)		,002	,562
	N	8	8	8
Znesek računa	Pearson Correlation	,910**	1	,351
	Sig. (2-tailed)	,002		,394
	N	8	8	8
Število večerij	Pearson Correlation	,243	,351	1
	Sig. (2-tailed)	,562	,394	
	N	8	8	8

** . Correlation is significant at the 0.01 level (2-tailed).

Pearsonov koeficient meri linearno povezanost (korelacijo) med spremenljivkami.

Pearsonov koeficient med spremenljivko *Višina napitnine* in *Znesek računa* znaša 0,910, kar kaže na močno povezanost med obema spremenljivkama. Stopnja značilnosti $p < 0,05$ (Sig. = 0,002) kaže na to, da obstaja povezanost med spremenljivkama *Višina napitnine* in *Znesek računa*.

Pearsonov koeficient med spremenljivko *višina napitnine* in *število večerij* znaša 0,243, kar kaže na slabo povezanost med obema spremenljivkama. Stopnja značilnosti $p > 0,05$ (Sig. = 0,562) kaže na to, da ne obstaja povezanost med spremenljivko *višina napitnine* in *število večerij*.

Rezultati multiple regresije:

Tabela 88: Rezultat – multipli korelacijski in prilagojeni determinacijski koeficient

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,913 ^a	,834	,768	1,524

a. Predictors: (Constant), Število večerij, Znesek računa

Tabela 89: Rezultat – F-test (ANOVA)

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	58,389	2	29,194	12,571	,011 ^b
	Residual	11,611	5	2,322		
	Total	70,000	7			

a. Dependent Variable: Višina napitnine

b. Predictors: (Constant), Število večerij, Znesek računa

Tabela 90: Rezultat – regresijski koeficienti

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-1,915	1,598		-1,199	,284
	Znesek računa	,223	,046	,940	4,834	,005
	Število večerij	-,184	,413	-,087	-,446	,674

a. Dependent Variable: Višina napitnine

Naloga 22

Analiziramo odvisnost začetniške plače diplomanta od povprečne ocene študija ter od števila dodatnih kompetenc, ki si jih je pridobil v času študija. V tabeli 91 so podatki za 10 diplomantov.

Tabela 91: Podatki za 10 diplomantov

Zaporedna številka	Začetniška plača (d.e.)	Povprečna ocena študija	Število dodatnih kompetenc
1	750	8,3	2
2	550	8,2	2
3	1050	9,0	5
4	1100	9,0	7
5	650	7,5	1
6	870	7,5	3
7	990	8,9	4
8	1100	8,9	6
9	1200	9,5	8
10	800	8,0	2

a) Vnesite podatke v SPSS.

b) Analizirajte rezultate multiple korelacije in regresije.

Rezultat multiple korelacije:

Tabela 92: Rezultat – multipla korelacija

Correlations				
		Začetniška plača	Povprečna ocena študija	Število dodatnih kompetenc
Začetniška plača	Pearson Correlation	1	,798**	,928**
	Sig. (2-tailed)		,006	,000
	N	10	10	10
Povprečna ocena študija	Pearson Correlation	,798**	1	,867**
	Sig. (2-tailed)	,006		,001
	N	10	10	10
Število dodatnih kompetenc	Pearson Correlation	,928**	,867**	1
	Sig. (2-tailed)	,000	,001	
	N	10	10	10

** . Correlation is significant at the 0.01 level (2-tailed).

Rezultati multiple regresije:

Tabela 93: Rezultat – multipli korelacijski in prilagojeni determinacijski koeficient

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,928 ^a	,861	,822	91,036

a. Predictors: (Constant), Število dodatnih kompetenc, Povprečna ocena študija

Tabela 94: Rezultat – F-test (ANOVA)

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	360627,290	2	180313,645	21,757	,001 ^b
	Residual	58012,710	7	8287,530		
	Total	418640,000	9			

a. Dependent Variable: Začetniška plača

b. Predictors: (Constant), Število dodatnih kompetenc, Povprečna ocena študija

Tabela 95: Rezultat – regresijski koeficienti

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	634,925	669,948		,948	,375
	Povprečna ocena študija	-8,272	89,044	-,026	-,093	,929
	Število dodatnih kompetenc	85,305	25,294	,951	3,373	,012

a. Dependent Variable: Začetniška plača

Naloga 23

Pri analizi pričakovanega števila gostov v smučarskem središču, je uprava hotela za mesec dni v smučarski sezoni, vsak drugi dan, zbrala podatke, ki so prikazani v tabeli 96.

Tabela 96: Zbrani podatki

Dan	Število smučarjev	Višina snega (cm)	Najvišja dnevna temperatura (stopinje celzija)	Vikend 0 - ne 1 - da
1	402	55	11	0
2	337	45	12	0
3	471	40	10	0
4	610	79	9	1
5	620	75	8	1
6	545	60	9	0
7	523	60	9	0
8	563	75	9	0
9	473	120	2	0
10	358	80	10	0
11	768	150	1	1
12	753	150	1	1
13	485	140	8	0
14	667	170	1	0
15	635	170	2	0

- a) Vnesite podatke v SPSS.
- b) Izračunajte in pojasnite vse rezultate opisne statistike za število smučarjev.
- c) Ali se povprečno število smučarjev statistično značilno razlikuje med vikendom in med tednom?
- d) Ali se število smučarjev statistično značilno razlikuje, če je višina snega do 1m oziroma če je več kot 1m?
- e) Analizirajte in pojasnite vse kazalce multiple korelacije in regresije.

Postopek pri točki d):

V tabeli 96 imamo podano višino snega v cm, ker pa naloga sprašuje ali se število smučarjev statistično značilno razlikuje, če je višina snega do 1m oziroma če je več kot 1m, smo morali višino snega izraženo v cm pretvoriti v m (pri opredelitvi nove spremenljivke, se namesto 1 m zapiše 100 cm).

Postopek:

Višino snega je potrebno preoblikovati tako, da bosta nastala dva neodvisna vzorca in sicer, višina snega do 1m in več kot 1m. Postopek: Kliknemo *Transform* in nato *Recode into different variable*. Spremenljivko *Sneg* prenesemo v desno okence, kjer piše »*Numeric Variable*→*Output Variable*«. V okencu *Output Variable* napišemo ime nove spremenljivke, ki jo bomo oblikovali tako, bo imela vrednost 0, če je višina snega do 100 cm in vrednost 1, če je višina snega več kot 100 cm. Poimenovali jo bomo na primer *Sneg2* ter kliknemo *Change*. Nato kliknemo okvirček *Old and new values*, v okence kjer piše »*Range, lowest trough value*« napišemo 100 in v okence kjer piše »*New value*« pod *Value* napišemo 0 ter kliknemo *add*. Nato kliknemo na »*Range, value through highest*«, ter napišemo 100 in pod »*New value*« v okence *Value* napišemo 1 ter kliknemo *add*. Obe spremembi se izpišeta v spodnjem okencu *Old* → *New*. Kliknemo *Continue*. V okencu *Data View* se prikaže nova spremenljivka, ki ima vrednost 0, če je višina snega do 1m in vrednost 1, če je višina snega več kot 1m. Nato lahko izvedemo t-test za dva neodvisna vzorca.

Izpis rezultata:

b)

Tabela 97: Opisna statistika za spremenljivko število smučarjev

N	Valid	15
	Missing	0
Mean		547,33
Std. Error of Mean		33,771
Median		545,00
Mode		337 ^a
Std. Deviation		130,793
Variance		17106,810
Skewness		,081
Std. Error of Skewness		,580
Kurtosis		-,677
Std. Error of Kurtosis		1,121
Range		431
Minimum		337
Maximum		768

a. Multiple modes exist. The smallest value is shown

c)

Tabela 98: Rezultati t-testa za dva neodvisna vzorca**Independent Samples Test**

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-t.)	Mean Dif.	Std. Error Dif.	95% Confidence Interval of the Diff.	
									Lower	Upper
Število smučarjev	Equal variances assumed	,091	,767	-3,255	13	,006	-191,477	58,823	-318,558	-64,397
	Equal variances not assumed			-3,629	6,710	,009	-191,477	52,757	-317,330	-65,624

d)

Tabela 99: Rezultati t-testa za dva neodvisna vzorca

		Independent Samples Test								
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Število smučarjev	Equal variances assumed	,168	,688	-2,285	13	,040	-138,056	60,426	-268,598	-7,513
	Equal variances not assumed			-2,196	9,411	,054	-138,056	62,873	-279,343	3,232

e)

Rezultati multiple korelacije:

Tabela 100: Rezultati multiple korelacije

		Correlations			
		Število smučarjev	Višina snega	Najvišja dnevna temperatura	Vikend
Število smučarjev	Pearson Correlation	1	,652**	-,755**	,670**
	Sig. (2-tailed)		,008	,001	,006
	N	15	15	15	15
Višina snega	Pearson Correlation	,652**	1	-,898**	,208
	Sig. (2-tailed)	,008		,000	,458
	N	15	15	15	15
Najvišja dnevna temperatura	Pearson Correlation	-,755**	-,898**	1	-,313
	Sig. (2-tailed)	,001	,000		,256
	N	15	15	15	15
Vikend	Pearson Correlation	,670**	,208	-,313	1
	Sig. (2-tailed)	,006	,458	,256	
	N	15	15	15	15

**. Correlation is significant at the 0.01 level (2-tailed).

Rezultati multiple regresije:

Tabela 101: Rezultat – multipli korelacijski in prilagojeni determinacijski koeficient

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,883 ^a	,780	,719	69,278

a. Predictors: (Constant), Vikend, Višina snega, Najvišja dnevna temperatura

Tabela 102: Rezultat – F-test (ANOVA)

ANOVA^a

Model	Sum of Squares	df	Mean Square	F	Sig.	
1	Regression	186701,744	3	62233,915	12,967	,001 ^b
	Residual	52793,589	11	4799,417		
	Total	239495,333	14			

a. Dependent Variable: Število smučarjev

b. Predictors: (Constant), Vikend, Višina snega, Najvišja dnevna temperatura

Tabela 103: Rezultat – regresijski koeficienti

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	619,803	162,498		3,814	,003
	Višina snega	,130	,913	,047	,142	,889
	Najvišja dnevna temperatura	-17,961	10,761	-,562	-1,669	,123
	Vikend	138,507	43,256	,485	3,202	,008

a. Dependent Variable: Število smučarjev

6 Študija primera - mobilno oglaševanje

6.1. Vprašalnik za mobilno oglaševanje

Podjetja se v velikem obsegu poslužujejo mobilnega oglaševanja. Za oblikovanje prave strategije mobilnega oglaševanja, pa mora podjetje vedeti, kako mobilno oglaševanje dojemajo njihovi potencialni kupci.

S tem namenom je bila zasnovana anketna raziskava, ki je temeljila na vprašalniku z merskimi lestvicami, s katerimi podjetje meri:

- Kako mladi zaznavajo uporabnost mobilnih oglasov,
- Kakšen odnos imajo mladi do mobilnega oglaševanja in kako ga sprejemajo,
- Ali se zdijo mladim mobilni oglasi igrivi, dinamični, zanimivi...
- Ali se zdijo mladim mobilni oglasi nadležni,
- Itd.

VPRAŠALNIK

V.1 Navajamo Vam trditve o zaznani uporabnosti mobilnih oglasov

V kolikšni meri se strinjate z naslednjimi trditvami. Trditve o zaznavni uporabnosti mobilnih oglasov ocenite od 1 do 5, pri čemer ocene pomenijo (izbrano oceno za vsako trditev v spodnji tabeli ustrezno obkrožite):

Sploh se ne strinjam	Se ne strinjam	Delno se strinjam	Se strinjam	Popolnoma se strinjam
1	2	3	4	5

Trditve zaznavno uporabnost	Ocene				
Uporaba mobilnega oglasa mi pomaga pri odločitvi nakupa dodatnih storitev/ izdelka	1	2	3	4	5
Skozi mobilno oglaševanje dobim točne informacije	1	2	3	4	5
Mobilno oglaševanje varuje naravno okolje	1	2	3	4	5
Uporaba mobilnih oglasov se mi zdi koristna	1	2	3	4	5
Mobilni oglasi me obveščajo o novostih	1	2	3	4	5
Preko mobilnih oglasov lažje dobim koristne informacije	1	2	3	4	5
Preko mobilnih oglasov dobim boljši opis oglaševanih izdelkov	1	2	3	4	5
Mobilno oglaševanje je prilagojeno mojim potrebam	1	2	3	4	5

V.2 Navajamo Vam trditve o odnosu in sprejemanju mobilnega oglaševanja na splošno

V kolikšni meri se strinjate z naslednjimi trditvami. Trditve o odnosu do sprejema mobilnih oglasov ocenite od 1 do 5, pri čemer ocene pomenijo (izbrano oceno za vsako trditev v spodnji tabeli ustrezno obkrožite):

Sploh se ne strinjam	Se ne strinjam	Delno se strinjam	Se strinjam	Popolnoma se strinjam
1	2	3	4	5

Trditve odnos do sprejema	Ocene				
Zdi se mi, da je ideja o uporabljanju mobilnih oglasov dobra	1	2	3	4	5
Imam pozitiven odnos do sprejema mobilnih oglasov	1	2	3	4	5
Za mobilne oglase mi je vseeno	1	2	3	4	5
Mobilno oglaševanje je zanimivo	1	2	3	4	5

V.3 Navajamo Vam trditve o sprejemanju mobilnih oglasov

V kolikšni meri se strinjate z naslednjimi trditvami. Trditve o sprejemu mobilnih oglasov ocenite od 1 do 5, pri čemer ocene pomenijo (izbrano oceno za vsako trditev v spodnji tabeli ustrezno obkrožite):

Sploh se ne strinjam	Se ne strinjam	Delno se strinjam	Se strinjam	Popolnoma se strinjam
1	2	3	4	5

Trditve o sprejemu mobilnega oglaševanja	Ocene				
Ko dobim mobilni oglas ga preberem	1	2	3	4	5
Ko dobim mobilni oglas ga ne zavrnem takoj	1	2	3	4	5
Ko dobim mobilni oglas ga v celoti preberem ter posredujem drugim	1	2	3	4	5
Pripravljen sem se odzvati na mobilni oglas	1	2	3	4	5

V.4 Navajamo Vam trditve o zaznani igrivosti mobilnih oglasov

V kolikšni meri se strinjate z naslednjimi trditvami. Trditve o zaznani igrivosti mobilnih oglasov ocenite od 1 do 5, pri čemer ocene pomenijo (izbrano oceno za vsako trditev v spodnji tabeli ustrezno obkrožite):

Sploh se ne strinjam	Se ne strinjam	Delno se strinjam	Se strinjam	Popolnoma se strinjam
1	2	3	4	5

Trditve o igrivosti mobilnih oglasov	Ocene				
Pomembno je, da mobilno oglaševanje nudi zabavno izkušnjo	1	2	3	4	5
Če mobilni oglasi, poleg klikanja in dotikanja omogočajo tudi nove načine interakcije, kot so: uporaba kamere in mikrofona, ipd., me zabavajo	1	2	3	4	5
Če mobilni oglasi vsebujejo zvok in glasbo, me zabavajo	1	2	3	4	5
Če mobilni oglasi vsebujejo animacije in/ali video, me zabavajo	1	2	3	4	5
Zabavni mobilni oglasi me sproščajo	1	2	3	4	5
Mobilni oglasi so zabavni predvsem v prostem času	1	2	3	4	5

V.5 Navajamo Vam trditve o zaznani nadležnosti mobilnih oglasov

V kolikšni meri se strinjate z naslednjimi trditvami. Trditve o zaznani nadležnosti mobilnih oglasov ocenite od 1 do 5, pri čemer ocene pomenijo (izbrano oceno za vsako trditev v spodnji tabeli ustrezno obkrožite):

Sploh se ne strinjam	Se ne strinjam	Delno se strinjam	Se strinjam	Popolnoma se strinjam
1	2	3	4	5

Trditve o zaznani nadležnosti mobilnih oglasov.	Ocene				
Prevelika količina oglasnih sporočil preko mobilnega telefona je zame moteče	1	2	3	4	5
Nenehno pojavljanje mobilnih oglasov, kljub zapiranju le-teh, je zame moteče	1	2	3	4	5
Pojav istih oglasnih sporočil me moti	1	2	3	4	5
Neustrezen čas prejemanja oglasnih sporočil me moti	1	2	3	4	5
Mobilni oglasi so nadležni ker so predolgi in nerazumljivi	1	2	3	4	5
Nadležni se mi zdijo, ker zavajajo kupce	1	2	3	4	5

V.6 Navajamo Vam trditve o zaznani preglednosti mobilnih oglasov

V kolikšni meri se strinjate z naslednjimi trditvami. Trditve o zaznani preglednosti mobilnih oglasov ocenite od 1 do 5, pri čemer ocene pomenijo (izbrano oceno za vsako trditev v spodnji tabeli ustrezno obkrožite):

Sploh se ne strinjam	Se ne strinjam	Delno se strinjam	Se strinjam	Popolnoma se strinjam
1	2	3	4	5

Trditve o zaznani preglednosti mobilnih oglasov	Ocene				
Velika količina vsebin v mobilnem oglasu me odvrne od oglasa	1	2	3	4	5
Iz mobilnih oglasov ne razberemo bistva	1	2	3	4	5
Prevelika izbira različnih barv v oglasu me moti	1	2	3	4	5
Iz mobilnih oglasov lahko hitro razberemo informacije	1	2	3	4	5
Preglednim mobilnim oglasom namenimo več pozornosti	1	2	3	4	5

V.7 Navajamo Vam trditve o informacijski pismenosti pri uporabi mobilnih naprav

V kolikšni meri se strinjate z naslednjimi trditvami. Trditve o informacijski pismenosti mobilnih naprav ocenite od 1 do 5, pri čemer ocene pomenijo (izbrano oceno za vsako trditev v spodnji tabeli ustrezno obkrožite):

Sploh se ne strinjam	Se ne strinjam	Delno se strinjam	Se strinjam	Popolnoma se strinjam
1	2	3	4	5

Trditve glede informacijske pismenosti pri uporabi mobilne telefonije	Ocene				
Moj nivo znanja o uporabi mobilne telefonije zadostuje mojim trenutnim potrebam	1	2	3	4	5
Brez težav namestim ali posodobim mobilni operacijski sistem (Android,iOS)	1	2	3	4	5
Nalaganje/posodabljanje aplikacij mi ne predstavlja težav	1	2	3	4	5

V.8 Navajamo Vam dejavnike uporabe mobilnih naprav

V kolikšni meri so za Vas pomembni naslednji dejavniki uporabe mobilnih naprav (izbrano oceno za vsak dejavnik v spodnji tabeli ustrezno obkrožite):

Zelo nepomemben	Nepomemben	Delno pomemben	Pomemben	Zelo pomemben
1	2	3	4	5

Dejavniki	Ocene				
	1	2	3	4	5
Blagovna znamka mobilnega telefona	1	2	3	4	5
Sistem mobilnega telefona (npr. android, Windows, iOS)	1	2	3	4	5
Velikost mobilnega telefona	1	2	3	4	5
Zmogljivost mobilnega telefona	1	2	3	4	5
Barva mobilnega telefona	1	2	3	4	5
Ločljivost kamere	1	2	3	4	5
Cena mobilnega telefona	1	2	3	4	5
Cena mobilnih storitev	1	2	3	4	5
Dizajn (oblika) mobilnega telefona	1	2	3	4	5
Zadnja različica (izdaja) mobilnega telefona	1	2	3	4	5
Ponudnik mobilnih storitev	1	2	3	4	5

Demografija

D1. Spol (z oznako X izberite odgovor): Moški Ženski

D2. Moje stalno prebivališče je (z oznako X izberite odgovor): na podeželju v mestu

D3. Napišite vašo starost: _____

6.2 Naloge

Naloga 1

a) Preglejte **priloženi vprašalnik** in poiščite primer za:

- Večdimenzionalno spremenljivko
- Dihotomno spremenljivko
- Številsko spremenljivko
- Opisno spremenljivko

- b) Spletno anketiranje – poglejte **priloženi link** in izpolnite spletno anketo.
- c) Rezultati anketiranja – poglejte **priloženo bazo podatkov** in opišite, kaj pomeni, da so vrednosti spremenljivke kodirane – na primer spremenljivka spol.
- d) Za spremenljivko “V1_4: Uporaba mobilnih oglasov se mi zdi koristna”, izračunajte in pojasnite:
 - a. Povprečno stopnjo strinjanja s trditvijo
 - b. Mediano
 - c. Modus
 - d. Variacijski razmik
- e) Za spremenljivko spol prikažite strukturo zbranih podatkov v strukturnem krogu.
- f) V tabeli prikažite strukturo zbranih podatkov po spolu in po kraju bivanja (podeželje ali mestno okolje).

Naloga 2

V priloženi bazi podatkov Mobilno oglaševanje:

- a) Izberite samo podatke za ženske (ukaz select cases) in opravite analizo, opisano pri nalogi 1d zgoraj.
- b) Za spremenljivko “V3_1: Ko dobim mobilni oglas ga preberem”, v rezultatu ločeno prikažite analizo za moške in ženske (ukaz split file).
- c) Spremenljivko starost prekodirajte v novo spremenljivko, ki bo imela vrednost 1, če je oseba stara do 20 let, ter vrednost 2, če je stara več kot 20 let (ukoz recode into different variable).
- d) Za spremenljivko “V3_1: Ko dobim mobilni oglas ga preberem”, v rezultatu ločeno prikažite analizo za obe starostni skupini (do 20 in več kot 20 let) (ukaz split file).

Naloga 3

V priloženi bazi podatkov Mobilno oglaševanje:

- a) Izmerite in komentirajte vrednost kazalca Cronbach alfa za mersko lestvico za spremenljivko "V_5: nadležnost mobilnih oglasov" iz vprašalnika.
- b) Analizirajte, kako bi lahko izboljšali zanesljivost merske lestvice (Item-total statistics).
- c) Katera merska lestvica, V_5 ali V_4 je bolj zanesljiva?

Naloga 4

V priloženi bazi podatkov Mobilno oglaševanje:

- a) Izmerite in komentirajte vrednost kazalca Cronbach alfa za mersko lestvico za spremenljivko "V_3: Sprejem mobilnih oglasov" iz vprašalnika.
- b) Na osnovi rezultatov analizirajte, ali bi lahko izboljšali zanesljivost merske lestvice.

Naloga 5

V priloženi bazi podatkov Mobilno oglaševanje:

- a) Preverite domnevo, da se moški in ženske ne razlikujejo glede povprečnega strinjanja s trditvijo "V7_1: Moj nivo znanja o uporabi mobilne telefonije zadostuje mojim trenutnim potrebam".
 - a. Opišite in pojasnite rezultate testiranja Levenovega testa enakosti varianc
 - b. Opišite in pojasnite rezultate t-testa za dve neodvisni skupini
- b) Ali se osebe, stare do 21 let, razlikujejo od oseb, ki so starejše, glede povprečnega strinjanja s trditvijo "V4_5: Zabavni mobilni oglasi me sproščajo".
- c) Ali se povprečne stopnja strinjanja s trditvijo V7_1 in povprečna stopnja strinjanja s trditvijo V4_5, statistično značilno razlikujeta? Kateri test ste uporabili? Zakaj?

Naloga 6

V priloženi bazi podatkov Mobilno oglaševanje:

- a) Preverite domnevo, da se moški in ženske ne razlikujejo glede povprečnega strinjanja s trditvijo “V3_4: Pripravljen sem se odzvati na mobilni oglas”.
- b) Preverite domnevo, da se povprečna stopnja strinjanja s trditvijo “V1_1: Uporaba mobilnega oglasa mi pomaga pri odločitvi nakupa dodatnih storitev/izdelka” in povprečna stopnja strinjanja s trditvijo “V1_8: Mobilno oglaševanje je prilagojeno mojim potrebam”, statistično značilno ne razlikujeta.

Naloga 7

Analizirajte ali je vrednost spremenljivke “V1_1: Uporaba mobilnega oglasa mi pomaga pri odločitvi nakupa dodatnih storitev/izdelka” odvisna od vrednosti spremenljivke “V1_5: Mobilni oglasi me obveščajo o novostih” ter pojasnite vse rezultate enostavne linerane regresije.

Naloga 1

Analizirajte in pojasnite, ali je vrednost spremenljivke “V1_4: Uporaba mobilnih oglasov se mi zdi koristna”, odvisna od vrednosti spremenljivke “V1_2: Skozi mobilno oglaševanje dobim točne informacije” ter od spola osebe:

- a) Pojasnite multipli korelacijski koeficient in prilagojeni determinacijski koeficient.
- b) Preverite kakovost modela kot celote.
- c) Opišite postopek preverjanja za vsako posamezno vključeno spremenljivko, ali statistično značilno vpliva na odvisno spremenljivko ali ne. Utemeljite odgovor.
- d) Kaj pomenijo ocenjene vrednosti regresijskih koeficientov v regresijski funkciji? Izpišite regresijsko funkcijo z enačbo.

Naloga 9

Analizirajte in pojasnite, ali je vrednost spremenljivke “V6_5: Preglednim mobilnim oglasom namenimo več pozornosti”, odvisna od vrednosti spremenljivke “V3_4: Pripravljen sem se odzvati na mobilni oglas” ter od starosti osebe:

- a) Pojasnite multipli korelacijski koeficient in prilagojeni determinacijski koeficient.
- b) Preverite kakovost modela kot celote.
- c) Opišite postopek preverjanja za vsako posamezno vključeno spremenljivko, ali statistično značilno vpliva na odvisno spremenljivko ali ne. Utemeljite odgovor.
- d) Kaj pomenijo ocenjene vrednosti regresijskih koeficientov v regresijski funkciji? Izpišite regresijsko funkcijo z enačbo.

7 Naloge za seminarske vaje

Naloga 1

Predstavljajte si, da se vaše podjetje ukvarja z oglaševanjem in sicer z oglaševanjem na oglasnih mestih ob ogledalih v frizerskih salonih.

Ker želite oceniti, kako primeren je tak način oglaševanja za oglaševanje hitro pripravljenih juh Podravka, želite anketirati 100 strank v frizerskem salonu. V podjetju Podravka seveda vedo, katere so značilnosti njihovih najpogostejših kupcev in želijo vedeti, ali z oglaševanjem v frizerskem salonu dosežejo svoje potencialne stranke.

Pripravite vprašalnik z največ 5 vprašanji, s katerim boste izvedeli odgovor na zgoraj zastavljeno vprašanje - Ali je tak način oglaševanja za oglaševanje hitro pripravljenih juh Podravka primeren?

Pojasnite, zakaj in s kakšnim namenom ste oblikovali posamezno vprašanje.

Naloga 2

Opišite tehnike čistega verjetnostnega vzorčenja, ki temeljijo na enostavnem slučajnem vzorčenju.

Naloga 3

Opišite tehnike neverjetnostnega vzorčenja.

Naloga 4

Opišite možne načine za izbiro enot v vzorec pri enostavnem slučajnem vzorčenju.

Naloga 5

Na osnovi izvlečka iz tabele slučajnih števil (Tabela A) izberite iz neke populacije $N = 300$ enot enostavni slučajni vzorec $n = 10$ enot.

Naloga 6

Vzemimo, da opazovano populacijo predstavljajo 4 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 za vse možne vzorce. V obeh primerih izračunajte še aritmetično sredino iz vseh izračunanih vrednosti za vse možne vzorce.
- Izračunajte vzorčne variance ter nepristranske ocene vzorčnih varianc iz osnovnih štirih podatkov. Kaj ugotovite v primerjavi z rezultati, ki ste jih dobili pri točki e?

Naloga 7

Z uporabo sistematičnega vzorčenja izberite iz populacije $N = 200$ študentov, v vzorec:

- a) $n = 5$ študentov,
- b) $n = 10$ študentov,
- c) $n = 20$ študentov.

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

Naloga 8

Populacija 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 9

Opišite vire podatkov in načine zbiranja podatkov.

Naloga 10

Vprašalniki – prednosti/pomanjkljivosti, vsebina, značilnosti dobrega vprašalnika, načini razdeljevanja, neodgovori.

Naloga 11

Kakšne so lahko oblike vprašalnikov in tipi vprašanj v vprašalniku?

Naloga 12

Kako v poslovnih raziskavah merimo stališča ljudi? Zapišite primer.

Naloga 13

Kako merimo zanesljivost vprašalnika in kdaj je vprašalnik zanesljiv?

Naloga 14

Vzemimo, da želimo izmeriti odnos ljudi v družbi do zaposlovanja žensk. Naše izhodiče je, da je v Sloveniji polovica žensk zaposlenih. Kako bi postopali? Pri tem mnenje ljudi pridobite na podlagi največ 6 trditev (na primer: Vsak človek, ne glede na spol, naj s svojim delom skrbi za svoj obstoj; Ženska naj bo zaposlena le, če ni poročena in nima otrok ipd.)

Naloga 15

V slučajnem vzorcu je bilo zajetih 200 gospodinjstev, ki so letno za potrošnja izdelka A porabila naslednje zneske, ki so prikazani v tabeli 104.

Tabela 104: Letna potrošnja izdelka A

Potrošnja v €	Število gospodinjstev
Nad 500 do 600	18
Nad 600 do 700	42
Nad 700 do 800	66
Nad 800 do 900	33
Nad 900 do 1000	31
Nad 1000 do 1100	10
Skupaj	200

a) Določite in pojasnite:

- dvostranski interval za povprečni letni znesek porabe izdelka A, s 95%-no verjetnostjo
- dvostranski interval za povprečni letni znesek porabe izdelka A, z 75%-no verjetnostjo
- enostranski interval za povprečni letni znesek porabe izdelka A, z 99%-no verjetnostjo.

- b) Na ravni značilnosti $\alpha = 0,10$ preizkusite domnevo, da je povprečni letni znesek porabe izdelka A:
- Enak 750 €
 - Največ 750 €.

Naloga 16

V vzorec je bilo na slučajen način izbranih 7 imetnikov mobilnih telefonov. Le-ti so mesečno za telefon porabili naslednje zneske: 42, 50, 66, 77, 60, 58 in 67 €.

- a) Določite in pojasnite:
- dvostranski interval za povprečni znesek računa za telefon, s 95%-no verjetnostjo
 - dvostranski interval za povprečni znesek računa za telefon, z 80%-no verjetnostjo
 - enostranski interval za povprečni znesek računa za telefon, z 99%-no verjetnostjo.
- b) Na ravni značilnosti $\alpha = 0,10$ preizkusite domnevo, da je povprečni znesek računa za telefon:
- Enak 58 €.
 - Najmanj 58 €.

Naloga 17

Raziskava, ki je imela namen ugotoviti, kakšne so potrebe oseb z normalno mobilnostjo in tistih z zmanjšano mobilnostjo, pri nakupovanju, je bila izvedena na vzorcu 789 oseb, starejših od 65 let. Rezultati statistične analize so prikazani v tabeli 105 (merjeno na skali od 1 – ni pomembno, do 5 – zelo pomembno).

Tabela 105: Rezultati statistične analize

Potreba	Povprečje – normalna mobilnost	Povprečje – zmanjšanja mobilnost	Raven značilnosti (α)
Dostava na dom.	1,787	2,000	0,023
Naročanje po telefonu.	2,030	2,335	0,003
Možen prevoz do trgovine.	2,188	3,098	0,000
Prikladno parkiranje.	4,001	4,095	0,305
Trgovina blizu doma.	3,177	3,325	0,137
Različne trgovine blizu skupaj	3,456	3,681	0,023

- Pojasnite, za kakšna vzorca gre v gornjem primeru, kakšna analiza je bila izvedena in kateri test je bil v tem primeru uporabljen?
- Ali so vse potrebe enako pomembne za obe skupini oseb? Za katere potrebe je bolj verjetno, da jih bodo potrebovale osebe z zmanjšano mobilnostjo?

Naloga 18

Med 40 anketiranimi študenti neke fakultete, ki smo jih zajeli v slučajni vzorec, je bil čas študija na dan pred izpitom v povprečju enak 1,2 uri, $s = 0,4$ ure. Med anketiranimi študenti jih je 26 zadovoljnih s študijskimi prostori v čitalnici.

- Najmanj koliko ur povprečno na dan pred izpitom, študenti študirajo – enostranski interval zaupanja ocenite z 99%-no verjetnostjo.
- Ali lahko trdimo, da študenti dan pred izpitom v povprečju študirajo več kot 1 uro (na ravni značilnosti $\alpha = 5\%$)?
- Ali lahko trdimo, da je najmanj polovica študentov zadovoljnih s študijskimi prostori v čitalnici (na ravni značilnosti $\alpha = 5\%$)?

Naloga 19

Pri analiziranju zgodnje podjetniške aktivnosti (zanima nas, koliko % ljudi je pričelo s kakršnokoli aktivnostjo za ustanovitev lastnega podjetja oziroma novega posla ali pa imajo »mlado« podjetje, ki ne izplačuje plač dalj kot 3,5 let), nas zanima tudi, ali družbeno okolje v Sloveniji z družbenimi in kulturnimi normami, ki vladajo v slovenskem prostoru, podpira podjetništvo.

Anketirane ljudi smo zato povprašali, kako se strinjajo s spodnjimi trditvami, na lestvici od 1-se ne strinjam do 5-zelo se strinjam:

Trditev 1: V Sloveniji večina ljudi želi, da bi vsi imeli enak življenjski standard.

Trditev 2: Ustanovitev novega podjetja je v Sloveniji dobra izbira kariere.

Trditev 3: Uspešni novi podjetniki imajo v Sloveniji visok družbeni status.

Trditev 4: V Sloveniji veliko medijev objavlja zgodbe o uspešnih novih podjetnikih.

Podane imamo rezultate analize zanesljivosti. Analizirajte in vsebinsko pojasnite dobljene rezultate:

Tabela 106: Rezultati statistične analize

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,728	,700	4

Item Statistics

	Mean	Std. Deviation	N
Življenjski standard	3,08	1,519	191
Izbira kariere	2,66	1,816	191
Visok družbeni status	2,65	1,835	191
Mediji	2,65	1,834	191

Inter-Item Correlation Matrix

	Življenjski standard	Izbira kariere	Visok družbeni status	Mediji
Življenjski standard	1,000	,026	,028	-,109
Izbira kariere	,026	1,000	,842	,637
Visok družbeni status	,028	,842	1,000	,783
Mediji	-,109	,637	,783	1,000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	2,762	2,649	3,084	,435	1,164	,046	4
Item Variances	3,085	2,309	3,366	1,057	1,458	,268	4

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Življenski standard	7,96	25,162	-,020	,045	,902
Izbira kariere	8,39	13,860	,740	,710	,522
Visok družbeni status	8,40	12,757	,842	,815	,445
Mediji	8,39	14,882	,630	,632	,595

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
11,05	27,161	5,212	4

Naloga 20

Proizvajalec opreme za kampiranje želi ugotoviti preference kupcev glede barve novega tipa spalnih vreč, ki jih je razvil. Spalne vreče so povprečne kvalitete in se bodo prodajale po temu primernih cenah. Iz izkušenj v preteklosti proizvajalec pričakuje, da bodo spalne vreče svetlejših barv kupovale starejše osebe, spalne vreče temnejših barva pa mlajše osebe. Proizvodne kapacitete pa zahtevajo odločitev o proizvodnji samo ene vrste barv – svetle ali temne. Za lažjo odločitev je bilo na slučajen način izbranih 5 trgovin, v kateri so se v preteklosti prodajale spalne vreče obeh barv – svetle in temne.

Tabela 107: Rezultati statistične analize

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	svetle_b	42,8000	5	7,08520	3,16860
	temne_b	74,6000	5	8,29458	3,70945

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	svetle_b – temne_b	-31,80	12,112	5,417	46,83901	-16,76099	-5,9	4	,004

- a) Za kakšna vzorca gre v proučevanem primeru in katera statistična analiza je bila izvedena?

- b) Ali se je na osnovi rezultatov analize mogoče odločiti za barvo spalnih vreč? Zapišite domnevo, ki smo jo preizkušali ter opišite postopek preizkušanja domneve.

Naloga 21

V okviru izobraževanja zaposlenih sta v opazovanem podjetju na voljo dve metodi izobraževanja. Vodja izobraževanja je na slučajen način izbral 14 zaposlenih za metodo A (v programu SPSS so zaposleni z metodo A razvrščeni v skupino 1) in 12 zaposlenih za metodo B (v programu SPSS so zaposleni z metodo B razvrščeni v skupino 2). Po končanem izobraževanju so testirali pridobljeno znanje, ki so ga zaposleni pridobili po obeh metodah. Znanje se je ocenjevalo na osnovi števila točk na testu.

Ali lahko trdimo, da sta obe metodi enako uspešni? Zapišite hipoteze, ki smo jih testirali in postopek sprejetja ali zavrnitve hipotez. Rezultati so prikazani na sliki 13.

Group Statistics

	A_B	N	Mean	Std. Deviation	Std. Error Mean
TEST	1	14	29,1818	10,17662	3,06837
	2	12	33,2000	16,43032	5,19572

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Upper	Lower
Equal variances assumed	4,406	,049	-,681	19	,504	-4,0182	5,90069	-16,37	8,332
Equal variances not assumed			-,666	14,8	,516	-4,0182	6,03411	-16,90	8,862

Slika 13: Rezultati statistične analize

Naloga 22²

Družbo za upravljanje (DZU) zanima, ali se njihove stranke, ki vlagajo v njihove sklade različno visoke zneske, med seboj razlikujejo po starosti.

Na osnovi rezultatov v preglednici opišite starosti vlagateljev v vseh treh skupinah.

Rezultati statistične analize so prikazani na sliki 14.

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1 skupina	13	31,46	12,467	3,458	23,93	39,00	20	65
2 skupina	12	53,00	7,544	2,178	48,21	57,79	43	65
3 skupina	5	64,20	1,643	,735	62,16	66,24	62	66
Total	30	45,53	16,068	2,934	39,53	51,53	20	66

Slika 14: Rezultati statistične analize

Naloga 23

V smučarskem središču so analizirali spreminjanje števila smučarjev v odvisnosti od višine snega. Rezultati statistične analize so prikazani na sliki 15.

Izpišite in vsebinsko pojasnite vse kazalce enostavne linearne korelacije in regresije.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,642(a)	,413	,392	105,08851

a Predictors: (Constant), visina_snega

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	217140,309	1	217140,309	19,662	,000(a)
	Residual	309220,658	28	11043,595		
	Total	526360,967	29			

a Predictors: (Constant), visina_snega

b Dependent Variable: st_smučarjev

² Naloga se je delno obravnavala pri nalogi 15 (poglavje 3. Preizkušanje domnev o statistično značilnih razlikah med povprečnimi vrednostmi spremenljivke v več vzorcih).

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta	B	Std. Error
1	(Constant)	388,313	53,052		7,319	,000
	visina_snega	1,570	,354	,642	4,434	,000

a Dependent Variable: st_smucarjev

Slika 15: Rezultati statistične analize

Naloga 24

Pri analizi pričakovanega števila gostov v smučarskem središču, je uprava hotela za 30 dni v smučarski sezoni zbrala podatke v tabeli 108.

- Izračunajte in pojasnite vse rezultate opisne statistike za število smučarjev.
- Ali se povprečno število smučarjev statistično značilno razlikuje med vikendom in med tednom?
- Analizirajte in pojasnite vse kazalce multiple korelacije in regresije.

Tabela 108: Podatki

Dan	Število smučarjev	Višina snega (cm)	Najvišja dnevna temperatura (stopinje celzija)	Vikend	
				0 – ne	1 - da
1	402	55	11	0	
2	337	45	12	0	
3	471	40	10	0	
4	610	79	9	1	
5	620	75	8	1	
6	545	60	9	0	
7	523	60	9	0	
8	563	75	9	0	
9	473	120	2	0	
10	358	80	10	0	
11	768	150	1	1	
12	753	150	1	1	
13	485	140	8	0	
14	667	170	1	0	
15	635	170	2	0	
16	648	160	5	0	
17	540	150	6	0	
18	814	170	5	1	
19	796	170	5	1	
20	477	160	9	0	
21	532	160	5	0	

Dan	Število smučarjev	Višina snega (cm)	Najvišja dnevna temperatura (stopinje celzija)	Vikend	
				0 – ne	1 - da
22	632	190	1	0	
23	618	185	4	0	
24	728	186	2	0	
25	820	186	1	1	
26	851	185	1	1	
27	616	210	2	0	
28	665	210	2	0	
29	650	200	1	0	
30	632	200	1	0	

Tabela 109: Izpis rezultatov

Descriptive Statistics

	N	Range	Minimum	Maximum	Sum	Mean		Std. Deviation	Variance
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
st_smučarjev	30	514,00	337,00	851,00	18229,00	607,6333	24,59700	134,72334	18150,378
Valid N (listwise)	30								

Group Statistics

	vikend	N	Mean	Std. Deviation	Std. Error Mean
st_smučarjev	1,00	8	754,0000	91,04002	32,18751
	,00	22	554,4091	105,70293	22,53594

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Upper	Lower
Equal variances assumed	,548	,465	4,729	28	,000	199,59091	42,20870	113,13030	286,05152
Equal variances not assumed			5,080	14,392	,000	199,59091	39,29255	115,53169	283,65013

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,915(a)	,837	,818	57,49038

a Predictors: (Constant), vikend, visina_snega, temp

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	440427,218	3	146809,073	44,418	,000(a)
	Residual	85933,748	26	3305,144		
	Total	526360,967	29			

a Predictors: (Constant), vikend, visina_snega, temp

b Dependent Variable: st_smucarjev

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta	B	Std. Error
1	(Constant)	525,039	80,697		6,506	,000
	visina_snega	,730	,376	,299	1,942	,063
	temp	-12,908	5,619	-,359	-2,297	,030
	vikend	172,719	24,667	,577	7,002	,000

a Dependent Variable: st_smucarjev

8 Naloge za samostojno delo

Naloga 1

V slučajnem vzorcu $n = 200$ kupcev je bila povprečna poraba izdelka 20,8 kosov izdelka v časovni enoti, nepristranska ocena variance pa 38,44 kosov². Izračunajte 95%-ni interval zaupanja za povprečno porabo izdelka v osnovni statistični množici.

Naloga 2

V knjižnem klubu so za 9 naključno izbranih oseb zbrali podatke o višini letnih zneskov za nakup knjig. Zneski so (v €): 600, 800, 1200, 900, 400, 800, 700, 500 in 400. Z 95 % verjetnostjo ocenite povprečen letni znesek za nakup knjig.

Naloga 3

Raziskovalci so želeli analizirati stopnjo izvoza v letu 2021, za deset srednje velikih podjetij v Sloveniji. V tabeli 110 so podatki o stopnji izvoza v letu 2021, za deset srednje velikih podjetij (v % od celotnih prihodkov).

Tabela 110: Podatki o stopnji izvoza v letu 2021, za deset srednje velikih podjetij (v % od celotnih prihodkov)

Podjetje	1	2	3	4	5	6	7	8	9	10
Stopnja izvoza	6,7	9,1	10,7	11,8	12,0	13,1	13,1	13,3	10,6	9,0

- a) Opredelite statistično enoto, statistično spremenljivko in moč vzorca.
- b) Izračunajte povprečno vrednost, mediano, varianco, standardni odklon
- c) Odčitajte modus
- d) Izračunajte variacijski razmik

Naloga 4

Z uporabo sistematičnega vzorčenja izberite iz populacije $N = 400$ podjetij, v vzorec:

- a) $n = 20$ podjetij,
- b) $n = 40$ podjetij.

Izračunajte in pojasnite za posamezni primer vzorčni delež, ter določite prvo in vsa nadaljnja podjetja v vzorec.

Naloga 5

Populacija obsega $N = 1800$ zaposlenih, ki jih opazujemo po osebnem dohodku:

Skupina A: 800 zaposlenih, ki dosegajo osebni dohodek do pod 900 €.

Skupina B: 600 zaposlenih, ki dosegajo osebni dohodek od 900 do pod 1700 €.

Skupina C: 400 zaposlenih, ki dosegajo osebni dohodek 1700 € ali več.

Skupno število statističnih enot v vzorcu naj bo $n = 300$.

Pri izbiri enot v vzorec uporabite Metodo stratificiranega disproporcionalnega vzorčenja.

Naloga 6

V slučajnem vzorcu je bilo zajetih 100 kupcev, ki so v trgovini X mesečno za potrošnjo določenih prehrabnih izdelkov porabili zneske, ki so prikazani v tabeli 111.

Tabela 111: Zneski potrošnje določenih prehrabnih izdelkov

Potrošnja v €	Število kupcev
Nad 40 do 50	30
Nad 50 do 60	22
Nad 60 do 70	19
Nad 70 do 80	16
Nad 80 do 90	13
Skupaj	100

Določite in pojasnite:

- Dvostranski interval za povprečno mesečno potrošnjo določenih prehrabnih izdelkov, s 95%-no verjetnostjo.
- Enostranski interval za povprečno mesečno potrošnjo določenih prehrabnih izdelkov, z 90%-no verjetnostjo.
- Na ravni značilnosti $\alpha = 10\%$ preizkusite domnevo, da je povprečna mesečna potrošnja določenih prehrabnih izdelkov enaka 70 €.

Naloga 7

V vzorec je bilo na slučajen način izbranih 7 študentov Ekonomsko-poslovne fakultete v Mariboru. Le-ti so pri testu pri predmetu X dosegli naslednje točke: 42, 36, 12, 29, 47, 34 in 10.

Določite in pojasnite:

- Dvostranski interval za povprečno doseženo število točk študentov na testu, z 80%-no verjetnostjo.
- Dvostranski interval za povprečno doseženo število točk študentov na testu, s 95%-no verjetnostjo.
- Enostranski interval za povprečno doseženo število točk študentov na testu, z 99%-no verjetnostjo.
- Na ravni značilnosti $\alpha = 5\%$ preizkusite domnevo, da je povprečno doseženo število točk študentov na testu največ 47 točk.
- Na ravni značilnosti $\alpha = 10\%$ preizkusite domnevo, da je povprečno doseženo število točk študentov na testu enako 36 točk.

Naloga 8

V podjetju X so želeli izmeriti motiviranost svojih zaposlenih. Anketirani zaposleni so na 5-stopenjski Likertovi lestvici označili svoje strinjaje s posameznimi trditvami od 1 do 5 (kjer pomeni 1 – popolnoma se ne strinjam in 5 – popolnoma se strinjam):

Trditev 1: Motivira me pohvala za dobro opravljeno delo.

Trditev 2: Motivira me možnost fleksibilnosti na delovnem mestu.

Trditev 3: Motivira me možnost raznolikega opravljanja delovnih nalog.

Trditev 4: Motivira me napredovanje.

Trditev 5: Motivirajo me dobri odnosi z zaposlenimi.

Rezultat analize zanesljivosti vprašalnika je podan v naslednjih dveh tabelah. Vsebinsko pojasnite rezultate.

Reliability Statistics

Cronbach's Alpha	N of Items
,887	5

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Trditev 1	13,28	11,457	,813	,842
Trditev 2	13,03	13,252	,823	,843
Trditev 3	13,82	13,641	,602	,892
Trditev 4	12,59	14,624	,666	,876
Trditev 5	13,27	12,890	,767	,852

Naloga 9

V podjetju X so želeli ugotoviti ali se zaposleni soočajo s stresom pri opravljanju delovnih nalog. Zaposleni so na 7-stopenjski Likertovi lestvici označili svoje strinjaje s posameznimi trditvami od 1 do 7 (kjer pomeni 1-sploš se ne strinjam in 7-popolnoma se strinjam):

Trditev 1: V podjetju se soočam s stresom pri opravljanju delovnih nalog.

Trditev 2: Zaradi stresa na delovnem mestu občutim pomanjkanje energije.

Trditev 3: Zaradi stresa na delovnem mestu imam težave z zbranostjo.

Trditev 4: Čutim pomanjkanje zmogljivosti pri opravljanju delovnih nalog.

Trditev 5: Zaradi stresa mi primanjkuje volje do dela.

Rezultat analize zanesljivosti vprašalnika je podani v naslednjih dveh tabelah. Vsebinsko pojasnite rezultate. Kako bi lahko izboljšali zanesljivost merilnega instrumenta?

Reliability Statistics

Cronbach's Alpha	N of Items
,662	5

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Trditev 1	12,54	8,274	,530	,572
Trditev 2	13,49	5,935	,767	,409
Trditev 3	12,78	7,435	,488	,574
Trditev 4	12,41	12,394	-,256	,845
Trditev 5	13,66	6,158	,744	,429

Naloga 10

V 50 podjetij v Sloveniji so želeli preveriti ali prihaja do statistično značilnih razlik med moškim in ženskim spolom glede povprečnega strinjanja z določeno trditvijo, ki se nanaša na stres. Na osnovi rezultatov preverite domnevo, da se moški in ženske ne razlikujejo glede povprečnega strinjanja s trditvijo: »V podjetju se soočam s stresom pri opravljanju delovnih nalog«. Pojasnite tudi kateri test je uporabljen in zakaj.

Group Statistics

		N	Mean	Std. Deviation	Std. Error Mean
1. V podjetju se soočam s stresom pri opravljanju delovnih nalog.	Moški	622	3,69	,835	,033
	Ženski	464	3,66	,867	,040

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Dif.	95% Confidence Interval of the Dif	
									Lower	Upper
1. V podjetju se soočam s stresom pri opravljanju delovnih nalog.	Equal variances assumed	1,908	,167	,539	1084	,590	,028	,052	-,074	,130
	Equal variances not assumed			,536	976,960	,592	,028	,052	-,075	,131

Naloga 11

V podjetju X so anketirali zaposlene o njihovem zadovoljstvu z delovnimi pogoji. V podjetju X so želeli preveriti ali se povprečna stopnja strinjanja s trditvijo T1:« V podjetju sem zadovoljen/na z omogočanjem fleksibilnega delovnega časa» in povprečna stopnja strinjanja s trditvijo T2:« V podjetju sem zadovoljen/na s količino programov v okviru izobraževanja, usposabljanja», statistično značilno razlikujeta. Pojasnite kateri test je uporabljen in zakaj. Prav tako še pojasnite rezultate v spodnjih tabelah.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	T1. V podjetju sem zadovoljen/na z omogočanjem fleksibilnega delovnega časa.	3,39	1086	,993	,030
	T2. V podjetju sem zadovoljen/na s količino programov v okviru izobraževanja, usposabljanja.	3,19	1086	1,250	,038

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	T1. V podjetju sem zadovoljen/na z omogočanjem fleksibilnega delovnega časa. & T2. V podjetju sem zadovoljen/na s količino programov v okviru izobraževanja, usposabljanja.	1086	,768	,000

		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	T1. V podjetju sem zadovoljen/na z omogočanjem fleksibilnega delovnega časa. – T2. V podjetju sem zadovoljen/na s količino programov v okviru izobraževanja, usposabljanja.	,196	,800	,024	,148	,244	8,075	1085	,000

Naloga 12

Analizirajte ali obstajajo statistično značilne razlike glede tedenskega števila potrjenih primerov s SARS-CoV-2 glede na dve različni lokaciji prenosa okužbe in sicer (1) delovno mesto in (2) trgovina. Podatki o tedenskem številu potrjenih primerov s SARS-CoV-2 na delovnem mestu in v trgovini so prikazani v tabeli 112.

Tabela 112: Podatki o tedenskem številu potrjenih primerov s SARS-CoV-2 na delovnem mestu in v trgovini

Teden	Delovno mesto	Trgovina
2020-47	2573	186
2020-48	2601	216
2020-49	2747	221
2020-50	2694	266
2020-51	2270	227
2020-52	1783	156
2020-53	1276	174
2021-01	1128	190
2021-02	1436	157
2021-03	1499	114
2021-04	1521	147
2021-05	1118	118
2021-06	744	88
2021-07	563	72
2021-08	644	53

Vir: NIJZ, 2021

Naloga 13

Podjetje, ki se ukvarja s prodajo računalniške opreme zanima, ali se njihove stranke, ki kupujejo računalniško opremo, med seboj statistično značilno razlikujejo glede panoge v kateri so zaposleni. Kupce so razdelili v tri skupine po panogah, ter primerjali višino nakupa (v €) kupcev, ki sodijo v določeno panogo. Pojasnite kateri test je uporabljen ter pojasnite prikazane rezultate.

Descriptives

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	32	108,13	74,723	13,209	81,18	135,07	20	300
2	33	112,73	83,751	14,579	83,03	142,42	10	400
3	36	189,17	144,980	24,163	140,11	238,22	20	500
Total	101	138,51	113,008	11,245	116,21	160,82	10	500

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	143860,182	2	71930,091	6,220	,003
Within Groups	1133217,045	98	11563,439		
Total	1277077,228	100			

Multiple Comparisons

Dependent Variable: Prodaja

Tukey HSD

(I) Panoga	(J) Panoga	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-4,602	26,679	,984	-68,09	58,89
	3	-81,042*	26,126	,007	-143,22	-18,87
2	1	4,602	26,679	,984	-58,89	68,09
	3	-76,439*	25,916	,011	-138,11	-14,76
3	1	81,042*	26,126	,007	18,87	143,22
	2	76,439*	25,916	,011	14,76	138,11

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

Naloga 14

Analizirajte ali obstajajo statistično značilne razlike glede tedenskega števila potrjenih primerov s SARS-CoV-2 glede na tri različne lokacije prenosa okužbe in sicer na (1) delovnem mestu, (2) bolnišnica, (3) trgovina. Podatki so prikazani v tabeli 113.

Tabela 113: Tedensko število potrjenih primerov s SARS-CoV-2 glede na tri različne lokacije prenosa okužbe

Teden	Delovno mesto	Bolnišnica	Trgovina
Skupaj	29412	2656	2783
2020-47	2573	191	186
2020-48	2601	254	216
2020-49	2747	239	221
2020-50	2694	257	266
2020-51	2270	177	227
2020-52	1783	186	156
2020-53	1276	103	174
2021-01	1128	98	190
2021-02	1436	175	157
2021-03	1499	149	114
2021-04	1521	144	147
2021-05	1118	105	118
2021-06	744	66	88
2021-07	563	48	72
2021-08	644	74	53
2021-09	565	55	39
2021-10	573	37	45
2021-11	609	60	41
2021-12	730	52	48
2021-13	765	30	61
2021-14	622	41	67
2021-15	514	50	59
2021-16	437	65	38

Vir: NIJZ, 2021

Naloga 15

Pri predmetu X so študenti pisali izpit, na katerem so lahko maksimalno zbrali 50 točk. Analizirajte ali je višina doseženega števila točk na izpitu odvisna od števila ur učenja. Izpišite in vsebinsko pojasnite vse kazalce enostavne linearne regresije, ki so prikazani v spodnjih tabelah.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,817 ^a	,668	,635	10,068

a. Predictors: (Constant), število ur učenja

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2037,363	1	2037,363	20,100	,001 ^b
	Residual	1013,637	10	101,364		
	Total	3051,000	11			

a. Dependent Variable: višina doseženega števila točk

b. Predictors: (Constant), število ur učenja

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	13,269	5,556		2,388	,038
	število ur učenja	,833	,186	,817	4,483	,001

a. Dependent Variable: višina doseženega števila točk

Naloga 16

Pri predmetu X so študenti pisali kolokvij, na katerem so lahko maksimalno zbrali 70 točk. Analizirajte ali je višina doseženega števila točk na kolokviju odvisna od števila ur učenja in števila dni učenja. Izpišite in vsebinsko pojasnite vse kazalce multiple linearne regresije, ki so prikazani v spodnjih tabelah.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,940 ^a	,884	,884	,34073388

a. Predictors: (Constant), število dni učenja, število ur učenja

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	948,870	2	474,435	4086,449	,000 ^b
	Residual	124,343	1071	,116		
	Total	1073,213	1073			

a. Dependent Variable: višina doseženega števila točk

b. Predictors: (Constant), število dni učenja, število ur učenja

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	,003	,010		,310	,757
	število ur učenja	,773	,025	,774	30,758	,000
	število dni učenja	,180	,025	,180	7,155	,000

a. Dependent Variable: višina doseženega števila točk

Naloga 17

Neko podjetje izdeluje digitalne merilnike krvnega pritiska. Z namenom testiranja njihove natančnosti, je bilo testiranih 10 ljudi: vsaki osebi je krvni pritisk najprej izmeril zdravnik, takoj nato pa si je vsak izmed testiranih izmeril krvni pritisk še sam z digitalnim merilnikom. V tabeli 114 so podani podatki meritev ter starost ljudi.

Tabela 114: Podatki meritev in starost ljudi

Oseba	Višina krvnega pritiska, kot jo je izmeril zdravnik	Višina krvnega pritiska, izmerjena z digitalnim merilnikom	Starost
1	112	125	45
2	109	108	40
3	139	116	65
4	141	123	65
5	120	138	45
6	99	123	40
7	128	118	35
8	118	122	35
9	116	116	30
10	120	118	40

- a) Podatke vnesite v SPSS.
- b) Za spremenljivko, ki opisuje višino krvnega pritiska, izmerjenega z digitalnim merilnikom, pojasnite vzorčno aritmetično sredino ter standardni odklon, pojasnite standardno napako ocene aritmetične sredine in 75%-ni interval zaupanja za povprečno vrednost.
- c) Preverite domnevo, da je povprečni krvni tlak, ki ga izmeri zdravnik, enak 110 enot. Zapišite H_0 in H_1 in pojasnite rezultat preverjanja.
- d) Preverite domnevo, da se ljudje, stari do 40 let, značilno razlikujejo glede povprečnega krvnega tlaka, ki jim ga izmeri zdravnik, od ljudi, ki so starejši od 40 let. Zapišite ustrezno postavljene domneve in pojasnite rezultat testiranja.
- e) Preverite domnevo, da se povprečni krvni tlak, ki ga izmeri zdravnik in povprečni krvni tlak, izmerjen z digitalnim merilnikom, med seboj statistično značilno razlikujeta. Zapišite H_0 in H_1 in pojasnite rezultat testiranja.

Naloga 18

1. Primer testa:

1. naloga

Operater mobilne telefonije ponuja aparate, za katere trdi, da je trajanje baterije dalj kot 120 ur. Za preverjanje te trditve so izbrali slučajni vzorec 10 baterij in zabeležili njihove čase trajanja. Dobili so naslednje podatke (v urah):

98	89	91	100	120
110	105	108	130	133

- Zapišite ustrezno postavljeno ničelno in alternativno domnevo.
- Statistični test naj bo izveden pri 5%-ni stopnji značilnosti preizkusa. Kakšen zaključek lahko podate?

2. naloga

Iz seznama z $N = 250$ elementov želimo izbrati slučajni vzorec z 8 elementi. Uporabite tabelo slučajnih števil in določite zaporedne številke iz seznama, če je začetna pozicija izbiranja zgornji desni kot tabele, izbiro pa nadaljujete vodoravno proti levi strani po prvi vrstici, nato vodoravno od desne proti levi po 2. vrstici itd.

3. naloga

Pri analizi odvisnosti višine začetniške plače diplomanta od povprečne ocene študija diplomanta in števila let študija, smo dobili rezultate v spodnjih preglednicah.

- Pojasnite multipli korelacijski koeficient in prilagojeni determinacijski koeficient.
- Preverite ustreznost modela kot celote. Zapišite ustrezno postavljen par hipotez in pojasnite rezultate.
- Preverite in vsebinsko pojasnite, kako neodvisne spremenljivke vplivajo na odvisno. Zapišite ustrezno postavljene pare hipotez in pojasnite rezultate.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,834(a)	,695	,608	134,99239

a Predictors: (Constant), ocena, stevilo_let_studija

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	291079,379	2	145539,690	7,987	,016(a)
	Residual	127560,621	7	18222,946		
	Total	418640,000	9			

a Predictors: (Constant), ocena, stevilo_let_studija

b Dependent Variable: zacetniska_placa

Coefficients^a

		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Err.	Beta		
1	(Constant)	-273,505	994,757		-,275	,791
	stevilo_let_studija	-49,011	42,086	-,356	-1,165	,002
	ocena	169,724	96,586	,537	1,757	,020

a Dependent Variable: zacetniska_placa

Naloga 18

2. Primer testa:

1. naloga

V vzorec je bilo na slučajen način izbranih 10 uporabnikov zavarovalnih storitev, ki jim je poteklo avtomobilsko zavarovanje. Ob ponovni sklenitvi avtomobilskega zavarovanja so uporabniki poravnali naslednje zavarovalne premije v EUR:

854 564 650 1240 850 560 420 1100 1050 874

Določite in pojasnite enostranski interval zaupanja (s spodnjo mejo) za povprečni znesek avtomobilskega zavarovanja, z 99%-no verjetnostjo.

2. naloga

Pojasnite razlike med slučajnim in neslučajnim vzorčenjem (v največ treh stavkih).

3. naloga

V spodnji tabeli prikazujemo rezultate analize, v kateri smo preverjali, ali se dodiplomski študenti in podiplomski študenti statistično značilno razlikujejo med seboj glede povprečne ocene strinjanja o uporabnosti znanja statistike (oznaka spremenljivke US, merjeno na Likertovi lestvici: 1-popolnoma se ne strinjam; 7-popolnoma se strinjam).

- Katero metodo analiziranja podatkov smo uporabili v danem primeru?
- Pojasnite vse podatke, ki so prikazani v tabeli (Group Statistics). Iz tabele za obe skupini študentov izpišite vse vrednosti in jih opišite.
- Zapišite ustrezno postavljeno hipotezo, s katero smo preverjali, ali se dodiplomski in podiplomski študenti med seboj statistično značilno razlikujejo glede povprečne ocene strinjanja o uporabnosti znanja statistike. Ustrezno zapišite hipoteze, s katerimi boste preverjali enakost povprečnih vrednosti za obe skupini študentov.

Group Statistics

	Skupina študentov	N	Mean	Std. Deviation	Std. Error Mean
US	Podiplomski študenti	112	4,77	1,492	0,160
	Dodiplomski študenti	188	3,96	1,767	0,252

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	95% Confidence Interval of the Difference	
							Lower	Upper
US	Equal variances assumed	4,301	0,040	2,845	134	0,005	0,247	1,375
	Equal variances not assumed			2,713	86,485	0,008	0,217	1,405

4. naloga

V spodnjih tabelah prikazujemo rezultate raziskave, s pomočjo katere smo preverjali odvisnost med vrednosmi zavarovalne storitve in zadovoljstvom z zavarovalnico ter zaupanjem v zavarovalnico. Spremenljivke »vrednost zavarovalnih storitev, zaupanje v zavarovalnico in zadovoljstvo uporabnika« so bile merjene na 5-stopenjski lestvici (od 1-zelo nizko do 5-zelo visoko ter od 1-zelo nezadovoljen do 5-zelo zadovoljen).

- Izpišite in pojasnite korelacijski in determinacijski koeficient za dani primer.
- Zapišite in vsebinsko pojasnite hipotezo za testiranje ustreznosti celotnega regresijskega modela.
- Zapišite in vsebinsko pojasnite vse hipoteze za preverjanje statistične značilnosti posameznih regresijskih koeficientov v regresijski funkciji.
- Izpišite regresijsko funkcijo in jo vsebinsko pojasnite.
- Kaj bi na podlagi rezultatov raziskave predlagali zavarovalnici?

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,802 ^a	,643	,636	,579

a. Predictors: (Constant), zadovoljstvo, zaupanje

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	92,387	3	30,796	91,879	,000 ^b
	Residual	51,282	153	,335		
	Total	143,669	156			

a. Dependent Variable: Vrednost zavarovalnih storitev

b. Predictors: (Constant), zadovoljstvo, zaupanje

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Er.	Beta		
1	(Constant)	,445	,215		2,073	,040
	Zaupanje	,275	,089	,291	3,074	,003
	Zadovoljstvo	,498	,098	,479	5,060	,000

a. Dependent Variable: Vrednost zavarovalnih storite

9 Rešitve računskih nalog

Naloga 5

$N = 300$, kar pomeni, da izbiramo trimestna števila (vendar tista, ki niso večja od 300).
 $n = 10$, kar pomeni, da naključno izberemo 10 elementov.

Tako na primer s pomočjo tabele naključnih števil naključno izberemo: 085, 143, 240, 119, itd.

Naloga 6

a)

Porazdelitev parametrov izračunanih iz vseh možnih kombinacij enot v vzorcih (vseh možnih vzorcev):

$$K_n^N = \binom{N}{n} = \frac{N!}{(N-n)!n!}$$

$$K = \frac{N}{n} = \frac{4!}{(4-2)! \cdot 2!} = 6 - \text{število vseh možnih vzorcev}$$

b)

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

c)

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.

e)

Vzorčne variance (σ^2) za vse možne vzorce izračunamo po formuli:

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

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 formuli:

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

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 vseh izračunanih vrednosti: $\frac{10}{6} = 1,667$

f)

Izračun σ^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 nepriustranskih ocen variance vseh možnih vzorcev je enaka nepriustranski oceni variance izračunane na osnovi podatkov iz celotne množice.

Naloga 7

$N = 200$ študentov

a)

$n = 5$ študentov

Vzorčni delež = $n/N = 5/200 = 1/40$, kar pomeni, da v vzorec izberemo vsakega 40-tega študenta iz populacije.

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.: $130 + 40 = 170$, itd.

b)

$n = 10$ študentov

Vzorčni delež = $n/N = 10/200 = 1/20$, kar pomeni, da v vzorec vključimo vsakega 20-tega študenta iz populacije.

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

c)

$n = 20$ študentov.

Vzorčni delež = $n/N = 20/200 = 1/10$, kar pomeni, da v vzorec vključimo vsakega 10-tega študenta iz populacije.

Naloga 8

Ponovitev korakov pri stratificiranem vzorčenju:

Korak 1	Oprelitev 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) · p_i
	Korak 9: Izbira izračunanega števila elementov iz vsakega stratuma s tehniko slučajnega vzorčenja

1. KORAK:

Oprelitev populacije: $N = 1.500$ zaposlenih oseb

Statistična enota ali element: 1 zaposleni

2. KORAK:

Oprelitev 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)

$$p = n/N$$

$$p_A = 700/1500 = 0,467 \quad p_B = 500/1500 = 0,333 \quad p_C = 300/1500 = 0,200$$

8. KORAK (določitev števila elementov, ki jih izberemo iz posameznega stratuma)

$n = p \cdot n$ (podano v nalogi: Skupno število statističnih enot v vzorcu naj bo $n = 200$)

$$n_A = 0,467 \cdot 200 = 93,4 = 93 \text{ (93 oseb)} \quad n_B = 0,333 \cdot 200 = 66,6 = 67 \quad n_C = 0,200 \cdot 200 = 40$$

9. KORAK (izbira izračunanega števila elementov iz vsakega stratuma s tehniko slučajnega vzorčenja – npr. s sistematičnim vzorčenjem)

Vzorčni delež = n/N

A: vzorčni delež = $93/700 = 1/7,52 = 1/8$ (izberemo vsakega 8. zaposlenega) \rightarrow npr. prvo izbrano št. je 2., nato 10, 18, 26, 34, itd.

B: vzorčni delež = $67/500 = 1/7,46 = 1/7$ (izberemo vsakega 7. zaposlenega) → npr. prvo izbrano št. je 4., nato 11, 18, 25, 32, itd.

C: vzorčni delež = $40/300 = 1/7,5 = 1/8$ (izberemo vsakega 8. zaposlenega)

6. KORAK: DISPROPORCIONALNA STRATIFIKACIJA (točka b)

7. KORAK (določitev deleža populacije ki je zajeta v stratum):

Število enot iz vsakega stratuma = velikost vzorca (n) / število stratumov (k) = $200 / 3 = 67$

8. KORAK:

Izbira elementov iz vsakega stratuma s tehniko sistematičnega vzorčenja:

A: vzorčni delež = $67/700 = 1/10,4 = 1/10$ (izberemo vsakega 10. zaposlenega)

B: vzorčni delež = $67/500 = 1/7,46 = 1/7$ (izberemo vsakega 7. zaposlenega)

C: vzorčni delež = $67/300 = 1/4,48 = 1/4$ (izberemo vsakega 4. zaposlenega)

Naloga 15

Potrošnja v €	Število gospodinjstev (f_k)	Sredina razreda (y_k)	$(y_k - \bar{y})^2$
Nad 500 do 600	18	550	49952,25
Nad 600 do 700	42	650	15252,25
Nad 700 do 800	66	750	552,25
Nad 800 do 900	33	850	5852,25
Nad 900 do 1000	31	950	31152,25
Nad 1000 do 1100	10	1050	76452,25
Skupaj	200		

Uporabimo formulo za aritmetično sredino iz frekvenčne porazdelitve:

$$\bar{y} = \frac{1}{N} \sum_{k=1}^r f_k y_k$$

$$\bar{y} = \frac{1}{200} \cdot 154700 = 773,5 \text{ €}$$

Uporabimo formulo za vzorčno varianco iz frekvenčne porazdelitve:

$$s^2 = \frac{1}{n-1} \sum_{k=1}^r f_k (y_k - \bar{y})^2$$

$$s^2 = \frac{1}{199} \cdot 3499550 = 17585,68$$

Standardni odklon:

$$s = \sqrt{17585,68} = 132,61$$

Nato izračunamo standardno napako ocene aritmetične sredine:

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

$$se_{\bar{y}} = 132,61 / \sqrt{200} = 9,377$$

A)

a) Dvostranski interval za povprečni letni znesek porabe izdelka A, s 95%-no verjetnostjo:

Imamo velik vzorec, ker je $n = 200$ ter dvostransko intervalno ocenjevanje aritmetične sredine:

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

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

$$P(773,5 - 1,96 \cdot 9,377 < \bar{y} < 773,5 + 1,96 \cdot 9,377) = 95 \%$$

$$P(755,12 < \bar{y} < 791,88) = 95 \%$$

S 95%-no verjetnostjo ocenjujemo, da je povprečni letni znesek porabe izdelka A med 755,12 € in 791,88 €.

b) Dvostranski interval za povprečni letni znesek porabe izdelka A, z 75%-no verjetnostjo:

$$\gamma = 75 \% \rightarrow \alpha = 25 \%, z = \pm 1,15$$

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

$$P(773,5 - 1,15 \cdot 9,377 < \bar{y} < 773,5 + 1,15 \cdot 9,377) = 75 \%$$

$$P(762,72 < \bar{y} < 784,28) = 75 \%$$

S 75%-no verjetnostjo ocenjujemo, da je povprečni letni znesek porabe izdelka A med 762,72 € in 784,28 €.

c) Enostranski interval za povprečni letni znesek porabe izdelka A, z 99%-no verjetnostjo:

Vzamemo formulo za enostransko intervalno ocenjevanje aritmetične sredine:

$$\gamma = 99 \% \rightarrow \alpha = 1 \%, z = 2,33$$

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

$$P(\bar{y} < 773,5 + 2,33 \cdot 9,377) = 99 \%$$

$$P(\bar{y} < 795,35) = 99 \%$$

Z 99%-no verjetnostjo ocenjujemo, da je povprečni letni znesek porabe izdelka A največ 795,35 €.

B)

a) Na ravni značilnosti $\alpha = 0,10$ preizkusimo domnevo, da je povprečni letni znesek porabe izdelka A enak 750 €:

– Imamo dvostransko preizkušanje domneve.

$$\alpha = 0,10 \text{ oz. } 10 \% \rightarrow z = \pm 1,645$$

Zastavimo ničelno in raziskovalno domnevo:

$$H_0: \bar{y}_D = 750 \text{ €}$$

$$H_1: \bar{y}_D \neq 750 \text{ €}$$

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

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

$$z = \frac{773,5 - 750}{9,377} = \mathbf{2,51} \text{ (vidimo, da dobljeno število ne pade v interval } \pm 1,645\text{)}$$

Na osnovi rezultata vidimo, da povprečni letni znesek porabe izdelka A ni enak 750 €, tako, da zavrnamo domnevo H_0 in sprejmemo raziskovalno domnevo H_1 .

b) Na ravni značilnosti $\alpha = 0,10$ preizkusimo domnevo, da je povprečni letni znesek porabe izdelka A največ 750 €.

– Imamo enostransko preizkušanje domneve.

$$\alpha = 0,10 \text{ oz. } 10 \% \rightarrow z = 1,28$$

Zastavimo ničelno in raziskovalno domnevo:

$$H_0: \bar{y}_D \leq 750 \text{ €}$$

$$H_1: \bar{y}_D > 750 \text{ €}$$

$$z = \frac{773,5 - 750}{9,377} = \mathbf{2,51} \text{ (vidimo, da dobljeno število ne pade v interval } + 1,28\text{)}$$

Na osnovi rezultata vidimo, da povprečni letni znesek porabe izdelka A ne znaša največ 750 €, tako, da zavrnamo domnevo H_0 in sprejmemo raziskovalno domnevo H_1 .

Naloga 16

Uporabimo formulo za aritmetično sredino iz posamičnih vrednosti:

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

$$\bar{y} = \frac{420}{7} = 60 \text{ €}$$

Uporabimo formulo za vzorčno varianco iz posamičnih vrednosti:

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

$$s^2 = \frac{1}{6} \cdot 802 = 133,67$$

Standardni odklon:

$$s = \sqrt{133,67} = 11,56$$

Standardna napaka ocene aritmetične sredine: $se_{\bar{y}} = 11,56 / \sqrt{7} = 4,370$

A)

a) Dvostranski interval za povprečni znesek računa za telefon, s 95%-no verjetnostjo:

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

$$P(\bar{Y} - t_{n-1; \alpha/2} \cdot se_{\bar{y}} < \bar{y} < \bar{Y} + t_{n-1; \alpha/2} \cdot se_{\bar{y}}) = \gamma$$

$$P(60 - 2,447 \cdot 4,370 < \bar{y} < 60 + 2,447 \cdot 4,370) = 95 \%$$

$$P(49,31 < \bar{y} < 70,69) = 95 \%$$

S 95%-no verjetnostjo ocenjujemo, da je povprečni znesek računa za telefon med 49,31 € in 70,69 €.

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

$$\alpha/2 = 0,05/2 = 0,025$$

$$(\gamma = 95 \% \rightarrow \alpha = 5\%)$$

b) Dvostranski interval za povprečni znesek računa za telefon, z 80%-no verjetnostjo:

$$P(\bar{Y} - t_{n-1; \alpha/2} \cdot se_{\bar{y}} < \bar{y} < \bar{Y} + t_{n-1; \alpha/2} \cdot se_{\bar{y}}) = \gamma$$

$$P(60 - 1,440 \cdot 4,370 < \bar{y} < 60 + 1,440 \cdot 4,370) = 80 \%$$

$$P(53,71 < \bar{y} < 66,29) = 80 \%$$

Z 80 %-no verjetnostjo ocenjujemo, da je povprečni znesek računa za telefon med 53,71 € in 66,29 €.

$$t_{n-1; \alpha/2} = t_{6; 0,1} = 1,440 \text{ (gledamo tabelo: Kritična vrednost za t porazdelitev)}$$

$$\alpha/2 = 0,2/2 = 0,1$$

$$(\gamma = 80 \% \rightarrow \alpha = 20\%)$$

c) Enostranski interval za povprečni znesek računa za telefon, z 99%-no verjetnostjo:

$$P(\bar{y} < \bar{Y} + t_{n-1; \alpha} \cdot se_{\bar{y}}) = \gamma$$

$$P(\bar{y} < 60 + 3,143 \cdot 4,370) = 99 \%$$

$$P(\bar{y} < 73,73) = 99 \%$$

Z 99 %-no verjetnostjo ocenjujemo, da je povprečni znesek računa za telefon največ 73,73 €.

$$t_{n-1; \alpha} = t_{6; 0,01} = 3,143 \text{ (gledamo tabelo: Kritična vrednost za t porazdelitev)}$$

$$(\gamma = 99 \% \rightarrow \alpha = 1 \%)$$

B)

a) Na ravni značilnosti $\alpha = 0,10$ preizkusimo domnevo, da je povprečni znesek računa za telefon enak 58 €.

– Imamo dvostransko preizkušanje domneve.

$\alpha = 0,10$ oz. 10 % $\rightarrow t_{n-1; \alpha/2} = t_{6; 0,05} = 1,943$ (gledamo tabelo: Kritična vrednost za t porazdelitev)

Zastavimo ničelno in raziskovalno domnevo:

$$H_0: \bar{y}_D = 58 \text{ €}$$

$$H_1: \bar{y}_D \neq 58 \text{ €}$$

Izračunamo t po formuli (izračun testne vrednosti pri preizkušanju domneve o aritmetični sredini – t -test):

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

$$t = \frac{60 - 58}{4,370} = \mathbf{0,46} \text{ (vidimo, da dobljeno število pade v interval } \pm 1,943\text{)}$$

Na osnovi rezultata vidimo, da je povprečni znesek računa za telefon enak 58 €, tako, da sprejmemo domnevo H_0 .

b) Na ravni značilnosti $\alpha = 0,10$ preizkusimo domnevo, da je povprečni znesek računa za telefon najmanj 58 €.

– Imamo enostransko preizkušanje domneve.

$\alpha = 0,10$ oz. 10 % $\rightarrow t_{n-1; \alpha} = t_{6; 0,10} = 1,440$ (gledamo tabelo: Kritična vrednost za t porazdelitev)

Zastavimo ničelno in raziskovalno domnevo:

$$H_0: \bar{y}_D \geq 58 \text{ €}$$

$$H_1: \bar{y}_D < 58 \text{ €}$$

$$t = \frac{60 - 58}{4,370} = \mathbf{0,46} \text{ (vidimo, da dobljeno število pade v interval } - 1,440\text{)}$$

Na osnovi rezultata vidimo, da povprečni znesek računa za telefon znaša najmanj 58 €, tako, da sprejmemo domnevo H_0 .

Naloga 18

$n = 40$ študentov (veliki vzorec)

$$\bar{Y} = 1,2 \text{ h}$$

$$s = 0,4 \text{ h}$$

Standardna napaka ocene aritmetične sredine: $se_{\bar{y}} = 0,4 / \sqrt{40} = 0,063$

a)

$$\gamma = 99 \% \rightarrow \alpha = 1 \%, z = -2,33$$

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

$$P(\bar{y} > 1,2 - 2,33 \cdot 0,063) = 99 \%$$

$$P(\bar{y} > 1,05) = 99 \%$$

Z 99%-no verjetnostjo ocenjujemo, da povprečno na dan pred izpitom študenti študirajo najmanj 1,05 ur.

b)

Enostransko preizkušanje domneve:

$$\alpha = 5 \% \rightarrow z = -1,645$$

Zastavimo ničelno in raziskovalno domnevo:

$$H_0: \bar{y}_D \geq 1 \text{ h}$$

$$H_1: \bar{y}_D < 1 \text{ h}$$

$$z = \frac{1,2 - 1}{0,063} = \mathbf{3,18}$$

Na osnovi rezultata vidimo, da študenti dan pred izpitom v povprečju študirajo 1 uro ali manj, sprejmemo domnevo H_0 .

c)

Enostransko preizkušanje domneve:

najmanj polovica študentov (50 % študentov)

 $n = 40$ študentov $n_a = 26$ zadovoljnih študentov $\alpha = 5\% \rightarrow -1,645$

Zastavimo ničelno in raziskovalno domnevo:

 $H_0: \pi_D \geq 50\%$ $H_1: \pi_D < 50\%$

Izračun strukturnega odstotka iz vzorca:

$$p = 100 \frac{n_a}{n}$$

$$p = \frac{26}{40} \cdot 100 = 65\%$$

Izračun standardne napake ocene strukturnega odstotka:

$$SE_\pi = \sqrt{\frac{p(100-p)}{n}}$$

$$se_\pi = \sqrt{\frac{65(100-65)}{40}} = 7,54$$

Izračun testne vrednosti pri preizkušanju domneve o strukturnem deležu – z-test:

$$z_{izr} = \frac{p - \pi_D}{se_\pi}$$

$$z = \frac{65-50}{7,54} = 1,99$$

Na osnovi rezultata vidimo, da je najmanj polovica študentov zadovoljnih s študijskimi prostori v čitalnici in sprejmemo domnevo H_0 .

Rešitve računskih nalog za samostojno delo:

Naloga 1

Dvostransko intervalno ocenjevanje aritmetične sredine

$n = 200$ kupcev (velik vzorec)

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

$se_{\bar{y}} = 6,2/\sqrt{200} = 0,438$

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

$P(20,8 - 1,96 \cdot 0,438 < \bar{y} < 20,8 + 1,96 \cdot 0,438) = 95\%$

$P(19,941 < \bar{y} < 21,658) = 95\%$

S 95%-no verjetnostjo ocenjujemo, da se povprečna poraba izdelka v osnovni statistični množici giblje med 19,941 in 21,658 enot.

Naloga 2

Dvostransko intervalno ocenjevanje aritmetične sredine

$n = 9$ (mali vzorec)

$\bar{y} = 700$ €

$s^2 = 67500$ €²

$s = 259,808$ €

$se_{\bar{y}} = 86,603$ €

$P(\bar{Y} - t_{n-1; \alpha/2} \cdot se_{\bar{y}} < \bar{y} < \bar{Y} + t_{n-1; \alpha/2} \cdot se_{\bar{y}}) = \gamma$

$P(700 - 2,306 \cdot 86,603 < \bar{y} < 700 + 2,306 \cdot 86,603) = 95\%$

$P(500,293 < \bar{y} < 899,707) = 95\%$

Z 95 %-no verjetnostjo ocenjujemo, da je povprečen letni znesek za nakup knjig med 500,293 € in 899,707 €.

$t_{n-1; \alpha/2} = t_{8; 0,025} = 2,306$ (gledamo tabelo: Kritična vrednost za t porazdelitev)

$\alpha/2 = 0,05/2 = 0,025$

($\gamma = 95\% \rightarrow \alpha = 5\%$)

Naloga 3

b)

Povprečna vrednost:

$$\bar{y} = 10,94 \% \text{ od celotnih prihodkov}$$

Mediana:

– Podatke uredimo v ranžirno vrsto ter nato odčitamo mediano:

Me: $\frac{n+1}{2} = \frac{10+1}{2} = 5,5$ (poiščemo vrednost v ranžirni vrsti in ugotovimo, da vrednost leži med 5. in 6. mestom)

$$\text{Me: } \frac{10,7+11,8}{2} = 11,25 \% \text{ od celotnih prihodkov}$$

Varianca:

$$\sigma^2 = 4,2064 \% \text{ od celotnih prihodkov}^2$$

Standardni odklon:

$$\sigma = 2,051 \% \text{ od celotnih prihodkov}$$

c)

Modus: 13,1 % od celotnih prihodkov

d)

Variacijski razmik: $13,3 - 6,7 = 6,6$ %-nih točk

Naloga 4

$N = 400$ podjetij

a)

$n = 20$ podjetij

Vzorčni delež = $n/N = 20/400 = 1/20$, kar pomeni, da v vzorec vključimo vsako 20-to podjetje iz populacije (na primer: 10, 30, 50, 70, 90, itd).

b)

$n = 40$ podjetij

Vzorčni delež = $n/N = 40/400 = 1/10$, kar pomeni, da v vzorec vključimo vsako 10-to podjetje iz populacije (na primer 5, 15, 25, 35, 45, itd).

Naloga 5

1. KORAK:

Oprelitev populacije: $N = 1.800$ zaposlenih oseb

Statistična enota ali element: 1 zaposleni

2. KORAK:

Oprelitev stratumov:

$k = 3$ stratumi:

Stratum A: osebni dohodek do pod 900 €

Stratum B: osebni dohodek od 900 € do pod 1700 €

Stratum C: osebni dohodek 1700 € ali več

3. in 4. KORAK:

Razvrstitev elementov po stratumih:

Stratum A: zaporedne št. od 1 do 800

Stratum B: zaporedne št. od 1 do 600

Stratum C: zaporedne št. od 1 do 400

5. KORAK

Določitev velikosti vzorca: $n = 300$

6. KORAK: Disproporcionalna stratifikacija**7. KORAK** (določitev deleža populacije ki je zajeta v stratum):

Število enot iz vsakega stratuma = velikost vzorca (n) / število stratumov (k) = $300 / 3 = 100$

8. KORAK:

Izbira elementov iz vsakega stratuma s tehniko sistematičnega vzorčenja:

A: vzorčni delež = $100/800 = 1/8$ (izberemo vsakega 8. zaposlenega)

B: vzorčni delež = $100/600 = 1/6$ (izberemo vsakega 6. zaposlenega)

C: vzorčni delež = $100/400 = 1/4$ (izberemo vsakega 4. zaposlenega)

Naloga 6

Potrošnja v €	Število kupcev (f_k)	Sredina razreda (y_k)	$(y_k - \bar{y})^2$
Nad 40 do 50	30	45	256
Nad 50 do 60	22	55	36
Nad 60 do 70	19	65	16
Nad 70 do 80	16	75	196
Nad 80 do 90	13	85	576
Skupaj	100		

Uporabimo formulo za aritmetično sredino iz frekvenčne porazdelitve: $\bar{y} = 61$ €

Uporabimo formulo za vzorčno varianco iz frekvenčne porazdelitve: $s^2 = 195,9596 \text{ €}^2$

Standardni odklon: $s = \sqrt{195,9596} = 13,9986 \text{ €}$

Standardna napaka ocene aritmetične sredine: $se_{\bar{y}} = 13,9986/\sqrt{100} = 1,39986 \text{ €}$

a) Dvostranski interval za povprečno mesečno potrošnjo določenih prehrabnih izdelkov, s 95%-no verjetnostjo:

Imamo velik vzorec, ker je $n = 100$ ter dvostransko intervalno ocenjevanje aritmetične sredine:

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

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

$$P(61 - 1,96 \cdot 1,39986 < \bar{y} < 61 + 1,96 \cdot 1,39986) = 95 \%$$

$$P(58,2563 < \bar{y} < 63,7437) = 95\%$$

S 95%-no verjetnostjo ocenjujemo, da je povprečna mesečna potrošnja določenih prehrabnih izdelkov med 58,2563 € in 63,7437 €.

b) Enostranski interval za povprečno mesečno potrošnjo določenih prehrabnih izdelkov, z 90%-no verjetnostjo:

$$\gamma = 90\% \rightarrow \alpha = 10\%, z = +1,28$$

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

$$P(\bar{y} < 61 + 1,28 \cdot 1,39986) = 90 \%$$

$$P(\bar{y} < 62,7918) = 90 \%$$

Z 90%-no verjetnostjo ocenjujemo, da je povprečna mesečna potrošnja določenih prehrabnih izdelkov največ 62,7918 €.

c) Na ravni značilnosti $\alpha = 10 \%$ preizkusimo domnevo, da je povprečna mesečna potrošnja določenih prehrabnih izdelkov enaka 70 €.

– Imamo dvostransko preizkušanje domneve.

$$\alpha = 10 \% \rightarrow z = \pm 1,645$$

Zastavimo ničelno in raziskovalno domnevo:

$$H_0: \bar{y}_D = 70 \text{ €}$$

$$H_1: \bar{y}_D \neq 70 \text{ €}$$

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

$$z = \frac{61 - 70}{1,39986} = -6,429$$

Na osnovi rezultata vidimo, da povprečna mesečna potrošnja določenih prehrabnih izdelkov ni enaka 70 €, zato zavrnemo ničelno domnevo H_0 in sprejmemo raziskovalno domnevo H_1 .

Naloga 7

$n = 7$ študentov (mali vzorec)

Uporabimo formula za aritmetično sredino iz posamičnih vrednosti: $\bar{y} = \frac{210}{7} = 35$ točk

Uporabimo formulo za vzorčno varianco iz posamičnih vrednosti: $s^2 = 230,833$ točk²

Standardni odklon: $s = 15,193$ točk

Standardna napaka ocene aritmetične sredine: $se_{\bar{y}} = 15,193/\sqrt{7} = 5,742$ točk

a)

$$P(\bar{Y} - t_{n-1; \alpha/2} \cdot se_{\bar{y}} < \bar{y} < \bar{Y} + t_{n-1; \alpha/2} \cdot se_{\bar{y}}) = \gamma$$

$$P(35 - 1,440 \cdot 5,742 < \bar{y} < 35 + 1,440 \cdot 5,742) = 80 \%$$

$$P(26,732 < \bar{y} < 43,268) = 80 \%$$

b)

$$P(\bar{Y} - t_{n-1; \alpha/2} \cdot se_{\bar{y}} < \bar{y} < \bar{Y} + t_{n-1; \alpha/2} \cdot se_{\bar{y}}) = \gamma$$

$$P(35 - 2,447 \cdot 5,742 < \bar{y} < 35 + 2,447 \cdot 5,742) = 95 \%$$

$$P(20,949 < \bar{y} < 49,051) = 95 \%$$

c)

$$P(\bar{y} < \bar{Y} + t_{n-1; \alpha} \cdot se_{\bar{y}}) = \gamma$$

$$P(\bar{y} < 35 + 3,143 \cdot 5,742) = 99 \%$$

$$P(\bar{y} < 53,047) = 99 \%$$

d)

– Imamo enostransko preizkušanje domneve:

$$\alpha = 5 \% \rightarrow t_{n-1; \alpha} = t_{6; 0,05} = 1,943 \text{ (gledamo tabelo: Kritična vrednost za t porazdelitev)}$$

Zastavimo ničelno in raziskovalno domnevo:

$$H_0: \bar{y}_D \leq 47 \text{ točk}$$

$$H_1: \bar{y}_D > 47 \text{ točk}$$

$$t = \frac{35 - 47}{5,742} = \mathbf{-2,089}$$

Na osnovi rezultata vidimo, da je povprečno doseženo število točk študentov na testu največ 47 točk, sprejmemo domnevo H_0 .

e)

- Imamo dvostransko preizkušanje domneve:

$$\alpha = 10 \% \rightarrow t_{n-1; \alpha/2} = t_{6; 0,05} = 1,943$$

Zastavimo ničelno in raziskovalno domnevo:

$$H_0: \bar{y}_D = 36 \text{ točk}$$

$$H1: \bar{y}_D \neq 36 \text{ točk}$$

$$t = \frac{35 - 36}{5,742} = -0,174$$

Na osnovi rezultata vidimo, da je da je povprečno doseženo število točk študentov na testu enako 36 točk, tako, da sprejmemo domnevo H_0 .

Naloga 18

1. Primer testa - 1. naloga:

– Imamo enostransko preizkušanje domneve.

$$\alpha = 5 \% \rightarrow t_{n-1; \alpha} = t_{9; 0,05} = -1,833$$

Zastavimo ničelno in raziskovalno domnevo:

$$H_0: \bar{y}_D \geq 120 \text{ ur}$$

$$H_1: \bar{y}_D < 120 \text{ ur}$$

$$\bar{y} = 108,4 \text{ ur}$$

$$s^2 = 230,93 \text{ ur}^2$$

$$s = 15,2 \text{ ur}$$

$$se_{\bar{y}} = 4,8 \text{ ur}$$

$$t = \frac{108,4 - 120}{4,8} = -2,4$$

Na osnovi rezultata zavrnemo domnevo H_0 in sprejmemo raziskovalno domnevo H_1 ($H_1: \bar{y}_D < 120 \text{ ur}$).

Naloga 19

2. Primer testa - 1. naloga:

$$\bar{y} = 816,2 \text{ €}$$

$$s^2 = 634563,6 \text{ €}^2$$

$$s = 796,59 \text{ €}$$

$$se_{\bar{y}} = 252,08 \text{ €}$$

$$P(\bar{y} > \bar{Y} - t_{n-1; \alpha} \cdot se_{\bar{y}}) = \gamma$$

$$P(\bar{y} > 816,2 - 2,821 \cdot 252,08) = 99 \%$$

$$P(\bar{y} > 105,08) = 99 \%$$

Z 99%-no verjetnostjo ocenjujemo, da je povprečni znesek avtomobilskega zavarovanja najmanj 105,08 €.

Tabele

Tabela naključnih števil

(Vir: Artenjak, 2003)

08508	32106	51974	62471	93713	14379	24030	05903	11939
45656	66295	23207	02395	92939	40798	05132	89590	34718
33928	10372	92243	15724	97214	10531	37551	62320	40291
69937	12892	88085	89939	14899	42287	40520	72675	17779
74112	74853	16691	24751	30053	59522	61124	41543	40300
90689	42932	11302	69784	78000	24339	96847	37829	86554
03254	33465	64264	51794	94993	95086	40582	67593	03775
85614	88771	22228	63325	44863	72073	24089	20052	45719
86498	38126	30831	94378	07600	91048	22347	65840	53920
83710	48753	64771	56284	35347	41171	06772	34141	16905
54812	28862	57363	61558	68877	33649	07137	14646	96808
46069	29295	73449	14992	81522	16786	72843	29487	91214
98806	00378	56935	02019	50106	80126	58851	70486	50363
39619	81264	51349	82951	40546	32112	97243	27570	73768
39231	19363	26973	14083	85996	72852	36584	83179	68575
72907	87771	85453	09976	00089	13250	60258	98907	96324
28128	75037	45104	10532	17943	19126	12928	68482	72721
15600	82635	88096	47220	94724	34436	50844	74834	16590
06705	62019	52549	89542	03013	94503	48711	81546	39549
61683	65631	85761	51897	28139	88612	29759	53081	20029
45001	65301	15429	57662	41544	88545	58665	41584	89122
21119	64390	15357	69790	44458	66118	76133	60562	84150
86844	36815	04236	61121	45315	29411	37089	63186	15953
74596	87178	40209	70162	77471	62551	05670	97891	63173
23717	76219	13359	40501	28224	82466	43233	91302	44408
45054	64532	75437	13650	49820	38253	59968	83895	14014
85289	00221	69338	51689	16090	81156	02011	16977	60327
26175	26872	90257	07236	71620	23072	88762	20341	25006
63259	18678	52021	78626	78021	57992	15867	27011	50842
02773	43851	59382	95793	49815	05937	71602	34895	34581
61252	80074	59276	84763	11223	22891	75671	46649	72755
93219	73585	44469	15082	25169	26028	87826	97213	78270
42027	52124	60859	18839	79449	28637	34028	20485	42414
41585	51383	22464	02665	12335	49753	77739	34741	27747
15259	06071	17126	72741	22580	65932	75921	58615	08893
28528	71994	84355	73719	53244	43648	68087	84599	27298
05750	02579	43123	81273	24023	79502	92791	88577	83368
12741	62322	24707	84707	94484	24336	35244	66572	40374
12578	12124	16639	12851	16623	23112	54056	79660	37044
39716	83950	96478	21089	37788	11160	06230	91654	55359
86057	28363	03497	34034	25556	29928	11818	89048	51863
42738	72520	83942	66757	17989	82728	38784	43394	92825
43634	11750	12528	28826	06678	68756	80090	26774	65589
95728	05421	39144	69996	94444	93051	42212	15502	95447
88615	29148	33922	64326	82835	59113	78432	17519	29045

Ploščine $H(z)$ za standardizirano normalno porazdelitev

(Vir: Artenjak, 2003)

$$z = \frac{y - \bar{y}}{\sigma}$$

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.

Kritične vrednosti za t porazdelitev

(Vir: Artenjak, 2003)

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

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

Literatura in viri

- Agresti, A., Finlay, B. (2009). *Statistical methods for the Social sciences*. Boston: Pearson
- Artenjak, J. (2003). *Poslovna statistika, Prenovljena in dopolnjena izdaja*. Maribor: UM Ekonomsko-poslovna fakulteta.
- Bastič, M. (2006). *Metode raziskovanja*. Maribor: UM Ekonomsko-poslovna fakulteta.
- Corder, G. W., Foreman, D.I. (2014). *Nonparametric Statistics: A Step-by-Step Approach, 2nd Edition*. New Jersey : John Wiley & Sons
- Heumann, C., Schomaker, M., Shalabh, S. (2016). *Introduction to Statistics and Data Analysis*. Singapore: Springer
- Jackson, S. L. (2012). *Research methods and statistics: A critical thinking approach(4thed.)*. California: Wadsworth Publishing Company
- Leskovar Špacapan, G., Tominc, P. (2008/2009), *Statistične poslovne raziskave in Poslovna statistika II, Učno gradivo za seminarske in laboratorijske vaje, interno gradivo, Ekonomsko-poslovna fakulteta Maribor*.
- Landau, S., Everitt, B. S. (2004). *A Handbook of Statistical Analyses using SPSS*. Washington: CHAPMAN & HALL/CRC
- NIJZ. (2021). *Dnevno spremljanje okužb s SARS-CoV-2 (COVID-19)*. Pridobljeno na spletni strani: <https://www.nijz.si/sl/dnevno-spremljanje-okuzb-s-sars-cov-2-covid-19>
- Rožman, M., Tominc, P. (2019). *Poslovna statistika 1 in Statistika (drugi del predmeta), Gradivo: Naloge za laboratorijske vaje, EPF UM*.
- Tominc, P. (2008/2009). *Izbrana poglavja iz poslovne statistike, Učno gradivo pri predmetih Statistične poslovne raziskave in Poslovna statistika II. Ekonomsko poslovna fakulteta Maribor, Univerza v Mariboru Maribor, študijsko leto 2008/2009*.
- Sakaran, U., Bougie, R. (2016). *Research Methods for Business, A Skill-Building Approach, 7th Ed.*, John Wiley & Sons.
- Selvamuthu, D., Das, D. (2018). *Introduction to Statistical Methods, Design of Experiments and Statistical Quality Control*. Singapore: Springer
- Tabachnick, B.G., Fidell, L.S. (2013). *Using multivariate statistics*. Boston: Pearson
- Tominc, P., Kramberger, T. (2007). *Statistične metode v logistiki*. Celje: UM Fakulteta za logistiko.
- Watkins, J. C. (2016). *An Introduction to the Science of Statistics: From Theory to Implementation*. Pridobljeno na spletni strani: <https://www.math.arizona.edu/~jwatkins/statbook.pdf>
- WHO - World health organization. (2021). *WHO Coronavirus Disease (COVID-19) Dashboard*. Pridobljeno na spletni strani: https://covid19.who.int/?gclid=CjwKCAiA1eKBBhBZEiwAX3gqlxLtVErKfcZX0nQOGAR8k_pTpzANn6MVmwjDhqCd_3C4Nd0Xy3xsQhoC78sQAvD_BwE

STATISTIČNA POSLOVNA ANALIZA: GRADIVA ZA SEMINARSKE IN LABORATORIJSKE VAJE

POLONA TOMINC IN MAJA ROŽMAN

Univerza v Mariboru, Ekonomsko-poslovna fakulteta, Maribor, Slovenija.

E-pošta: polona.tominc@um.si. maja.rozman@um.si

Povzetek Gradivo Statistična poslovna analiza vsebuje naslednje vsebinske sklope: opisna statistika in vzorčni pristop, preizkušanje domnev o statistično značilnih razlikah med povprečnimi vrednostmi spremenljivke v več vzorcih ter regresijska analiza. Študenti v okviru predmeta na katerega se nanaša gradivo, spoznajo uporabnost statističnih metod pri reševanju poslovnih problemov ter utrdijo in nadgradijo teoretično znanje na področju statističnih tehnik in metod, ki omogočajo spremeniti različne podatke v uporabne informacije za poslovno odločanje. Študenti usvojijo analitičen matematično statističen pristop k preučevanju poslovnih problemov, ki se sestoji iz naslednjih korakov: formulacija problema na statističen način, izbira ustrezne statistične metode, reševanje problema in interpretacija rezultatov v smislu možnih rešitev problema. V tej zbirki nalog opisujemo nekatere statistične kvantitativne metode, ki predstavljajo orodje v poslovnih raziskavah, pri čemer uporabljamo programsko orodje IBM SPSS.

Ključne besede:

opisna statistika, vzorčni pristop, preizkušanje domnev, enostavna regresija, multipla regresija

STATISTICAL BUSINESS ANALYSIS: SEMINAR AND LABORATORY TUTORIALS

POLONA TOMINC & MAJA ROŽMAN

University of Maribor, Faculty of Economics and Business, Maribor, Slovenia.

E-mail: polona.tominc@um.si. maja.rozman@um.si

Abstract The Statistical Business Analysis contains the following chapters: descriptive statistics and probability sampling, hypothesis testing about statistically significant differences between the arithmetic means of a variable in several samples and regression analysis. Within the course to which these tutorials refer, students learn about the usefulness of statistical methods in solving business problems and consolidate and upgrade theoretical knowledge in the field of statistical techniques and methods that allow to change various data into useful information for business decision making. Students acquire an analytical mathematical-statistical view to the study of business problems, which consists of the following steps: formulation of the problem in a statistical way, selection of appropriate statistical methods, problem solving and interpretation of results in terms of possible solutions. In these tutorials we describe some statistical quantitative methods that are tools in business research, using the IBM SPSS software.

Keywords:

descriptive
statistics,
sampling,
hypothesis
testing,
simple
regression,
multiple
regression







Univerza v Mariboru

Ekonomsko-poslovna fakulteta