



ZBIRKA VAJ ZA PREDMET METODE RAZISKOVANJA

POLONA TOMINC

VESNA ČANČER

MAJA ROŽMAN



Univerzitetna založba
Univerze v Mariboru



Univerza v Mariboru

Ekonomsko-poslovna fakulteta

Zbirka vaj za predmet Metode raziskovanja

Avtorice

dr. Polona Tominc

dr. Vesna Čančer

Maja Rožman

December 2018

Naslov Zbirka vaj za predmet Metode raziskovanja
Title A Collection of Exercises for the Research Methods Course

Avtorice red. prof. dr. Polona Tominc
Authors (Univerza v Mariboru, Ekonomsko-poslovna fakulteta)

red. prof. dr. Vesna Čančer
(Univerza v Mariboru, Ekonomsko-poslovna fakulteta)

asist. Maja Rožman
(Univerza v Mariboru, Ekonomsko-poslovna fakulteta)

Recenzija red. prof. dr. Majda Bastič
Review (MLC Fakulteta za management in pravo Ljubljana)

izr. prof. dr. Tomaž Kramberger
(Univerza v Mariboru, Fakulteta za logistiko)

Tehnični urednik Jan Perša, mag. inž. prom.
Technical editor (Univerzitetna založba Univerze v Mariboru)

Oblikovanje ovitka Jan Perša, mag. inž. prom.
Cover designer (Univerzitetna založba Univerze v Mariboru)

Grafične priloge Avtorice
Graphic material

Izdajatelj / Co-published by
Univerza v Mariboru, Ekonomsko-poslovna fakulteta
Razlagova ulica 14, 2000 Maribor, Slovenija
<http://epf.um.si>, epf@um.si

Založnik / Published by
Univerzitetna založba Univerze v Mariboru
Slomškov trg 15, 2000 Maribor, Slovenija
<http://press.um.si>, zalozba@um.si

Izdaja Prva izdaja
Edition

© Univerza v Mariboru,
Univerzitetna založba

Vrsta publikacije e-knjiga
Publication type

Dostopno na <http://press.um.si/index.php/ump/catalog/book/377>
Available at

Izdano Maribor, december 2018
Published

Vse pravice pridržane. Brez pisnega dovoljenja založnika je prepovedano reproduciranje, distribuiranje, predelava ali druga uporaba tega dela ali njegovih delov v kakršnemkoli obsegu ali postopku, vključno s fotokopiranjem, tiskanjem ali shranjevanjem v elektronski obliki.

CIP - Kataložni zapis o publikaciji
Univerzitetna knjižnica Maribor

004.42:311.1(076)

TOMINC, Polona

Zbirka vaj za predmet Metode raziskovanja / avtorice Polona Tominc, Vesna Čančer, Maja Rožman. - 1. izd. - El. knjiga. - V Mariboru : Univerzitetna založba Univerze, 2018

Način dostopa (URL): <http://press.um.si/index.php/ump/catalog/book/377>. - Nasl. v kolofonu: A Collection of Exercises for the Research Methods Course

ISBN 978-961-286-223-7

doi: 10.18690/978-961-286-223-7

1. Dr. vzp. stv. nasl. 2. Čančer, Vesna 3. Rožman, Maja, 1988-

COBISS.SI-ID [95769601](https://nbn-resolving.org/urn:nbn:si:coibiss:SI-ID-95769601)

ISBN 978-961-286-223-7 (PDF)

DOI <https://doi.org/10.18690/978-961-286-223-7>

Cena Brezplačen izvod

Odgovorna oseba založnika red. prof. dr. Zdravko Kacič, rektor Univerze v Mariboru

Zbirka vaj za predmet Metode raziskovanja

POLONA TOMINC, VESNA ČANČER IN MAJA ROŽMAN

Povzetek Zbirka vaj je namenjena predvsem uporabi statističnega programskega paketa SPSS za izbrane vsebinske sklope, ki sodijo v okvir kvantitativnega dela predmeta Metode raziskovanja, na magistrskem študijskem programu Ekonomske in poslovne vede, na UM EPF. Gradivo je dopolnitev k osnovni študijski literaturi pri kvantitativnem delu predmetu in k sklopom, obravnavanim v okviru predavanj. Gradivo je namenjeno študentom pri študiju in razumevanju klasifikacije kvantitativnih metod, glede na značilnosti zbranih podatkov in glede na postavljene raziskovalne domneve. Vsebinska analiza je prikazana z namenom pravilne razlage dobljenih rezultatov, pridobljenih s programom SPSS. Sestavni del aktivnega dela študentov na vajah pa je interpretacija dobljenih rezultatov za podporo poslovnemu odločanju.

Študentom želimo uspešno uporabo obravnavanih metod.

Ključne besede: • Statistične kvantitativne metode • deskriptivna statistika • univariatne metode • multivariatne metode • SPSS •

NASLOVI AVTORIC: dr. Polona Tominc, redna profesorica, Univerza v Mariboru, Ekonomsko-poslovna fakulteta, Razlagova ulica 14, 2000 Maribor, Slovenija, e-pošta: polona.tominc@um.si. dr. Vesna Čančer, redna profesorica, Univerza v Mariboru, Ekonomsko-poslovna fakulteta, Razlagova ulica 14, 2000 Maribor, Slovenija, e-pošta: vesna.cancer@um.si. Maja Rožman, asistentka, Univerza v Mariboru, Ekonomsko-poslovna fakulteta, Razlagova ulica 14, 2000 Maribor, Slovenija, e-pošta: maja.rozman1@um.si.

Kazalo

1	DESKRIPTIVNA STATISTIKA IN VZORČNI PRISTOP	1
1.1	Statistična enota, statistična množica, slučajni vzorec	1
1.2	Statistične spremenljivke	2
1.3	Osnovni statistični parametri in statistike	3
1.4	Analiza podatkov	5
1.4.1	Vnos podatkov v program SPSS	5
1.4.2	Transformacija podatkov	8
1.4.3	Uvoz podatkov iz Excela v program SPSS	10
1.4.4	Vzorčenje in intervalno ocenjevanje parametrov	13
2	NORMALNA PORAZDELITEV	21
2.1	Opredeleitev in značilnosti normalne porazdelitve	21
2.2	Kolmogorov-Smirnov test in Shapiro-Wilk W test	22
3	PARAMETRIČNI IN NEPARAMETRIČNI UNIVARIATNI STATISTIČNI TESTI	29
3.1	Parametrični test za odvisna vzorca: t-test za odvisna vzorca	30
3.2	Parametrični test za neodvisna vzorca: t-test za neodvisna vzorca	34
3.3	Parametrični test za neodvisne vzorce: ANOVA	38
3.4	Neparametrični test: χ^2 –test za analizo povezanosti dveh nominalnih spremenljivk	43
3.5	Neparametrični test za odvisna vzorca: Wilcoxon signed ranks test	49
3.6	Neparametrični test za neodvisna vzorca: Mann-Whitney U test	53
4	FAKTORSKA ANALIZA	59
5	REGRESIJSKA ANALIZA	71
5.1	Enostavna linearna regresija	71
5.2	Multipla regresijska analiza	75
6	DISKRIMINANTNA ANALIZA	81
	LITERATURA IN VIRI	87

1 DESKRIPTIVNA STATISTIKA IN VZORČNI PRISTOP

Deskriptivna statistika ali opisna statistika je vrsta analize, katera kvantitativno opisuje ali povzema značilnosti zbranih podatkov. V deskriptivno statistiko sodijo metode, ki se nanašajo na prikazovanje in analiziranje značilnosti danega niza podatkov. O opisnih statističnih analizah govorimo takrat, kadar predstavljamo npr. velikost vzorca (n), ponazarjamo strukturo zbranih podatkov (strukturni %), pogostost pojavljanja posameznih vrednosti spremenljivke - frekvence (f), minimalne (Min) in maksimalne (Max) vrednosti, srednje vrednosti, razpršenost podatkov okoli povprečja in drugo (Tominc & Kramberger, 2007; Goos & Meintrup, 2015).

Na drugi strani inferenčna ali matematična statistika uporablja vzorčni pristop in je osredotočena na testiranje domnev o statističnih parametrih statistične množice na osnovi zbranih podatkov slučajnega vzorca (Tabachnick & Fidell, 2013).

1.1 Statistična enota, statistična množica, slučajni vzorec

Statistična enota je vsak posamezen element statistične množice, ki je predmet opazovanja (npr. podjetje, gospodinjstvo, delavec, študent ipd.), v danem časovnem trenutku oziroma intervalu. Statistična enota je pojav, ki je opredeljen s stvarnega, krajevnega in časovnega vidika (Artenjak, 2003).

Statistična množica je množica statističnih enot, ki izpolnjujejo določene opredeljujoče kriterije oziroma lastnosti. Statistično množico imenujemo tudi populacija.

Vzorec je del celotne populacije, na osnovi katerega izvedemo sklepanje o celotni populaciji. Temeljno načelo pri tem je, da mora biti vzorec slučajen. Za *slučajni 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 & Kramberger, 2007).

1.2 Statistične spremenljivke

Statistična spremenljivka opisuje lastnost statistične enote (Artenjak, 2003).

Vrste spremenljivk (ibid):

Opisne (atributne) spremenljivke so tiste spremenljivke, katerih vrednosti lahko opišemo le z besedami (npr. spol: moški, ženski).

Številске (numerične) spremenljivke so tiste spremenljivke, katerih vrednosti lahko izrazimo s števili (npr. starost, dobiček podjetja). Med njimi ločimo:

- zvezne spremenljivke, ki so številске spremenljivke, ki lahko zavzamejo katerokoli vrednost na intervalu (npr. dolžina, čas, teža) in
- diskretne spremenljivke (ali nezvezne spremenljivke), ki so številске spremenljivke, ki lahko zavzamejo le določene končne, najpogosteje celoštevilčne, vrednosti (npr. število članov v gospodinjstvu).

Vrste spremenljivk glede na tip merjenja (merske lestvice) (Bastič, 2006):

Opisne spremenljivke merimo na nominalni in ordinalni merski lestvici, medtem ko številске spremenljivke merimo na intervalni in razmernostni lestvici.

1. Nominalne spremenljivke merimo na nominalni merski lestvici, katera omogoča razvrščanje enot po določeni skupni značilnosti. Statistične enote so razvrščene v skupine tako, da imajo enote, razvrščene v isto skupino, isto značilnost (npr. spol: 1-moški, 2-ženski; odgovor: 1-da, 2-ne).

2. Ordinalne spremenljivke merimo na ordinalni merski lestvici, katera omogoča, da so skupine razvrščene po določenem kriteriju, kar pomeni, da lahko vrednosti uredimo od najmanjše do največje (npr. stopnja izobrazbe, uspeh, velikost podjetja: 1-malo podjetje, 2-srednje veliko, 3-veliko podjetje).

3. Intervalne spremenljivke merimo na intervalni merski lestvici, za katero je značilno, da uporablja enoto mere. Med svojo začetno in končno točko je razčlenjena na enako velike intervale (npr. temperatura v °C).

4. Razmernostne spremenljivke merimo na razmernostni merski lestvici, za katero je značilno, da je njena začetna točka 0 in se ne spreminja. Navedena merska lestvica je absolutna in se razlika vedno meri od točke nič (npr. število turistov, starost, prihodek).

1.3 Osnovni statistični parametri in statistike

Parameter je številska ali opisna vrednost, ki opisuje neko značilnost statistične množice ali populacije (Artenjak, 2003).

Statistika je številska ali opisna vrednost, ki ocenjuje neko značilnost statistične množice in jo pridobimo iz vzorca (ibid).

Mere centralne tendence so vrednosti, ki jih največkrat uporabljamo pri osnovni opisni statistični analizi. Mere centralne tendence so predstavniki vseh opazovanih vrednosti. Med najpomembnejše ***mere centralne tendence*** uvrščamo aritmetično sredino, mediano in modus:

Aritmetična sredina (ang. *Mean*) je srednja vrednost, ki jo dobimo tako, da vsoto vseh vrednosti spremenljivke delimo s številom enot v zbranih podatkih (n).

Mediana (ang. *Median*) je srednja vrednost, od katere ima polovica enot manjše ali enake vrednosti, polovica pa večje. Označili jo bomo z *Me*.

Modus (ang. *Mode*) ali gostiščnica je enaka tisti vrednosti spremenljivke, ki se najpogosteje pojavlja. Označili ga bomo z *Mo* (ibid).

Med najpogostejše ***mere variabilnosti*** uvrščamo variacijski razmik, varianco in standardni odklon:

Variacijski razmik (ang. *Range*) je razlika med največjo in najmanjšo vrednostjo spremenljivke.

Varianca (ang. *Variance*) meri odstopanja posameznih vrednosti spremenljivke od aritmetične sredine. Definirana je kot povprečje kvadratov odklonov posameznih vrednosti od aritmetične sredine.

Standardni odklon (ang. *Standard deviation*) je definiran kot kvadratni koren iz variance. S standardnim odklonom lahko izmerimo, kako so razpršene vrednosti okoli aritmetične sredine zbranih podatkov; izraža pa se v enakih merskih enotah kot opazovana spremenljivka. Če opazujemo več skupin enot po isti spremenljivki, velja, da višja kot je vrednost standardnega odklona, bolj so enote v vzorcu razpršene in obratno, nižja vrednost kaže na manjšo razpršenost enot in večjo koncentracijo enot okoli vrednosti aritmetične sredine (ibid).

Asimetričnost in sploščenost

Asimetričnost in sploščenost merimo z merami asimetrije in merami sploščenosti.

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. Bolj kot se koeficient

asimetrije razlikuje od vrednosti 0, večja je jakost asimetrije, koeficient asimetrije pa pri večini empiričnih porazdelitev lahko zavzame vrednost med -3 in +3 (Artenjak, 2003).

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* (ima daljša repa in ožji osrednji del). 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 tako koeficienta asimetričnosti in sploščenosti enaka 0. Vrednosti obeh koeficientov (med drugim) pokažeta, kako se proučevana porazdelitev vrednosti spremenljivke razlikuje od normalne porazdelitve (Bastič, 2006).

1.4 Analiza podatkov

1.4.1 Vnos podatkov v program SPSS

Naloga 1.

Raziskovalci so želeli analizirati stopnjo izvoza v letu 2017, za deset srednje velikih podjetij v Sloveniji. V tabeli 1 so podatki o stopnji izvoza v letu 2017, za deset srednje velikih podjetij (v % od celotnih prihodkov).

Tabela 1: Podatki o stopnji izvoza v letu 2017, 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	131	12,3	10,6	9,0

- Opreделите statistično enoto, statistično spremenljivko in moč vzorca.
- Podatke o stopnji izvoza v letu 2017, za teh deset srednje velikih podjetij vnesite v program SPSS. Izračunajte in pojasnite rezultate opisne statistike za stopnjo izvoza v letu 2017 teh desetih podjetij.
- Na osnovi mere asimetričnosti in mere sploščenosti pojasnite ali so vrednosti navedene spremenljivke normalno porazdeljene.
- Za spremenljivko *stopnja izvoza v letu 2017*, za deset srednje velikih podjetij prikažite frekvenčno preglednico.
- Narišite še frekvenčni histogram z normalno krivuljo.

Postopek vnosa podatkov: Podatke o stopnji izvoza v letu 2017 desetih srednje velikih podjetij vnesemo v SPSS tako, da spodaj v levem kotu kliknemo desno okence *Variable View*. V okence *Name* vpišemo ime spremenljivke *stopnja_izvoza*, v okencu *Type* določimo vrsto spremenljivke (numeric), v okencu *Decimals* določimo število decimalnih mest, v okencu *Label* zapišemo celotno ime spremenljivke (*stopnja izvoza v letu 2017 desetih srednje velikih podjetij*). Nato kliknemo spodnje levo okence *Data View* in zapišemo vrednosti spremenljivk v stolpec.

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

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

Odgovori in izpisi rezultatov:

a)

Statistična enota: 1 srednje veliko podjetje.

Statistična spremenljivka: *Stopnja izvoza* (numerična, zvezna spremenljivka).

b in c)

Tabela 2: Opisna statistika za spremenljivko stopnja izvoza v letu 2017, desetih srednje velikih podjetij

N	Valid	10
	Missing	0
Mean		10,940
Std. Error of Mean		0,6837
Median		11,250
Mode		13,1
Std. Deviation		2,1619
Variance		4,674
Skewness		- 0,752
Std. Error of Skewness		0,687
Kurtosis		- 0,091
Std. Error of Kurtosis		1,334
Range		6,6
Minimum		6,7
Maximum		13,3

V tabeli 2 so prikazani rezultati opisne statistike za spremenljivko *stopnja izvoza v letu 2017* desetih srednje velikih podjetij. Vidimo, da je bilo v vzorec vključenih 10 srednje velikih podjetij (n) in ni nobene manjkajoče vrednosti (Missing = 0). Povprečna vrednost (Mean) stopnje izvoza v letu 2017 desetih podjetij znaša 10,940 (% od prihodka). Standardna napaka ocene aritmetične sredine (ang. Standard error of mean) znaša 0,684 (% od prihodka) in pomeni, da manjša kot je njena vrednost, manjša je variabilnost med vzorčnimi povprečnimi vrednostmi in boljši predstavnik statistične množice je vzorec. Mediana (Median) znaša 11,250 (% od prihodka), kar pomeni, da ima polovica (50 %) podjetij stopnjo izvoza manjšo ali enako od 11,250, ter polovica (50 %) podjetij več kot 11,250 (% od prihodka) stopnjo izvoza. Modus (Mode) je 13,1 (% od prihodka) in predstavlja tisto vrednost spremenljivke *stopnja izvoza v letu 2017* desetih srednje velikih podjetij, ki se najpogosteje pojavlja (ta vrednost se pojavlja dvakrat, vse ostale po enkrat). Standardni odklon je kvadratni koren iz variance in v našem primeru znaša 2,162 (% od prihodka), ki opisuje razpršenost vrednosti spremenljivke okoli vrednosti aritmetične sredine. Varianca znaša 4,674 (% od prihodka)². Koeficient asimetrije (Skewness) znaša -0,752, kar pomeni, da je porazdelitev asimetrična v levo (negativna asimetrična porazdelitev). Koeficient sploščenosti (Kurtosis) znaša -0,091, kar kaže na sploščeno porazdelitev (negativna vrednost). Variacijski razmik (Range) znaša 6,6 (%-nih točk) in je enak razliki med največjo (Maximum = 13,3 (% od prihodka)) in najmanjšo (Minimum = 6,7 (% od prihodka)) vrednostjo spremenljivke.

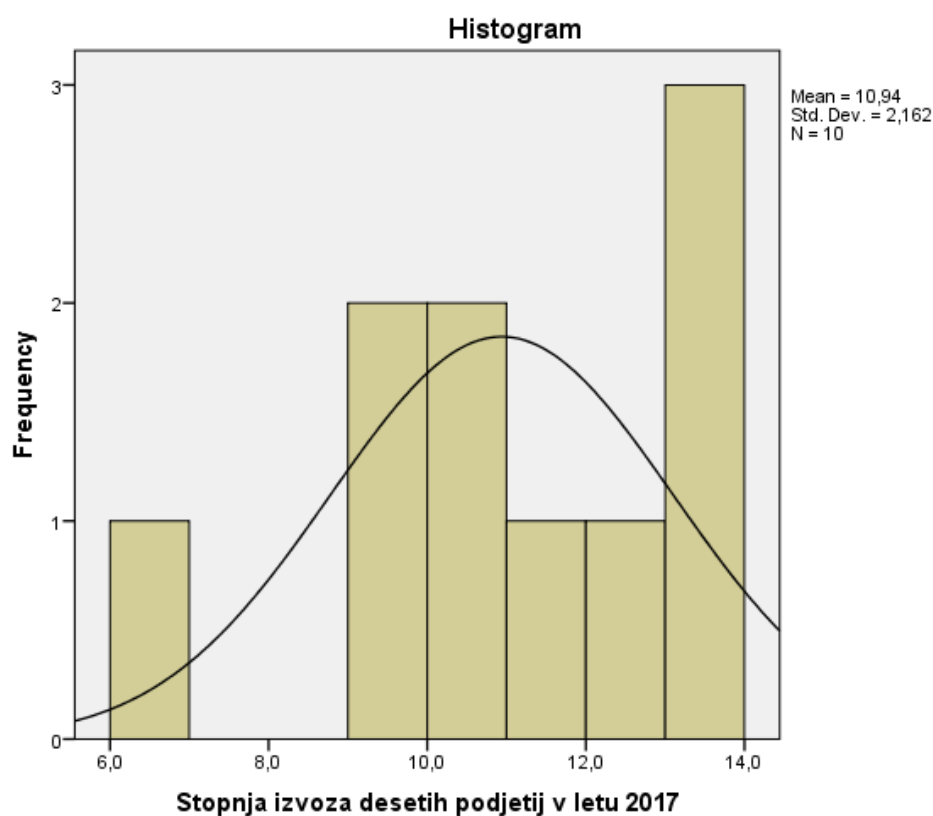
d)

Tabela 3: Frekvenčna preglednica - Stopnja izvoza v letu 2017, desetih srednje velikih podjetij

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	6,7	1	10,0	10,0	10,0
	9,0	1	10,0	10,0	20,0
	9,1	1	10,0	10,0	30,0
	10,6	1	10,0	10,0	40,0
	10,7	1	10,0	10,0	50,0
	11,8	1	10,0	10,0	60,0
	12,0	1	10,0	10,0	70,0
	13,1	2	20,0	20,0	90,0
	13,3	1	10,0	10,0	100,0
	Total	10	100,0	100,0	

Tabela 3 kaže, da je imelo eno srednje veliko podjetje od desetih srednje velikih podjetij stopnjo izvoza 6,7 (% od prihodka). Prav tako je imelo eno srednje veliko podjetje od desetih srednje velikih podjetij stopnjo izvoza 9,0 (% od prihodka) in enako velja za ostale primere podjetij; le pri dveh srednje velikih podjetij od desetih srednje velikih podjetij vidimo, da sta imeli stopnji izvoza enaki, 13,1 (% od prihodka).

e)



Histogram 1: Frekvenčni histogram s krivuljo prilagojene normalne porazdelitve

1.4.2 Transformacija podatkov

Pretvorbo podatkov je mogoče doseči z uporabo transformacije. V primeru, kadar med vrednostmi proučevane spremenljivke nastopajo osamelci (tiste vrednosti spremenljivke, ki se bistveno razlikuje od ostalih vrednosti, *angl. outlier*) ali porazdelitev vrednosti spremenljivke ni normalna, si lahko pomagamo s transformacijo podatkov. Najpogosteje uporabljene transformacije so logaritmiranje podatkov, korenjenje vrednosti, recipročna transformacija ipd (Bastič, 2006).

Naloga 2.

Želimo zmanjšati negativno asimetrijo pri *stopnji izvoza v letu 2017* desetih podjetij iz naloge 1. V našem primeru bomo uporabili logaritmiranje stopenj izvoza.

Postopek: Kliknemo *Transform* in nato *Compute Variable*. V okencu *Target Variable* poimenujemo novo spremenljivko, ki jo bomo z logaritmiranjem oblikovali (npr. *log_izvoz*). V okencu *Type & Label* zapišemo opis novo poimenovane spremenljivke (npr. *log stopnja izvoza*). V desnem okencu *Function group* kliknemo na ustrezno transformacijo in sicer *Arithmetic*. V okencu *Functions and Special variables* kliknemo *Lg10*. Kliknemo na gumb s puščico navzgor. V okencu *Numeric Expression*, se izpiše *Lg(?)*, kjer namesto vprašaja prenesemo ime spremenljivke (spremenljivka je napisana v okencu na levi strani), ki jo želimo logaritmirati ter kliknemo *OK*.

V okencu *Data View* se izpiše nova spremenljivka (*log_izvoz*) z novimi vrednostmi, ki so prikazane v tabeli 3.

Tabela 4: Nova spremenljivka (*log_izvoz*) z novimi vrednostmi

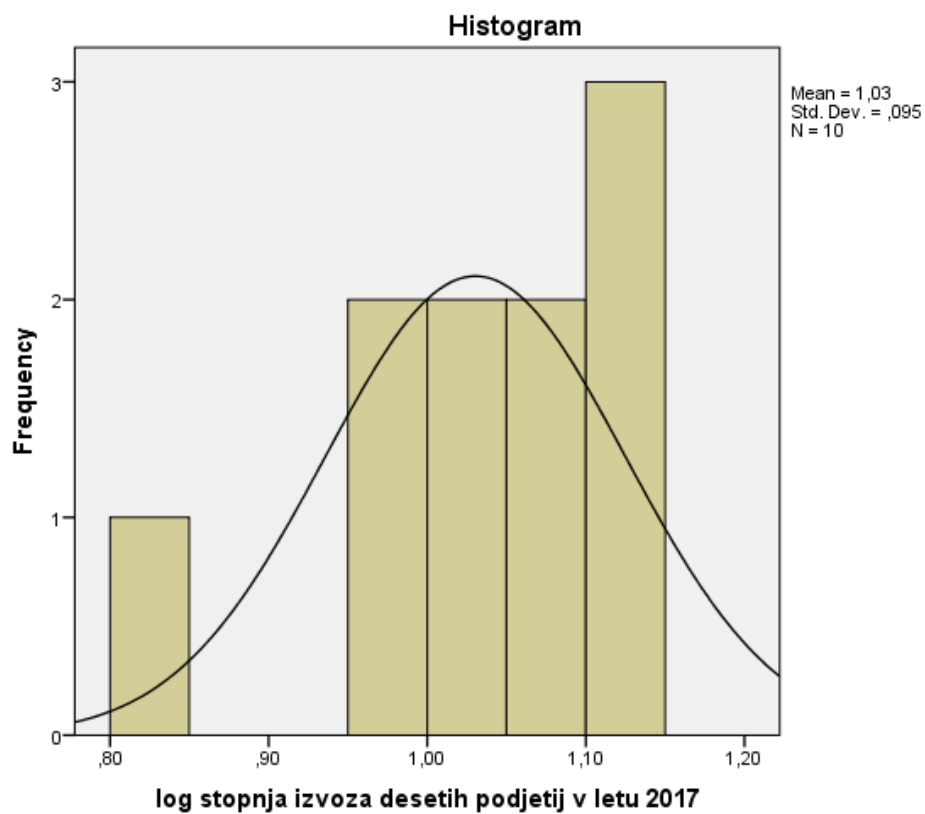
Podjetje	1	2	3	4	5	6	7	8	9	10
Stopnja izvoza v letu 2017	6,7	9,1	10,7	11,8	12,0	13,1	13,1	12,3	10,6	9,0
Nova spremenljivka: <i>log_izvoz</i>	0,83	0,96	1,03	1,07	1,08	1,12	1,12	1,12	1,03	0,95

Izračunajte in pojasnite rezultate opisne statistike za novo spremenljivko. Narišite še frekvenčni histogram z normalno porazdelitvijo in komentirajte rezultat.

Odgovori in izpisi rezultatov:

Tabela 5: Opisna statistika za spremenljivko log_izvoz

N	Valid	10
	Missing	0
Mean		1,0304
Std. Error of Mean		0,02992
Median		1,0506
Mode		1,12
Std. Deviation		0,09463
Variance		0,009
Skewness		-1,149
Std. Error of Skewness		0,687
Kurtosis		1,104
Std. Error of Kurtosis		1,334
Range		0,30
Minimum		0,83
Maximum		1,12



Histogram 2: Frekvenčni histogram s krivuljo prilagojene normalne porazdelitve (log_izvoz)

1.4.3 Uvoz podatkov iz Excela v program SPSS

Naloga 3.

V datoteki *Deskriptivna statistika in vzorčni pristop_Excel* so podani podatki o številu zaposlenih v podjetjih, ki so razvrščena po standardni klasifikaciji dejavnosti v letu 2017 v devetih geografskih regijah. Prenesite datoteko *Deskriptivna statistika_Excel* v program SPSS.

Postopek uvoza podatkov iz Excela v program SPSS: Odpremo program SPSS in kliknemo *File*, ter izberemo *Open* in nato še *Data*. Odpre se okence *Open data* v katerem poiščemo Excelovo datoteko (v okencu *Files of type* še označimo *All files*) in na koncu kliknemo *Open*.

- a) Izračunajte in pojasnite rezultate opisne statistike za spremenljivko *število zaposlenih v predelovalni dejavnosti*.
- b) Izračunajte in primerjajte rezultate opisne statistike za naslednje tri spremenljivke: *število zaposlenih v gostinstvu*, *število zaposlenih v dejavnosti poslovanje z nepremičninami* ter *število zaposlenih v drugih raznovrstnih poslovnih dejavnostih*. Vsebinsko pojasnite povprečno vrednost in standardno napako ocene aritmetične sredine vseh treh spremenljivk.
- c) Izračunajte koeficient asimetrije in koeficient sploščenosti ter narišite frekvenčni histogram z normalno krivuljo za spremenljivko *število zaposlenih v informacijski in komunikacijski dejavnosti*.
- d) Za spremenljivko *število zaposlenih v rudarstvu* narišite stolpični diagram.

Odgovori in izpisi rezultatov:

a)

Tabela 6: Opisna statistika - število zaposlenih v predelovalni dejavnosti

N	Valid	9
	Missing	0
Mean		18309,78
Std. Error of Mean		288,786
Median		17690,00
Mode		17419 ^a
Std. Deviation		866,358
Variance		750575,694
Skewness		0,454
Std. Error of Skewness		0,717
Kurtosis		-1,979
Std. Error of Kurtosis		1,400
Range		2137
Minimum		17419
Maximum		19556

a. Multiple modes exist. The smallest value is shown

b)

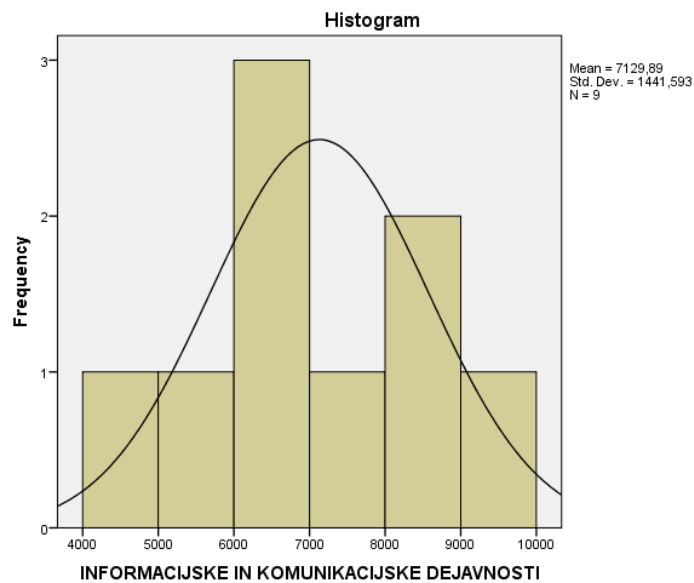
Tabela 7: Opisna statistika - število zaposlenih v dejavnostih gostinstvo, poslovanje z nepremičninami in druge raznovrstne poslovne dejavnosti

		Gostinstvo	Poslovanje z nepremičninami	Druge raznovrstne poslovne dejavnosti
N	Valid	9	9	9
	Missing	0	0	0
Mean		9761,00	2428,33	5793,67
Std. Error of Mean		474,678	99,051	430,134
Median		9093,00	2414,00	5625,00
Std. Deviation		1424,033	297,152	1290,403
Variance		2027871,000	88299,250	1665139,250
Skewness		0,169	0,593	-0,008
Std. Error of Skewness		0,717	0,717	0,717
Kurtosis		-2,074	-0,235	-1,221
Std. Error of Kurtosis		1,400	1,400	1,400
Range		3616	926	3767
Minimum		7965	2049	3823
Maximum		11581	2975	7590

c)

Tabela 8: Koeficient asimetrije in koeficient sploščenosti – število zaposlenih v informacijski in komunikacijski dejavnosti

	N	Skewness		Kurtosis	
	Statistic	Statistic	Std. Error	Statistic	Std. Error
Informacijske in komunikacijske dejavnosti	9	0,062	0,717	-0,899	1,400
Valid N (listwise)	9				



Histogram 3: Frekvenčni histogram z normalno krivuljo – število zaposlenih v informacijski in komunikacijski dejavnosti

d)

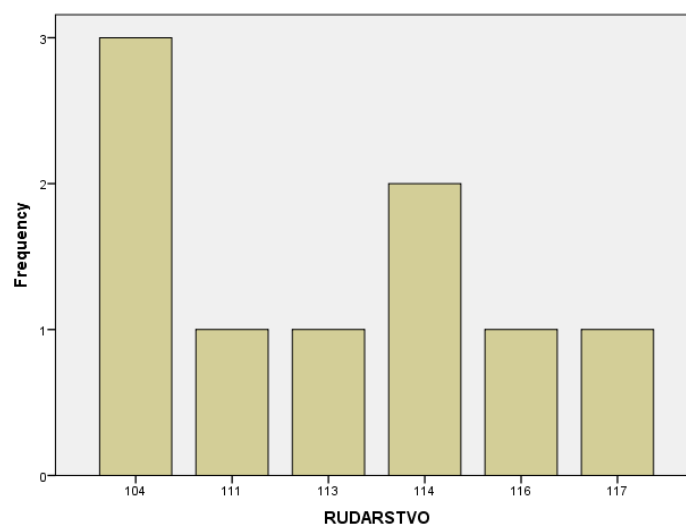


Diagram 1: Frekvenčni histogram – število zaposlenih v rudarstvu.

1.4.4 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 & 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).

Intervalno ocenjevanje aritmetične sredine

Vzorčna ocena aritmetične sredine statistične množice, je točkovna ocena. To je ocena vrednosti statističnega parametra, podana z eno samo vrednostjo, ki je izračunana iz slučajnega vzorca. Poleg točkovne ocene pa ocenjujemo vrednost statističnega parametra tudi z intervalno oceno, ki temelji na točkovni vzorčni oceni statističnega parametra (rezultat iz vzorca posplošimo na statistično množico). Na podlagi vzorčnih podatkov določimo interval, v katerem je z določeno stopnjo verjetnosti mogoče pričakovati, da se nahaja vrednost obravnavanega statističnega parametra. To verjetnost imenujemo stopnja zaupanja (običajno 90 %, 95 % ali 99 %), interval pa *interval zaupanja*. Verjetnost 100 % — stopnja zaupanja, imenujemo *stopnja tveganja* %. Če določimo meji intervala za vrednost statističnega parametra iz vzorčnih podatkov tako, da določimo zgornjo in spodnjo mejo intervala, ob dani stopnji zaupanja, govorimo o dvostranskem ocenjevanju statističnega parametra. Če postavimo samo eno od obeh mej, bodisi zgornjo ali spodnjo, ob dani stopnji zaupanja, govorimo o enostranskem ocenjevanju statističnega parametra (Artenjak, 2003).

Naloga 4.

V datoteki *Deskriptivna statistika in vzorčni pristop_Naloga4.sav* so podatki o podjetju, ki opravlja storitve na tržišču (Bastič in drugi, 2008). Podatki prikazujejo podatke slučajnega vzorca za vrednost opravljenih storitev v določenem letu, število sklenjenih pogodb, podatke o tem, s katerimi oddelki v teh partnerskih podjetjih anketirana poslujejo in kolikšna je donosnost v podjetjih, s katerimi poslujejo.

a) Izračunajte in pojasnite 95 %-ni in 80 %-ni interval zaupanja za spremenljivko *vrednost storitev*.

Potek: Kliknemo na *Analyze*, nato na *Descriptive Statistic* ter izberemo funkcijo *Explore*. Iz levega okenca prenesemo spremenljivko *vrednost storitev* v desno okence *Dependent List* in kliknemo na okvirček *Statistics* ter obkljukamo *Descriptives* – pri *Confidence Interval for Mean* napišemo 95 %, ker želimo pojasniti 95 %-ni interval zaupanja za spremenljivko *vrednost storitev* (v primeru, če želimo pojasniti 80 %-ni interval zaupanja za spremenljivko napišemo 80 %).

b) Vsa podjetja razdelite na dve skupini glede na vrednost mediane za spremenljivko *donosnost*.

Postopek:

V prvem koraku je potrebno ugotoviti vrednost mediane za spremenljivko *donosnost* - koraki si sledijo po naslednjem vrstnem redu: *Analyze*, nato *Descriptive statistic* in *Explore*.

Vsa podjetja razdelite na dve skupini glede na vrednost mediane za spremenljivko *donosnost*: Kliknemo na *Transform* in izberemo ukaz *Recode into Different Variables* (kar pomeni, da ohranimo spremenljivko *donosnost* in tvorimo novo spremenljivko, z novim imenom, ki bo shranjena v datoteki s podatki). V desno okence *Numeric Variable* prenesemo spremenljivko, ki jo želimo preoblikovati – *donosnost*. V okencu *Output Variable* pod vrstico *Name*, vpišemo ime nove spremenljivke (npr. *donosnost1*), kliknemo na gumbek *Change* in nato na gumbek *Old and New Values*. Odpre se okno, kjer imamo na levi strani stare vrednosti (*Old Value*), na desni strani pa nove vrednosti (*New Value*). Vse vrednosti spremenljivke, ki so manjše ali enake od vrednosti 12 pri spremenljivki *donosnost* morajo biti enake 0 pri novi spremenljivki *donosnost1*. Vse vrednosti spremenljivke, ki so večje od vrednosti 12 pri spremenljivki *donosnost* morajo biti enake 1 pri novi spremenljivki *donosnost1*. Zato izberemo na levi strani (*Old Value*): *Range, lowest through value*: vtipkamo 12 in na desni strani (*New Value*) izberemo 0 in kliknemo *Add*. Nato ponovimo postopek za vse vrednosti, ki so večje od 12 in sicer: Na levi strani (*Old Value*): *Range, value through highest* vtipkamo 12 in na desni strani (*New Value*) izberemo 1 in kliknemo *Add* ter *Continue*, *OK*.

c) Uvedite novo spremenljivko in jo poimenujte s *Test*, ki bo za posamezno podjetje, s katerim organizacija sodeluje, pojasnjevala povprečno vrednost *storitev na pogodbo*. Izračunajte deskriptivno statistiko za tako dobljeno spremenljivko.

Postopek uvedbe oziroma oblikovanje nove spremenljivke: Kliknemo na *Transform* in izberemo ukaz *Compute Variable*. Odpre se okno, kjer na levi strani zgoraj pod *Target Variable* vpišemo ime nove spremenljivke, ki jo želimo oblikovati, na primer *Test*. V okence *Numeric expression* vpišemo računski izraz za izračun nove spremenljivke (v našem primeru vrednost storitev delimo s številom pogodb in tako dobimo novo spremenljivko – *Test*) in kliknemo *OK*. V datoteki s podatki *Data View* lahko v

zadnji koloni najdemo zapisane vrednosti novo oblikovane spremenljivke, ki smo jo imenovali *Test*. Tudi pri *Variable View* lahko na koncu opazimo zapisano novo spremenljivko *Test*.

d) Z uporabo generatorja slučajnih števil izberite cca 35 statističnih enot v slučajni vzorec. Izračunajte deskriptivno statistiko za spremenljivko *vrednost storitev* v tako dobljenem vzorcu.

Postopek uporabe generatorja slučajnih števil: Kliknemo na *Data* in izberemo ukaz *Select Cases*. Kliknemo na gumb *Random sample of cases* in kliknemo okence *Sample* ter napišemo želeno vrednost pod *Approximately ___ % of all cases* (v našem primeru smo napisali 50 %). Kliknemo na *Continue* in *OK*.

Odgovori in izpisi rezultatov:

a)

Tabela 9: 95 %-ni interval zaupanja za spremenljivko vrednost storitev

		Statistic	Std. Error	
Vrednost storitev	Mean	55,45	12,423	
	95 % Confidence Interval for Mean	Lower Bound	30,67	
		Upper Bound	80,23	
	5 % Trimmed Mean	38,47		
	Median	24,00		
	Variance	10803,400		
	Std. Deviation	103,939		
	Minimum	6		
	Maximum	777		
	Range	771		
	Interquartile Range	41		
	Skewness	5,325	0,287	
	Kurtosis	34,292	0,566	

V tabeli 9 vidimo, da spodnja meja (Lower Bound) intervala zaupanja znaša 30,67 in zgornja meja (Upper Bound) znaša 80,23, kar pomeni, da s 95 % verjetnostjo ocenjujemo, da je povprečna vrednost storitev v statistični množici med 30,67 in 80,23 enotami.

Tabela 10: 80 %-ni interval zaupanja za spremenljivko vrednost storitev

		Statistic	Std. Error	
Vrednost storitev	Mean	55,45	12,423	
	80 % Confidence Interval for Mean	Lower Bound	39,38	
		Upper Bound	71,52	
	5 % Trimmed Mean	38,47		
	Median	24,00		
	Variance	10803,400		
	Std. Deviation	103,939		
	Minimum	6		
	Maximum	777		
	Range	771		
	Interquartile Range	41		
	Skewness	5,325	0,287	
Kurtosis	34,292	0,566		

Z 80 % verjetnostjo ocenjujemo, da je povprečna vrednost storitev v statistični množici med 39,38 in 71,52 enot.

b)

Tabela 11: Izračun vrednosti mediane za spremenljivko donosnost

		Statistic	Std. Error	
Donosnost	Mean	12,40	0,492	
	80% Confidence Interval for Mean	Lower Bound	11,76	
		Upper Bound	13,04	
	5% Trimmed Mean	12,52		
	Median	12,00		
	Variance	16,910		
	Std. Deviation	4,112		
	Minimum	-3		
	Maximum	22		
	Range	25		
	Interquartile Range	5		
	Skewness	-0,657	0,287	
Kurtosis	1,897	0,566		

V tabeli 11 vidimo, da mediana za spremenljivko *donosnost* (vzorčni podatki) znaša 12,00 (v %).

c)

Tabela 12: Opisna statistika za novo spremenljivko Test

N	Valid	70
	Missing	0
Mean		12,9383
Std. Error of Mean		4,35546
Median		3,4186
Mode		48,50
Std. Deviation		36,44037
Variance		1327,900
Range		295,71
Minimum		0,29
Maximum		296,00

Naloga 5.

V tabeli 13 so podani podatki za $n = 15$ anketiranih oseb, ki so odgovorile na vprašanje: Ali bi vas strah pred neuspehom odvrnil od tega, da bi ustanovili svoje podjetje? Anketirani so odgovarjali na vprašanje z možnimi odgovori: 0 – Ne in 1 – Da.

Spol anketiranega je označen z 1 – Moški in 2 – Ženski spol. Starost je merjena v dopolnjenih letih.

Tabela 13: Podatki o odgovorih anketiranih oseb

Spol	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
Starost	45	45	58	62	64	55	57	20	23	20	21	33	25	33	30
Strah pred neuspehom	0	0	0	0	0	1	1	0	0	1	1	1	1	1	1

a) Vnesite podatke v program SPSS. Razdelite datoteko na dva dela, glede na spol anketiranega in opravite ločeno analizo za moške in ženske o odstotku tistih, ki bi jih strah pred neuspehom odvrnil od ustanovitve novega podjetja.

Postopek razdelitve datoteke na dva dela (glede na spol): Kliknemo *Data*, nato *Split File* (ukaz bo datoteko razdelil na 2 dela in sicer glede na spol), kliknemo na *Organize output by groups* in prenesemo spremenljivko *spol* v desni okvirček (nad katerim piše *Groups Based on..*) ter kliknemo *OK*.

Nato opravimo ločeno analizo za moške in ženske: kliknemo *Analyze*, nato izberemo ukaz *Descriptive Statistic* in kliknemo *Descriptives* (lahko tudi *Frequencies* ali *Explore*).

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

Postopek oblikovanja nove spremenljivke glede starosti: Kliknemo na *Transform* in izberemo ukaz *Compute Variable*. Odpre se okno, kjer na levi strani zgoraj pod *Target Variable* vpišemo ime nove spremenljivke, ki jo želimo oblikovati, na primer *Starost1*. V okence *Numeric expression* vpišemo računski izraz za izračun nove spremenljivke (v našem primeru vpišemo: $\text{starost} - 40$ in tako dobimo novo spremenljivko – *Starost1*) in kliknemo *OK*.

V datoteki s podatki *Data View* lahko v zadnji koloni najdemo vrednosti novo oblikovane spremenljivke, ki smo jo imenovali *Starost1*. Tudi pri *Variable View* lahko na koncu opazimo zapisano novo spremenljivko *Starost1*.

Postopek pretvorbe 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: Kliknemo na *Transform* in izberemo ukaz *Recode into Different Variables*. V desno okence *Numeric Variable* prenesemo spremenljivko, ki jo želimo preoblikovati – *Starost1*. V okencu *Output Variable* pod vrstico *Name*, vpišemo ime nove spremenljivke (npr. *Starost2*), kliknemo na gumbek *Change* in nato na gumbek *Old and New Values*. Odpre se okno,

kjer imamo na levi strani stare vrednosti (*Old Value*), na desni strani pa nove vrednosti (*New Value*). Vse vrednosti spremenljivke, ki so manjše ali enake 0 pri spremenljivki *Starost1* morajo biti enake 0 pri novi spremenljivki *Starost2*. Zato izberemo na levi strani (*Old Value*): Range, lowest through value: vtipkamo 0 in na desni strani (*New Value*) izberemo 0 in kliknemo *Add*. Nato ponovimo postopek za vse vrednosti, ki so večje od 0 in sicer: Na levi strani (*Old Value*) izberemo *All other values*, na desni strani pa *Value:1* ter kliknemo *Add*, nato continue in *OK*.

c) Izračunajte 80 %-ni interval zaupanja za povprečno starost oseb v statistični množici.

Odgovori in izpisi rezultatov:

a)

Tabela 14: Opisna statistika za anketirane osebe moškega spola

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
Strah pred neuspehom	7	1	0	1	0,29	0,184	0,488	0,238
Valid N (listwise)	7							

Tabela 15: Opisna statistika za anketirane osebe ženskega spola

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
Strah pred neuspehom	8	1	0	1	0,75	0,164	0,463	0,214
Valid N (listwise)	8							

Med moškimi bi jih 29 % strah pred neuspehom odvrnil od ustanovitve podjetja, pri ženskah ta odstotek znaša 75 %.

c)

Tabela 16: 80 %-ni interval zaupanja za povprečno starost anketiranih oseb

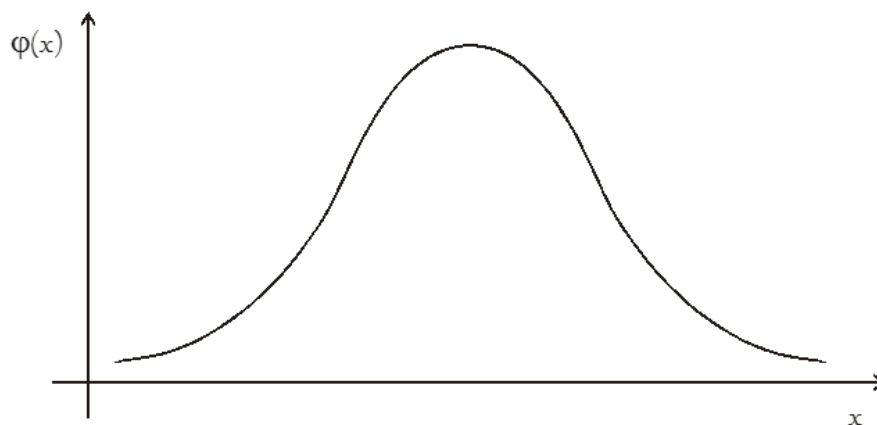
		Statistic	Std. Error	
Starost anketiranih oseb	Mean	39,40	4,263	
	80 % Confidence Interval for Mean	Lower Bound	33,67	
		Upper Bound	45,13	
	5 % Trimmed Mean	39,11		
	Median	33,00		
	Variance	272,543		
	Std. Deviation	16,509		
	Minimum	20		
	Maximum	64		
	Range	44		
	Interquartile Range	34		
	Skewness	0,236	0,580	
	Kurtosis	-1,656	1,121	

2 NORMALNA PORAZDELITEV

2.1 Opredelitev in značilnosti normalne porazdelitve

Normalna porazdelitev je najpomembnejša porazdelitev v teoriji in praksi statističnih postopkov. Na centralnem limitnem izreku in lastnostih normalne porazdelitve temelji inferenčna statistika (ali matematična statistika, statistično sklepanje o lastnostih in značilnostih statistične množice na osnovi značilnosti slučajnega vzorca) (Agresti & Finlay, 2009).

Normalna porazdelitev je verjetnostna porazdelitev vrednosti statističnih enot v statistični populaciji in jo shematično prikazujemo na sliki 1.



Slika 1: Prikaz normalne porazdelitve. (vir: Kožuh, 2005)

Krivulja normalne porazdelitve je zvonasta, unimodalna, simetrična in se asimptotično približuje osi x . Znana je tudi pod imenom Gaussova krivulja. Oblika normalne porazdelitve je odvisna od aritmetične sredine in od standardnega odklona (variance). Sprememba aritmetične sredine prestavlja krivuljo levo ali desno po osi x . Normalna porazdelitev ima vrh pri aritmetični sredini. Okoli te točke je gostota relativne frekvenca največja. Z oddaljenostjo od aritmetične sredine pa pada. Standardni odklon spreminja razpon normalne porazdelitve (z naraščanjem standardnega odklona se krivulja širi in znižuje). Standardni odklon meri razpršenost vrednosti okoli aritmetične sredine, zato vpliva na sploščenost krivulje (večji kot je odklon, bolj sploščena je krivulja) (Kožuh, 2005).

Pri normalni porazdelitvi je modus enak aritmetični sredini, zaradi simetričnosti pa je tema dvema enaka tudi mediana. Vrednosti levo in desno (navzgor in navzdol) od aritmetične sredine imajo vse manjšo verjetnostno gostoto.

Za normalno porazdelitev velja, da sta koeficienta asimetričnosti in sploščenosti enaka 0. Za proučevano (neko empirično) porazdelitev nam vrednosti teh dveh koeficientov nakazujeta, kako je proučevana porazdelitev podobna normalni porazdelitvi (Bastič, 2006).

2.2 Kolmogorov-Smirnov test in Shapiro-Wilk W test

Kolmogorov-Smirnov test in Shapiro-Wilk W test uporabimo takrat, kadar želimo preveriti trditev, da je obravnavana spremenljivka v statistični množici porazdeljena po normalni porazdelitvi, kar pomeni, da primerjamo pogostost pojavljanja proučevane spremenljivke na vzorčnih podatkih (empirične frekvence) s pogostostjo pojavljanja normalno porazdeljene spremenljivke z enako aritmetično sredino in standardnim odklonom, kot ga ima proučevana empirična porazdelitev (teoretične frekvence) (Bastič, 2006).

V primeru, ko je test neznačilen (stopnja značilnosti preizkusa ali stopnja tveganja, $p > 0,05$), tedaj ničelne domneve (H_0 : Obravnavani spremenljivki je dopustno prilagoditi normalno porazdelitev) ne zavrnamo in lahko sklepamo, da se porazdelitev proučevane spremenljivke v statistični množici ne razlikuje od normalne porazdelitve. Sklepamo torej, da proučevana spremenljivka ni normalno porazdeljena, ko je test statistično značilen ($p < 0,05$).

Naloga 1.

V podjetju so želeli izmeriti motiviranost svojih zaposlenih. Zaposleni so na 5-stopenjski Likertovi lestvici izrazili stopnjo strinjanja (kjer pomeni 1-sploh se ne strinjam in 5-popolnoma se strinjam) s štirimi trditvami o motiviranosti:

1. Na delovnem mestu in pri opravljanju svojega dela sem motiviran.
2. Vodstvo mi nudi ustrezno/primerno plačilo za moj dosežen uspeh.
3. Za doseganje uspešnih rezultatov v podjetju prejmemo zaposleni pohvalo ali priznanje.
4. Imam možnost kariernega razvoja.

Podatki so v datoteki *Normalna porazdelitev_Motiviranost.sav*.

a) Zapišite ničelno domnevo o normalni porazdelitvi vrednosti obravnavanih spremenljivk.

b) Preverite, ali so štiri spremenljivke - komponente motiviranosti zaposlenih, porazdeljene po normalni porazdelitvi in narišite frekvenčni histogram za spremenljivko *Za doseganje uspešnih rezultatov v podjetju prejmemo zaposleni pohvalo ali priznanje*.

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

Izpis rezultata:

Odgovori in izpisi rezultatov:

a)

H0: Obravnavani i -ti spremenljivki je dopustno prilagoditi normalno porazdelitev; $i=1,2,3,4$.

H1: Obravnavani i -ti spremenljivki ni dopustno prilagoditi normalne porazdelitve; $i=1,2,3,4$.

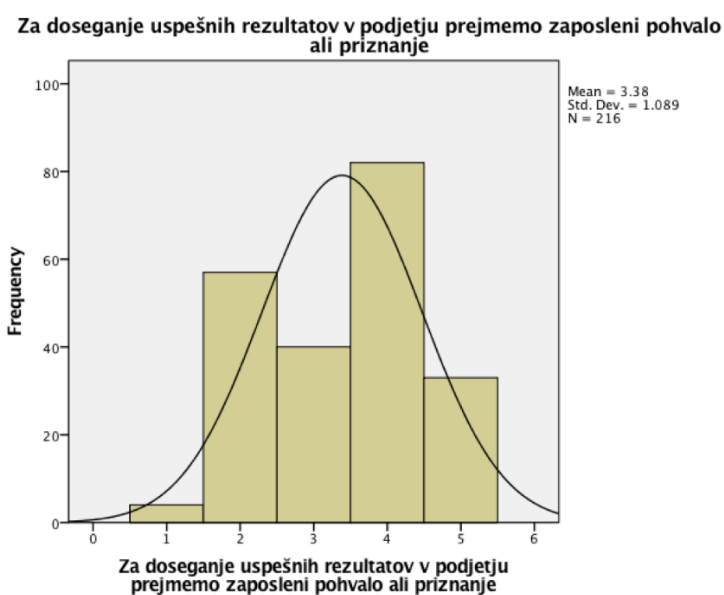
b)

Tabela 17: Kolmogorov-Smirnov test in Shapiro-Wilk W test (angl. Tests of Normality).

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Na delovnem mestu in pri opravljanju svojega dela sem motiviran.	0,221	216	0,000	0,881	216	0,000
Vodstvo mi nudi ustrezno/primerno plačilo za moj dosežen uspeh.	0,252	216	0,000	0,871	216	0,000
Za doseganje uspešnih rezultatov v podjetju prejmemo zaposleni pohvalo ali priznanje.	0,246	216	0,000	0,876	216	0,000
Imam možnost kariernega razvoja.	0,181	216	0,000	0,913	216	0,000

a. Lilliefors Significance Correction

Kolmogorov-Smirnov test in Shapiro-Wilk W test (tabela 17) kažeta, da nobena od štirih obravnavanih spremenljivk – komponent motiviranosti zaposlenih, ni porazdeljena po normalni porazdelitvi, saj je stopnja tveganja v vseh primerih manjša od 0,001 ($p < 0,001$), kar pomeni, da ničelno domnevo zavrnilo in potrdimo raziskovalno domnevo, da i -ta spremenljivka ni normalno porazdeljena; $i=1,2,3,4$. Odstopanja od oblike normalne porazdelitve, ki so opazna na sliki frekvenčnega histograma za eno od obravnavanih trditev, so statistično značilna.



Histogram 4: Frekvenčni histogram za spremenljivko Za doseganje uspešnih rezultatov v podjetju prejmemo zaposleni pohvalo ali priznanje s krivuljo prilagojene normalne porazdelitve.

Naloga 2.

Podjetje je na tržišču uvedlo nov izdelek. V podjetju želijo ugotoviti, koliko novih izdelkov je bilo prodanih v prvem tednu, na osnovi slučajnega vzorca 30-ih različnih trgovin. Podatki o prodaji novega izdelka v prvem tednu so v datoteki *Normalna porazdelitev_Prodaja izdelka.sav*.

a) Preverite ali je spremenljivka v statistični množici porazdeljena po normalni porazdelitvi in narišite frekvenčni histogram.

Odgovori in izpisi rezultatov:

a)

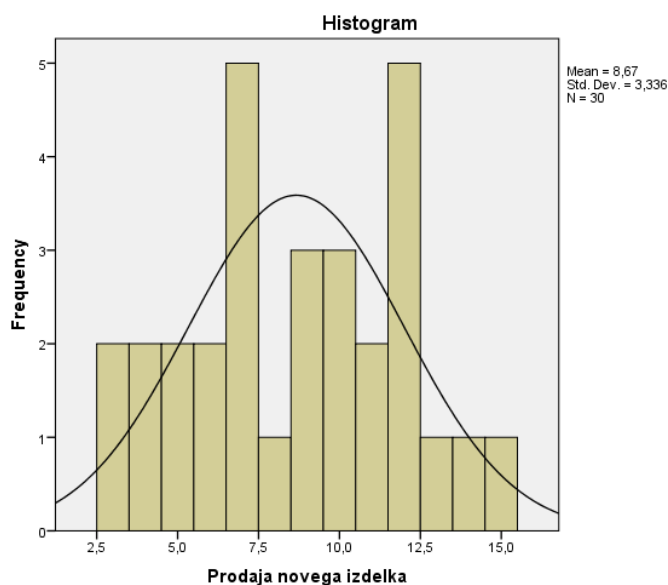
H0: Spremenljivki *prodaja izdelka* je dopustno prilagoditi normalno porazdelitev.

H1: Spremenljivki *prodaja izdelka* ni dopustno prilagoditi normalne porazdelitve.

Tabela 18: Kolmogorov-Smirnov test in Shapiro-Wilk W test (angl. Tests of Normality).

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Prodaja novega izdelka	0,125	30	0,200	0,964	30	0,384

Kolmogorov-Smirnov test in Shapiro-Wilk W test (tabela 18) kažeta, da ne zavrnamo ničelne domneve, da je obravnavana spremenljivka v statistični množici porazdeljena po normalni porazdelitvi, saj je stopnja tveganja pri Kolmogorov-Smirnovem testu 0,200 ter Shapiro-Wilkovem W testu 0,384, kar pa pomeni, da sta obe vrednosti stopnje tveganja večji od 0,05 (Kolmogorov-Smirnov 0,200 in Shapiro -Wilk 0,384).



Histogram 5: Frekvenčni histogram za spremenljivko prodaja izdelka

Naloga 3.

V datoteki *Normalna porazdelitev_Naloga3.sav* so podatki o podjetju, ki opravlja storitve na tržišču. Podatki prikazujejo vrednost opravljenih storitev v določenem letu.

- a) Preverite, ali lahko sklepamo, da je spremenljivki *vrednost storitev* dopustno prilagoditi normalno porazdelitev.
- b) Na osnovi mer centralne tendence, koeficienta asimetrije in koeficienta sploščenosti pojasnite ali je porazdelitev obravnavane spremenljivke (empirični podatki), po svojih značilnostih podobna normalni porazdelitvi.
- c) Narišite še frekvenčni histogram.

Odgovori in izpisi rezultatov:

a)

Tabela 19: Kolmogorov-Smirnov test in Shapiro-Wilk W test (angl. Tests of Normality).

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Vrednost storitev	0,317	70	0,000	0,440	70	0,000

b)

Tabela 20: Opisna statistika za spremenljivko vrednost storitev

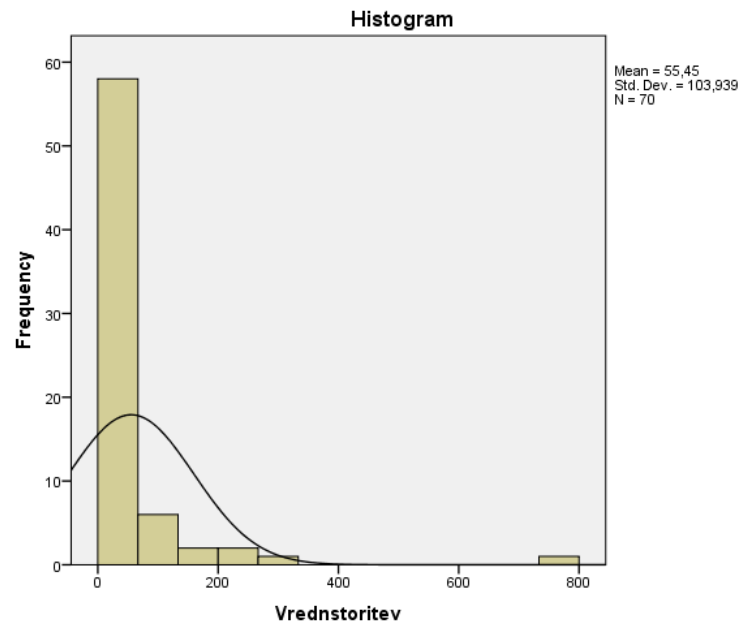
N	Valid	70
	Missing	0
Mean		55,45
Std. Error of Mean		12,423
Median		24,00
Mode		8^a
Std. Deviation		103,939
Variance		10803,400
Skewness		5,325
Std. Error of Skewness		0,287
Kurtosis		34,292
Std. Error of Kurtosis		0,566
Range		771
Minimum		6
Maximum		777

a. Multiple modes exist. The smallest value is shown

V tabeli 20 vidimo, da aritmetična sredina ni enaka mediani, prav tako tudi modus ni enak aritmetični sredini in mediani (za normalno porazdelitev velja: aritmetična sredina = mediana = modus).

Pomanjkljivost obravnavane empirične porazdelitve se kaže tudi v tem, da ima več kot en modus. Koeficient asimetričnosti kaže, da je porazdelitev asimetrična v desno (pozitivna asimetrična porazdelitev) in pozitivne vrednosti koeficienta sploščenosti kažejo na koničasto porazdelitev.

c)



Histogram 6: Frekvenčni histogram za spremenljivko vrednost storitev

3 PARAMETRIČNI IN NEPARAMETRIČNI UNIVARIATNI STATISTIČNI TESTI

Parametrični test za ugotavljanje statistično značilnih razlik med povprečnimi vrednostmi spremenljivk v vzorcih uporabljamo, kadar je podatkom za odvisno spremenljivko na intervalni ali razmernostni lestvici dopustno prilagoditi normalno porazdelitev (Corder & 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 vzorcih 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 vzorcih bomo uporabili analizo variance ANOVA.

Neparametrične ekvivalente parametričnim testom 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 & Foreman, 2014). Za primerjavo dveh neodvisnih vzorcev bomo uporabili 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.

3.1 Parametrični test za odvisna vzorca: t-test za odvisna vzorca

Naloga 1.

V podjetju X pri preverjanju kakovosti izdelkov ugotavljajo le, ali izdelek ustreza postavljenim merilom kakovosti ali ne. V prvem nadzornem obdobju so 25-krat preverili po 300 naključno izbranih izdelkov in ugotavljali število slabih izdelkov v vsaki podskupini. Ob koncu prvega obdobja so izvedli nekatere ukrepe za izboljšanje kakovosti izdelkov in nato v drugem nadzornem obdobju 25-krat preverili po 300 naključno izbranih izdelkov ter beležili število slabih izdelkov. V datoteki *t-test za odvisna vzorca_Nadzor kakovosti.sav* so v prvem stolpcu podatki o številu slabih izdelkov v prvem nadzornem obdobju, v drugem stolpcu pa so podatki o številu slabih izdelkov v drugem nadzornem obdobju. Ugotoviti želimo, ali so ukrepi za izboljšanje kakovosti uspešni. V ta namen bomo preverili naslednji domnevi:

H0: Povprečno število slabih izdelkov v prvem obdobju je enako povprečnemu številu slabih izdelkov drugem obdobju ($\mu_1 = \mu_2$).

H1: Povprečno število slabih izdelkov v prvem obdobju je večje kot povprečno število slabih izdelkov v drugem obdobju ($\mu_1 > \mu_2$).

Postopek:

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

V izboru *Analyze* izberemo *Compare Means*, nato pa *Paired-Samples T Test*. Označimo spremenljivko *število slabih izdelkov v prvem obdobju [slabi_1]* in jo s klikom na gumb s puščico prenesemo v okvir *Paired Variables*, in sicer v celico pod *Variable 1*, spremenljivko *število slabih izdelkov v drugem obdobju [slabi_2]* pa prenesemo v celico pod *Variable 2*. Kliknemo *OK*, da izvedemo t-test za dva odvisna vzorca.

Odgovori in izpisi rezultatov:

Tabela 21: Rezultati testov za preverjanje, ali je podatkom o številu slabih izdelkov v obeh nadzornih obdobjih v podjetju X dopustno prilagoditi normalno porazdelitev

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
število slabih izdelkov v prvem obdobju	0,115	25	0,200*	0,958	25	0,379
število slabih izdelkov v drugem obdobju	0,115	25	0,200*	0,958	25	0,371

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

a. Lilliefors Significance Correction

Iz tabele 21 je razvidno, da je pri obeh spremenljivkah $p > 0,05$, zato ničelne domneve Kolmogorov-Smirnovskega testa in Shapiro-Wilkovega testa ne zavrnamo. Številu slabih izdelkov v prvem obdobju je dopustno prilagoditi normalno porazdelitev. Normalno porazdelitev je dopustno prilagoditi tudi številu slabih izdelkov v drugem obdobju.

Tabela 22: Povprečno število slabih izdelkov v prvem in v drugem obdobju

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	število slabih izdelkov v prvem obdobju	14,88	25	5,442	1,088
	število slabih izdelkov v drugem obdobju	11,80	25	5,172	1,034

Iz tabele 22 je razvidno, da je povprečno število slabih izdelkov v podjetju X v prvem obdobju 14,88 (zaokroženo na cela mesta: 15 kosov), v drugem obdobju pa je 11,80 (zaokroženo na cela mesta: 12 kosov).

Tabela 23: 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	število slabih izdelkov v prvem obdobju - število slabih izdelkov v drugem obdobju	3,080	3,415	0,683	1,670	4,490	4,510	24	0,000

Iz rezultatov v zadnjem stolpcu v tabeli 23 je razvidno, da je izračunana dvostranska stopnja tveganja $p < 0,001$, kar velja tudi za enostransko stopnjo tveganja. Ker je $p < 0,001$, ničelno domnevo, ki smo jo zapisali pri Nalogi 1, zavrnamo. Povprečno število slabih izdelkov v prvem obdobju ni enako povprečnemu številu slabih izdelkov v drugem obdobju. Ob hkratnem upoštevanju rezultatov v tabeli 22 sprejememo raziskovalno domnevo: H1: Povprečno število slabih izdelkov v prvem obdobju je večje kot povprečno število slabih izdelkov v drugem obdobju.

Naloga 2.

Uspešnost ukrepov ekonomske politike v regiji smo proučevali na vzorcu 70 naključno izbranih podjetij. V ta namen smo zbrali podatke o donosnosti teh podjetij v letu pred uvedbo ukrepov ekonomske politike in v letu po uvedbi teh ukrepov. Ugotoviti želimo, ali je povprečna donosnost v letu po uvedbi ukrepov ekonomske politike statistično značilno višja kot v letu pred uvedbo teh ukrepov. Podatki so v datoteki *t-test za odvisna vzorca_Donosnost.sav*.

- Preverite, ali je podatkom za donosnost v letu pred uvedbo ukrepov ekonomske politike in v letu po uvedbi le-teh dopustno prilagoditi normalno porazdelitev.
- Utemeljite uporabo t-testa za odvisna vzorca, če želimo preveriti, ali je povprečna donosnost v letu po uvedbi ukrepov ekonomske politike statistično značilno višja kot v letu pred uvedbo teh ukrepov.
- Zapišite ničelno in raziskovalno domnevo.
- Preverite, ali je povprečna donosnost v letu po uvedbi ukrepov ekonomske politike statistično značilno višja kot v letu pred uvedbo teh ukrepov.

Odgovori in izpisi rezultatov:

a)

Tabela 24: Rezultati testov za preverjanje, ali je podatkom o donosnosti podjetij v letu pred ukrepi ekonomske politike in v letu po njih dopustno prilagoditi normalno porazdelitev

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Donosnost podjetja v letu pred ukrepi ekonomske politike	0,103	70	0,062	0,986	70	0,631
Donosnost podjetja v letu po ukrepih ekonomske politike	0,104	70	0,058	0,966	70	0,057

a. Lilliefors Significance Correction

Iz tabele 24 je razvidno, da je pri obeh spremenljivkah $p > 0,05$: podatkom za donosnost podjetij v letu pred ukrepi ekonomske politike in za donosnost podjetij v letu po ukrepih ekonomske politike je dopustno prilagoditi normalno porazdelitev.

b)

Spremenljivki *donosnost podjetja v letu pred ukrepi ekonomske politike* in *donosnost podjetja v letu po ukrepih ekonomske politike* temeljita na razmernostni lestvici. Podatkom je dopustno prilagoditi normalno porazdelitev. Enote obeh vzorcev so iz iste statistične množice.

c)

H0: Povprečna donosnost podjetij v letu pred ukrepi ekonomske politike in povprečna donosnost podjetij v letu po ukrepih ekonomske politike se statistično značilno ne razlikujeta ($\mu_1 = \mu_2$).

H1: Povprečna donosnost podjetij v letu po ukrepih ekonomske politike je statistično značilno višja kot povprečna donosnost v letu pred uvedbo le-teh ($\mu_2 > \mu_1$).

d)

Tabela 25: Povprečna donosnost v letu pred ukrepi ekonomske politike in v letu po njih

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Donosnost podjetja v letu pred ukrepi ekonomske politike	10,7429	70	4,05286	0,48441
	Donosnost podjetja v letu po ukrepih ekonomske politike	12,4000	70	4,11219	0,49150

Tabela 26: Rezultati t-testa za odvisna vzorca za donosnost

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95 % Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Donosnost podjetja v letu pred ukrepi ekonomske politike - Donosnost podjetja v letu po ukrepih ekonomske politike	-1,65714	5,93415	0,70927	-3,07209	-0,24220	-2,336	69	0,022

Ker je v raziskovalni domnevi zapisana stroga neenakost ($\mu_2 > \mu_1$), iz rezultatov v zadnjem stolpcu v tabeli 26 izračunamo enostransko stopnjo tveganja, ki je $p/2 = 0,022/2 = 0,011$, $p/2 < 0,05$. Povprečna donosnost podjetij v letu pred ukrepi ekonomske politike in povprečna donosnost v letu po njih se statistično značilno razlikujeta. Ob hkratnem upoštevanju rezultatov v tabeli 25 sprejmemo raziskovalno domnevo: H1: Povprečna donosnost podjetij v letu po ukrepih ekonomske politike je statistično značilno višja kot povprečna donosnost v letu pred uvedbo le-teh.

3.2 Parametrični test za neodvisna vzorca: t-test za neodvisna vzorca

Naloga 1.

V podjetju X želijo v drugem obdobju po sprejetju ukrepov za izboljšanje kakovosti primerjati kakovost svojih izdelkov s kakovostjo izdelkov v konkurenčnem podjetju Y. Tudi v podjetju Y pri preverjanju kakovosti izdelkov ugotavljajo le, ali izdelek ustreza postavljenim merilom kakovosti ali ne. V istem časovnem obdobju kot v podjetju X so v podjetju Y 30-krat preverili po 300 izdelkov in ugotavljali število slabih izdelkov v vsaki podskupini. V datoteki *t-test za neodvisna vzorca_Nadzor kakovosti.sav* so za drugo nadzorno obdobje v prvem stolpcu podatki o številu slabih izdelkov v podjetju X, v drugem stolpcu pa so podatki o številu slabih izdelkov v podjetju Y. Ugotoviti želimo, ali se kakovost izdelkov v podjetju X statistično značilno razlikuje od kakovosti izdelkov v konkurenčnem podjetju Y. V ta namen bomo preverili naslednji domnevi:

H0: Povprečno število slabih izdelkov v podjetju X je enako povprečnemu številu slabih izdelkov v podjetju Y ($\mu_X = \mu_Y$).

H1: Povprečno število slabih izdelkov v podjetju X se statistično značilno razlikuje od povprečnega števila slabih izdelkov v podjetju Y ($\mu_X \neq \mu_Y$).

Postopek:

S Kolmogorov-Smirnovim in Shapiro-Wilkovim testom smo najprej ugotovili, da je podatkom dopustno prilagoditi normalno porazdelitev (tabela 27). Zato za preverjanje zapisanih domnev uporabimo parametrični test za neodvisna vzorca, tj. t-test za neodvisna vzorca.

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

Odgovori in izpisi rezultatov:

Tabela 27: Rezultati testov za preverjanje, ali je podatkom o številu slabih izdelkov dopustno prilagoditi normalno porazdelitev

	oznaka podjetja	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
število slabih izdelkov	podjetje X	0,115	25	0,200*	0,958	25	0,371
	podjetje Y	0,097	30	0,200*	0,981	30	0,850

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

a. Lilliefors Significance Correction

Tabela 28: Povprečno število slabih izdelkov v podjetju X in v podjetju Y

	oznaka podjetja	N	Mean	Std. Deviation	Std. Error Mean
število slabih izdelkov	podjetje X	25	11,80	5,172	1,034
	podjetje Y	30	11,43	4,974	0,908

Tabela 29: Rezultati t-testa za neodvisna vzorca – podjetji X in Y

		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 slabih izdelkov	Equal variances assumed	0,000	0,991	0,267	53	0,790	0,367	1,371	-2,384	3,117
	Equal variances not assumed			0,266	50,448	0,791	0,367	1,376	-2,397	3,131

V tabeli 29 z Levenovim testom enakosti varianc najprej preverimo, ali lahko predpostavimo enake variance v obeh neodvisnih vzorcih:

H0: Predpostavimo enake variance.

H1: Enakih varianc ne predpostavimo.

Iz izpisa rezultatov Levenovega testa enakosti varianc vidimo, da je $p = 0,991$, torej $p > 0,05$, zato ničelne domneve ne zavrnamo, torej predpostavimo enake variance. V izpisu rezultatov t-testa za primerjavo povprečnih vrednosti neodvisnih vzorcev zato upoštevamo prvo vrstico izpisa. Odčitamo, da je $p = 0,790$, torej $p > 0,05$, torej ničelne domneve o enakosti povprečnih vrednosti ne zavrnamo: Povprečno število slabih izdelkov v podjetju X je enako povprečnemu številu slabih izdelkov v podjetju Y ($\mu_X = \mu_Y$). Kakovost v podjetju X se torej statistično značilno ne razlikuje od kakovosti v konkurenčnem podjetju Y.

Naloga 2.

Podjetje iz naloge 2 v poglavju o normalni porazdelitvi je nov izdelek uvedlo ne le na domačem tržišču A, ampak tudi na tujem tržišču B. V datoteki *t-test za neodvisna vzorca_Prodaja izdelka.sav* so podatki o prodaji v prvem tednu v 30 slučajno izbranih trgovinah na domačem tržišču A in o prodaji v drugih, 30 slučajno izbranih trgovinah na tujem tržišču B. V podjetju želijo ugotoviti, ali se je povprečna prodaja novega izdelka na domačem tržišču A statistično značilno razlikovala od povprečne prodaje na tujem tržišču B.

- Ugotovite, ali je podatkom o prodaji novega izdelka dopustno prirediti normalno porazdelitev.
- Utemeljite uporabo t-testa za neodvisna vzorca, če želimo preveriti, ali se je povprečna prodaja novega izdelka na domačem tržišču A statistično značilno razlikovala od povprečne prodaje na tujem tržišču B.
- Zapišite ničelno in raziskovalno domnevo.
- Ugotovite, ali se je povprečna prodaja novega izdelka na domačem tržišču A statistično značilno razlikovala od povprečne prodaje na tujem tržišču B.

Odgovori in izpisi rezultatov:

a)

Tabela 30: Rezultati testov za preverjanje, ali je podatkom o prodaji izdelkov dopustno prilagoditi normalno porazdelitev

	Tržišče	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Prodaja novega izdelka	Tržišče A	0,125	30	0,200*	0,964	30	0,384
	Tržišče B	0,109	30	0,200*	0,939	30	0,085

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

a. Lilliefors Significance Correction

b)

Trgovine so iz različnih statističnih množic: trgovine iz domačega tržišča A in trgovine iz tujega tržišča B. Podatki o prodaji so numerični in dopustno jim je prilagoditi normalno porazdelitev.

c)

H₀: Povprečna prodaja novega izdelka na domačem tržišču A je enaka kot povprečna prodaja novega izdelka na tujem tržišču B ($\mu_A = \mu_B$).

H₁: Povprečna prodaja novega izdelka na domačem tržišču A se statistično značilno razlikuje od povprečne prodaje novega izdelka na tujem tržišču B ($\mu_A \neq \mu_B$).

d)

Tabela 31: Opisna statistika neodvisnih vzorcev: domačega tržišča A in tujega tržišča B

	Tržišče	N	Mean	Std. Deviation	Std. Error Mean
Prodaja novega izdelka	Tržišče A	30	8,67	3,336	0,609
	Tržišče B	30	10,40	4,116	0,751

Tabela 32: Rezultati t-testa za neodvisna vzorca – domačega tržišča A in tujega tržišča B

		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
Prodaja novega izdelka	Equal variances assumed	2,195	0,144	-1,792	58	0,078	-1,733	0,967	-3,669	0,203
	Equal variances not assumed			-1,792	55,615	0,079	-1,733	0,967	-3,671	0,205

Iz izpisa rezultatov Levenovega testa enakosti varianc v tabeli 32 vidimo, da je $p > 0,05$ (izračunana stopnja tveganja je 0,144), torej predpostavimo enake variance. V izpisu rezultatov t-testa za primerjavo povprečnih vrednosti neodvisnih vzorcev zato upoštevamo prvo vrstico izpisa. Odčitamo, da je $p > 0,05$, torej ničelne domneve o enakosti povprečnih vrednosti ne zavrnemo: Povprečna prodaja novega izdelka na domačem tržišču A je enaka kot povprečna prodaja novega izdelka na tujem tržišču B ($\mu_A = \mu_B$). Ugotovimo, da se povprečna prodaja novega izdelka na domačem tržišču A statistično značilno ne razlikuje od povprečne prodaje novega izdelka na tujem tržišču B.

3.3 Parametrični test za neodvisne vzorce: ANOVA

Naloga 1.

V podjetju želijo ugotoviti, kateri ukrepi za izboljšanje kakovosti so najbolj uspešni. Ukrepi za izboljšanje kakovosti so: izboljšanje dizajna, zamenjava materiala in izobraževanje delavcev. V ta namen so 25-krat preverili po 300 izdelkov in ugotavljali število slabih izdelkov v vsaki podskupini. V datoteki *ANOVA_Nadzor kakovosti.sav* so v prvem stolpcu podatki o številu slabih izdelkov v podjetju, v drugem stolpcu je oznaka ukrepa za izboljšanje kakovosti. Ugotoviti želimo, ali se kakovost izdelkov v podjetju statistično značilno razlikuje glede na uporabljen ukrep. V ta namen smo oblikovali domnevi:

H0: Povprečno število slabih izdelkov je enako po vsakem izmed uporabljenih ukrepov za izboljšanje kakovosti ($\mu_1 = \mu_2 = \mu_3$).

H1: Vsaj dve povprečni števili slabih izdelkov se statistično značilno razlikujeta glede na uporabljen ukrep za izboljšanje kakovosti.

Postopek:

Najprej smo preverili, da je spremenljivki *število slabih izdelkov* dopustno prilagoditi normalno porazdelitev.

Nato za izvedbo ANOVA v izboru *Analyze* izberemo *Compare Means, One-way ANOVA* in odpre se okno *One-way ANOVA*, v katerem sta zapisani obe spremenljivki. Spremenljivko *število slabih izdelkov* [*slabi_izdelki*] prenesemo v okno pri *Dependent List*. Spremenljivko *ukrepi za izboljšanje kakovosti izdelkov* [*ukrep*] prenesemo v okno pri *Factor*. Kliknemo na *Options*, nato na *Descriptive* in potem *Continue*. Po kliku na *OK* dobimo izpis rezultatov.

Odgovori in izpisi rezultatov:

Tabela 33: Rezultati testov za preverjanje, ali je spremenljivkam za število slabih izdelkov po uporabi posameznih ukrepov (izboljšanje dizajna, zamenjava materiala in izobraževanje delavcev) dopustno prilagoditi normalno porazdelitev

	ukrepi za izboljšanje kakovosti izdelkov	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
število slabih izdelkov	izboljšanje dizajna	0,154	25	0,127	0,961	25	0,442
	zamenjava materiala	0,089	25	0,200*	0,967	25	0,563
	izobraževanje delavcev	0,082	25	0,200*	0,971	25	0,675

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

a. Lilliefors Significance Correction

Tabela 34: Opisna statistika - število slabih izdelkov po uporabi posameznih ukrepov

število slabih izdelkov								
	N	Mean	Std. Deviation	Std. Error	95 % Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
izboljšanje dizajna	25	7,36	4,672	0,934	5,43	9,29	0	17
zamenjava materiala	25	11,20	4,699	0,940	9,26	13,14	3	19
izobraževanje delavcev	25	9,84	4,819	0,964	7,85	11,83	2	19
Total	75	9,47	4,933	0,570	8,33	10,60	0	19

Tabela 35: Rezultati analize variance

število slabih izdelkov					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	189,547	2	94,773	4,235	0,018
Within Groups	1611,120	72	22,377		
Total	1800,667	74			

Iz izpisa rezultatov analize variance v tabeli 35 je razvidno, da je stopnja tveganja 0,018, torej je $p < 0,05$, zato ničelno domnevo zavrnejo in sprejmemo raziskovalno domnevo: H1: Vsaj dve povprečni števili slabih izdelkov se statistično značilno razlikujeta glede na uporabljen ukrep za izboljšanje kakovosti.

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 Tukeyev metodo tako, da izberemo *Tukey* (Bastič, 2006) in dobimo izpis v tabeli 36.

Tabela 36: Rezultati post-hoc testa pri analizi variance

Dependent Variable: število slabih izdelkov						
Tukey HSD						
(I) ukrepi za izboljšanje kakovosti izdelkov	(J) ukrepi za izboljšanje kakovosti izdelkov	Mean Difference (I-J)	Std. Error	Sig.	95 % Confidence Interval	
					Lower Bound	Upper Bound
izboljšanje dizajna	zamenjava materiala	-3,840*	1,338	0,015	-7,04	-0,64
	izobraževanje delavcev	-2,480	1,338	0,160	-5,68	0,72
zamenjava materiala	izboljšanje dizajna	3,840*	1,338	0,015	0,64	7,04
	izobraževanje delavcev	1,360	1,338	0,569	-1,84	4,56
izobraževanje delavcev	izboljšanje dizajna	2,480	1,338	0,160	-0,72	5,68
	zamenjava materiala	-1,360	1,338	0,569	-4,56	1,84

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

Iz tabele 36 je razvidno, da se povprečno število slabih izdelkov razlikuje po ukrepih izboljšanje dizajna in zamenjava materiala, stopnja tveganja je 0,015 ($p < 0,05$).

Naloga 2.

Po 50 naključno izbranih tujih turistov iz Avstrije, Italije in Nemčije smo povprašali, koliko denarja na dan nameravajo porabiti v Sloveniji. V datoteki *ANOVA_Predvidena dnevna poraba tujega turista.sav* so v stolpcu *poraba* zbrani njihovi odgovori o predvideni povprečni dnevni porabi v Sloveniji, izraženi v EUR, v stolpcu *država* pa je označeno, iz katere države prihajajo. Ugotoviti želimo, ali v Sloveniji tuji turisti v povprečju različno trošijo glede na državo, iz katere prihajajo.

- Opreделите vrsto spremenljivke *poraba* in vrsto spremenljivke *država*.
- Preverite, ali je spremenljivki *poraba* v vsakem posameznem neodvisnem vzorcu dopustno prilagoditi normalno porazdelitev.
- Zapišite ustrezni domnevi za zgoraj opisani problem, če boste uporabili ANOVA.
- Preverite domnevi.

Odgovori in izpisi rezultatov:

a)

Spremenljivka *poraba*, definirana kot »predvidena povprečna dnevna poraba tujega turista v Sloveniji, izražena v EUR« je numerična spremenljivka in temelji na razmernostni lestvici. Spremenljivka *država* je opisna spremenljivka in temelji na nominalni lestvici.

b)

Tabela 37: Rezultati testov za preverjanje, ali je spremenljivki poraba dopustno prilagoditi normalno porazdelitev

	država	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
predvidena povprečna dnevna poraba tujega turista v Sloveniji, izražena v EUR	Avstrija	0,107	50	0,200*	0,972	50	0,267
	Italija	0,088	50	0,200*	0,983	50	0,664
	Nemčija	0,091	50	0,200*	0,963	50	0,118

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

a. Lilliefors Significance Correction

c)

H0: Predvidena povprečna dnevna poraba tujih turistov v Sloveniji se glede na državo njihovega izvora ne razlikuje.

H1: Vsaj dve predvideni povprečni dnevni porabi tujih turistov v Sloveniji se glede na državo njihovega izvora statistično značilno razlikujeta.

d)

Tabela 38: Opisne statistike za predvideno dnevno porabo tujih turistov

predvidena povprečna dnevna poraba tujega turista v Sloveniji, izražena v EUR								
	N	Mean	Std. Deviation	Std. Error	95 % Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Avstrija	50	81,6872	26,86346	3,79907	74,0527	89,3217	30,00	135,00
Italija	50	104,9723	44,94987	6,35687	92,1977	117,7470	16,30	205,90
Nemčija	50	71,3170	27,98168	3,95721	63,3647	79,2693	10,00	120,60
Total	150	85,9922	36,85973	3,00958	80,0452	91,9392	10,00	205,90

Tabela 39: Rezultati ANOVA za predvideno dnevno porabo tujih turistov

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	29706,962	2	14853,481	12,641	0,000
Within Groups	172730,405	147	1175,037		
Total	202437,368	149			

Tabela 40: Rezultati Tukeyevega post hoc testa za predvideno dnevno porabo tujih turistov

Dependent Variable: predvidena povprečna dnevna poraba tujega turista v Sloveniji, izražena v EUR						
Tukey HSD						
(I) država	(J) država	Mean Difference (I-J)	Std. Error	Sig.	95 % Confidence Interval	
					Lower Bound	Upper Bound
Avstrija	Italija	-23,28518*	6,85576	,003	-39,5175	-7,0528
	Nemčija	10,37012	6,85576	,288	-5,8622	26,6025
Italija	Avstrija	23,28518*	6,85576	,003	7,0528	39,5175
	Nemčija	33,65530*	6,85576	,000	17,4230	49,8876
Nemčija	Avstrija	-10,37012	6,85576	,288	-26,6025	5,8622
	Italija	-33,65530*	6,85576	,000	-49,8876	-17,4230

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

3.4 Neparametrični test: χ^2 –test za analizo povezanosti dveh nominalnih spremenljivk

Naloga 1.

Ugotoviti želimo, ali sta spol in nakupovalne navade potrošnikov pri nakupu počitnic v hotelu povezani. 200 naključno izbranih potrošnikov je v vprašalniku poleg spola označilo tudi, ali so počitnice kupili preko spletnega ponudnika potovanj in povezanih storitev ali preko organizatorja potovanj ali kar v hotelu samem. Podatki so v datoteki *Hi-kvadrat_Nakupovalne navade.sav*. Preverjali smo naslednjo ničelno domnevo:

H0: Spremenljivki *spol* in *način nakupa počitnic* nista povezani.

Zapisali smo tudi raziskovalno domnevo:

H1: Spremenljivki *spol* in *način nakupa počitnic* sta povezani.

Postopek:

Izberemo *Analyze, Descriptive Statistics, Crosstabs*. Iz okna *Variables* prenesemo spremenljivko *Spol* v *Rows*, spremenljivko *Način nakupa počitnic* pa v *Columns* v okno na desni strani. Pri *Statistics* izberemo *Chi-square, Phi & Cramer's V, Lambda* in nadaljujemo s *Continue*. Pri *Cells* v okvirju *Count* izberemo *Observed in Expected*, v okvirju *Percentages* pa *Row* in nadaljujemo s *Continue*. Izbore potrdimo z *OK*.

Odgovori in izpisi rezultatov:

Tabela 41: Rezultati χ^2 – testa za preverjanje povezanosti spremenljivk spol in način nakupa počitnic

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	28,862 ^a	2	0,000
Likelihood Ratio	29,686	2	0,000
Linear-by-Linear Association	14,926	1	0,000
N of Valid Cases	200		

a. 0 cells (0,0 %) have expected count less than 5. The minimum expected count is 20,45.

Iz tabele 41 je v vrstici *Pearson Chi-Square* razvidno, da je $p < 0,001$, zato ničelno domnevo zavrnamo in sprejmemo raziskovalno domnevo: H1: Spremenljivki spol in nakupovalne navade potrošnikov pri nakupu počitnic v hotelu sta povezani.

Tabela 42: Frekvence po spolu in načinu nakupa počitnic

		Način nakupa počitnic			Total	
		Spletni ponudnik potovanj in povezanih storitev	Organizator potovanj	Hotel		
Spol	Moški	Count	54	16	17	87
		Expected Count	36,1	30,5	20,4	87,0
		% within Spol	62,1 %	18,4 %	19,5 %	100,0 %
	Ženski	Count	29	54	30	113
		Expected Count	46,9	39,6	26,6	113,0
		% within Spol	25,7 %	47,8 %	26,5 %	100,0 %
Total		Count	83	70	47	200
		Expected Count	83,0	70,0	47,0	200,0
		% within Spol	41,5 %	35,0 %	23,5 %	100,0 %

Tabela 43: Meri usmerjenosti pri χ^2 -testu za preverjanje povezanosti spremenljivk spol in način nakupa počitnic

			Value	Asymptotic Standard Error ^a	Approximate T ^b	Approximate Significance
Nominal by Nominal	Lambda	Symmetric	0,245	0,067	3,438	0,001
		Spol Dependent	0,287	0,088	2,797	0,005
		Način nakupa počitnic Dependent	0,214	0,069	2,797	0,005
	Goodman and Kruskal tau	Spol Dependent	0,144	0,049		0,000 ^c
		Način nakupa počitnic Dependent	0,085	0,029		0,000 ^c

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on chi-square approximation

Iz tabele 43 je razvidno, da je vrednost *lambda* večja od 0, kar nakazuje, da neodvisna spremenljivka pripomore k napovedovanju odvisne spremenljivke. Vrednosti *Goodman and Kruskal tau* v tabeli 43 pomenijo, da je zmanjšanje v napaki napovedovanja zaradi napačnega razvrščanja pri spremenljivki spol, ko poznamo vrednost spremenljivke način nakupa počitnic, 14,4 %, zmanjšanje v napaki napovedovanja način nakupa počitnic, ko poznamo vrednost spremenljivke spol, pa je 8,5 %.

Tabela 44: Meri simetričnosti pri χ^2 -testu za preverjanje povezanosti spremenljivk spol in način nakupa počitnic

		Value	Approximate Significance
Nominal by Nominal	Phi	0,380	0,000
	Cramer's V	0,380	0,000
N of Valid Cases		200	

Ker je kontingenčna tabela spol – način nakupa počitnic dimenzij 2 x 3 (tabela 42) in je torej večja kot 2 x 2, kot mero simetričnosti za oceno moči povezave med spremenljivkama uporabimo statistiko Cramer's V. Iz tabele 44 je razvidno, da je njena vrednost 0,38, kar nakazuje na srednje močno do močno statistično značilno ($p < 0,001$) povezavo med spremenljivkama spol in način nakupa počitnic (Corder in Foreman, 2014).

Naloga 2.

Z namenom ciljnega oglaševanja na družbenih omrežjih želimo ugotoviti, ali sta spremenljivki *generacija* in *družbeno omrežje* povezani. V datoteki *Hi-kvadrat_Družbena omrežja.sav* je za 150 naključno izbranih potrošnikov poleg oznake generacije, ki je X, Y in Z, tudi oznaka družbenega omrežja, ki ga posameznik največ uporablja – to je lahko družbeno omrežje T, družbeno omrežje F in družbeno omrežje I.

- a) Zapišite ustrezni domnevi.
- b) Preverite ničelno domnevo z ustreznim testom.

Odgovori in izpisi rezultatov:

a)

Preverjali smo naslednjo ničelno domnevo:

H0: Spremenljivki *generacija* in *družbeno omrežje* nista povezani.

Zapisali smo tudi raziskovalno domnevo:

H1: Spremenljivki *generacija* in *družbeno omrežje* sta povezani.

b)

Tabela 45: Frekvence po generaciji in družbenem omrežju, ki ga posameznik največ uporablja

			Družbeno omrežje, ki ga posameznik največ uporablja			Total
			Družbeno omrežje T	Družbeno omrežje F	Družbeno omrežje I	
Generacija	X	Count	33	16	9	58
		Expected Count	13,1	21,5	23,5	58,0
		% within Generacija	56,9 %	27,6 %	15,5 %	100,0 %
	Y	Count	7	35	28	70
		Expected Count	15,8	25,9	28,4	70,0
		% within Generacija	10,0 %	50,0 %	40,0 %	100,0 %
	Z	Count	5	23	44	72
		Expected Count	16,2	26,6	29,2	72,0
		% within Generacija	6,9 %	31,9 %	61,1 %	100,0 %
Total	Count	45	74	81	200	
	Expected Count	45,0	74,0	81,0	200,0	
	% within Generacija	22,5 %	37,0 %	40,5 %	100,0 %	

Tabela 46: Rezultati χ^2 –testa za preverjanje povezanosti spremenljivk generacija in družbeno omrežje, ki ga posameznik največ uporablja

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	64,681 ^a	4	0,000
Likelihood Ratio	61,283	4	0,000
Linear-by-Linear Association	47,204	1	0,000
N of Valid Cases	200		

a. 0 cells (0,0 %) have expected count less than 5. The minimum expected count is 13,05.

Tabela 47: Meri usmerjenosti pri χ^2 –testu za preverjanje povezanosti spremenljivk generacija in družbeno omrežje, ki ga posameznik največ uporablja

			Value	Asymptotic Standard Error ^a	Approximate T ^b	Approximate Significance
Nominal by Nominal	Lambda	Symmetric	0,287	0,064	4,021	0,000
		Generacija Dependent	0,313	0,063	4,264	0,000
		Družbeno omrežje, ki ga posameznik največ uporablja Dependent	0,261	0,074	3,097	0,002
	Goodman and Kruskal tau	Generacija Dependent	0,153	0,036		0,000 ^c
		Družbeno omrežje, ki ga posameznik največ uporablja Dependent	0,141	0,036		0,000 ^c

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on chi-square approximation

Tabela 48: Meri simetričnosti pri χ^2 –testu za preverjanje povezanosti spremenljivk generacija in družbeno omrežje, ki ga posameznik največ uporablja

		Value	Approximate Significance
Nominal by Nominal	Phi	0,569	0,000
	Cramer's V	0,402	0,000
N of Valid Cases		200	

3.5 Neparametrični test za odvisna vzorca: Wilcoxon signed ranks test

Naloga 1.

Proučevali smo uspešnost ukrepov za izboljšanje neugodnih delovnih razmer v organizaciji Y. V ta namen so naključno izbrani zaposleni v tej organizaciji na 5-stopenjski ordinalni lestivici (rang 1 – ne povzroča stresa, rang 2 – rahlo stresno, rang 3 – zmerno stresno, rang 4 – zelo stresno, rang 5 – izjemno stresno) označili *stopnjo stresa pri dejavniku neugodne delovne razmere* v letu pred uvedbo ukrepov in eno leto po uvedbi le-teh. Podatki so v datoteki *Wilcoxon signed ranks test_Neugodne delovne razmere.sav*. Preverili bomo naslednjo ničelno domnevo:

H0: Povprečni rang *stopnje stresa zaradi neugodnih delovnih razmer* se v letih pred in po uvedbi ukrepov ne razlikuje.

Zapisali smo tudi raziskovalno domnevo:

H1: Povprečni rang *stopnje stresa zaradi neugodnih delovnih razmer* je v letu pred uvedbo ukrepov statistično značilno višji kot eno leto po uvedbi le-teh.

Postopek:

Da bi izračunali mediano za prvi in drugi odvisni vzorec, v izboru *Analyze* izberemo *Descriptive Statistics* in nato *Frequencies*, prenesemo obe spremenljivki v desno okno *Variable(s)*, kliknemo na gumb *Statistics* in v razdelku *Central Tendency* označimo *Median*. Dobimo tabelo 47.

Da bi preverili ničelno domnevo, v izboru *Analyze* izberemo *Nonparametric Tests*, nato pa *Legacy Dialogs* in *2 Related Samples*. Prenesemo spremenljivko *stopnja stresa zaradi neugodnih delovnih razmer pred uvedbo ukrepov* kot *Variable 1*, spremenljivko *stopnja stresa zaradi neugodnih delovnih razmer po uvedbi ukrepov* kot *Variable 2*, da oblikujemo *Pair 1*. V okviru *Test Type* naj bo označeno *Wilcoxon*. Kliknemo *Options* in označimo *Descriptives* ter se s klikom na *Continue* vrnemo v glavno pogovorno okno. Kliknemo *OK*, da izvedemo Wilcoxon Signed Rank Test za dva odvisna vzorca, katerega rezultati so izpisani v tabelah 48 in 49.

Odgovori in izpisi rezultatov:

Tabela 49: Mediana ranga stopnje stresa zaradi neugodnih delovnih razmer pred uvedbo ukrepov in po njej

		stopnja stresa zaradi neugodnih delovnih razmer pred uvedbo ukrepov	stopnja stresa zaradi neugodnih delovnih razmer po uvedbi ukrepov
N	Valid	80	80
	Missing	0	0
Median		4,00	3,00

Iz tabele 49 je razvidno, da je mediana *stopnje stresa zaradi neugodnih delovnih razmer* pred uvedbo ukrepov 4, po uvedbi ukrepov pa je 3.

Tabela 50: Rangji pri izvedbi Wilcoxon Signed-Rank testa

		N	Mean Rank	Sum of Ranks
stopnja stresa zaradi neugodnih delovnih razmer po uvedbi ukrepov - stopnja stresa zaradi neugodnih delovnih razmer pred uvedbo ukrepov	Negative Ranks	51 ^a	28,22	1439,00
	Positive Ranks	4 ^b	25,25	101,00
	Ties	25 ^c		
	Total	80		

a. stopnja stresa zaradi neugodnih delovnih razmer po uvedbi ukrepov < stopnja stresa zaradi neugodnih delovnih razmer pred uvedbo ukrepov

b. stopnja stresa zaradi neugodnih delovnih razmer po uvedbi ukrepov > stopnja stresa zaradi neugodnih delovnih razmer pred uvedbo ukrepov

c. stopnja stresa zaradi neugodnih delovnih razmer po uvedbi ukrepov = stopnja stresa zaradi neugodnih delovnih razmer pred uvedbo ukrepov

Tabela 51: Rezultati Wilcoxon Signed-Rank testa za rang stopnje stresa zaradi neugodnih delovnih razmer

	stopnja stresa zaradi neugodnih delovnih razmer po uvedbi ukrepov - stopnja stresa zaradi neugodnih delovnih razmer pred uvedbo ukrepov
Z	-5,835 ^b
Asymp. Sig. (2-tailed)	0,000

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

Iz izpisov v tabeli 51 je razvidno, da je $|Z| > 1,96$ in da je $p/2 < 0,001$. Preverimo še rezultate v tabelah 49 in 50. Zavrնemo ničelno domnevo in sprejmemo raziskovalno domnevo: H1: Povprečni rang *stopnje stresa zaradi neugodnih delovnih razmer* je v letu pred uvedbo ukrepov statistično značilno višji kot eno leto po uvedbi le-teh.

Naloga 2.

V poglavju o normalni porazdelitvi smo pri nalogi 3 že ugotovili, da vrednosti opravljenih storitev na določenem tržišču ni dopustno prilagoditi normalne porazdelitve. Na tem trgu se je na začetku novega leta pojavil nov konkurent (Bastič in drugi, 2008). V datoteki *Wilcoxon Signed Rank_Vrednost storitev.sav* so poleg podatkov o vrednosti opravljenih storitev na tem tržišču v prvem opazovanem letu še podatki o vrednosti opravljenih storitev na tem tržišču v drugem opazovanem letu. Preveriti želimo, ali se povprečni rang vrednosti opravljenih storitev v prvem opazovanem letu statistično značilno razlikuje od povprečnega ranga vrednosti opravljenih storitev v drugem opazovanem letu.

- a) Zapišite ničelno in raziskovalno domnevo.
- b) Preverite ničelno domnevo in zapišite izid preverjanja. Pri tem upoštevajte, da so bili vrednostim spremenljivke prirejani rangi.

Odgovori in izpisi rezultatov:

a)

H0: Povprečni rang vrednosti opravljenih storitev v prvem opazovanem letu se statistično značilno ne razlikuje od povprečnega ranga vrednosti opravljenih storitev v drugem opazovanem letu.

H1: Povprečni rang vrednosti opravljenih storitev v prvem opazovanem letu se statistično značilno razlikuje od povprečnega ranga vrednosti opravljenih storitev v drugem opazovanem letu.

b)

Tabela 52: Opisna statistika - vrednosti storitev, za dva odvisna vzorca

	N	Mean	Std. Deviation	Minimum	Maximum
vrednost storitev na prvem tržišču v prvem letu	70	55,4500	103,93940	6,00	776,50
vrednost storitev na prvem tržišču v drugem letu	70	52,0747	83,30983	5,00	567,00

Tabela 53: Rangi za vrednost storitev za dva odvisna vzorca

		N	Mean Rank	Sum of Ranks
vrednost storitev na prvem tržišču v drugem letu - vrednost storitev na prvem tržišču v prvem letu	Negative Ranks	39 ^a	34,03	1327,00
	Positive Ranks	31 ^b	37,35	1158,00
	Ties	0 ^c		
	Total	70		

a. vrednost storitev na prvem tržišču v drugem letu < vrednost storitev na prvem tržišču v prvem letu

b. vrednost storitev na prvem tržišču v drugem letu > vrednost storitev na prvem tržišču v prvem letu

c. vrednost storitev na prvem tržišču v drugem letu = vrednost storitev na prvem tržišču v prvem letu

Tabela 54: Rezultati Wilcoxon Signed-Rank testa za vrednost storitev

	vrednost storitev na prvem tržišču v drugem letu - vrednost storitev na prvem tržišču v prvem letu
Z	-0,495 ^b
Asymp. Sig. (2-tailed)	0,621

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

Ker je $p/2 = 0,621/2 = 0,3105$, oz. je $p > 0,05$, ne zavrnamo ničelne domneve in zaključimo, da se povprečni rang vrednosti opravljenih storitev v prvem opazovanem letu statistično značilno ne razlikuje od povprečnega ranga vrednosti opravljenih storitev v drugem opazovanem letu.

3.6 Neparametrični test za neodvisna vzorca: Mann-Whitney U test

Naloga 1.

V raziskavi o uporabi kvantitativnih metod v podjetjih smo proučevali pogostost uporabe statističnih metod za napovedovanje prodaje (1 – nikoli, 2 – dvakrat letno, 3 – mesečno, 4 – tedensko, 5 – dnevno). Podatki o pogostosti uporabe statističnih metod za napovedovanje prodaje, ki smo jih pridobili v naključno izbranih 62 domačih in 137 tujih podjetjih, so razvidni iz datoteke *Mann-Whitney U test_Pogostost uporabe.sav*. Ugotoviti želimo, ali se povprečna ranga pogostosti uporabe statističnih metod za napovedovanje prodaje domačih in tujih podjetij statistično značilno razlikujeta. V ta namen smo preverili naslednjo ničelno domnevo:

H0: Povprečna ranga pogostosti uporabe statističnih metod za napovedovanje prodaje domačih in tujih podjetij se statistično značilno ne razlikujeta.

Zapisali smo tudi domnevo:

H1: Povprečna ranga pogostosti uporabe statističnih metod za napovedovanje prodaje domačih in tujih podjetij se statistično značilno razlikujeta.

Postopek:

V izboru *Analyze* izberemo *Nonparametric Tests*, nato *Legacy Dialogs* in nato *2 Independent Samples*. Odvisno spremenljivko *pogostost uporabe statističnih metod za napovedovanje prodaje* prenesemo v desno okno *Test Variable List*. Nato spremenljivko *podjetje* prenesemo v okno *Grouping Variable*. Kliknemo na *Define Groups* in definiramo vrednosti skupin: pri Group 1 vpišemo 1, pri Group 2 pa vpišemo 2 in kliknemo *Continue*. V okviru *Test Type* naj bo označeno *Mann-Whitney U*. Kliknemo *OK*, da izvedemo Mann-Whitney U test za dva neodvisna vzorca in dobimo rezultate v tabelah 55 in 56.

Odgovori in izpisi rezultatov:

Tabela 55: Rangi za pogostost uporabe statističnih metod za napovedovanje prodaje za dva neodvisna vzorca

	podjetje	N	Mean Rank	Sum of Ranks
pogostost uporabe statističnih metod za napovedovanje prodaje	domače podjetje	62	86,83	5383,50
	tuje podjetje	137	105,96	14516,50
	Total	199		

Iz izpisa rezultatov v tabeli 55 je razvidno, da je povprečni rang *pogostosti uporabe statističnih metod za napovedovanje prodaje* v tujih podjetjih višji kot v domačih. Enako velja tudi za vsoto rangov.

Tabela 56: Rezultati Mann-Whitneyevega U testa za pogostost uporabe

	pogostost uporabe statističnih metod za napovedovanje prodaje
Mann-Whitney U	3430,500
Wilcoxon W	5383,500
Z	-2,231
Asymp. Sig. (2-tailed)	0,026

a. Grouping Variable: podjetje

Iz izpisa rezultatov v tabeli 56 je razvidno, da je $|Z| > 1,96$ in da je $p < 0,05$. Na osnovi rezultatov zavrnemo ničelno domnevo in sprejmemo raziskovalno domnevo: H1: Povprečna ranga pogostosti uporabe statističnih metod za napovedovanje prodaje domačih in tujih podjetij se statistično značilno razlikujeta.

Naloga 2.

V raziskavi o zadovoljstvu zaposlenih v podjetju X smo na 5-stopenjski Likertovi lestvici (1 – sploh se ne strinjam, 5 – popolnoma se strinjam) merili zadovoljstvo zaposlenih v oddelkih Informatika in Prodaja. Stopnjo strinjanja s trditvijo *V podjetju sem zadovoljen z delovnimi pogoji* je izrazilo 15 naključno izbranih zaposlenih v oddelku Informatika in 23 naključno izbranih zaposlenih v oddelku Prodaja. Odgovori zaposlenih so zbrani v datoteki *Mann-Whitney U test_Zadovoljstvo z delovnimi pogoji.sav*. Podatki o zadovoljstvu zaposlenih temeljijo torej na intervalni lestvici. S Kolmogorov-Smirnovim in Shapiro-Wilkovim testom smo najprej ugotovili, da podatkom za spremenljivko *zadovoljstvo zaposlenih z delovnimi pogoji* v oddelku Prodaja ni dopustno prilagoditi normalne porazdelitve, kar je razvidno iz tabele 57. Zato se vrednostim spremenljivke priredijo rangi.

- a) Zapišite ustrezni domnevi za preverjanje, ali v povprečnih rangih o zadovoljstvu zaposlenih z delovnimi pogoji v dveh omenjenih oddelkih obstajajo statistično značilne razlike.
- b) Preverite pri nalogi a) zapisano ničelno domnevo.

Postopek:

a)

H₀: Povprečna ranga o zadovoljstvu zaposlenih z delovnimi pogoji v oddelkih Informatika in Prodaja se statistično značilno ne razlikujeta.

H₁: Povprečna ranga o zadovoljstvu zaposlenih z delovnimi pogoji v oddelkih Informatika in Prodaja se statistično značilno razlikujeta.

b)

Za preverjanje pri nalogi a) zapisane ničelne domneve uporabimo neparametrični test za neodvisna vzorca, tj. Mann-Whitney U test. Pri tem so vrednostim spremenljivke prirejeni rangi.

V izboru *Analyze* izberemo *Nonparametric Tests*, nato *Legacy Dialogs* in nato *2 Independent Samples*. Odvisno spremenljivko *zadovoljstvo zaposlenih z delovnimi pogoji* prenesemo v desno okno *Test Variable List*. Nato spremenljivko *oddelek v podjetju* prenesemo v okno *Grouping Variable*. Kliknemo na *Define Groups* in definiramo vrednosti skupin: pri Group 1 vpišemo 1, pri Group 2 pa vpišemo 2 in kliknemo *Continue*. V okviru *Test Type* naj bo označeno *Mann-Whitney U*. Kliknemo *OK*, da izvedemo Mann-Whitney U test za dva neodvisna vzorca in dobimo rezultate v tabelah 58 in 59.

Odgovori in izpisi rezultatov:

a)

Tabela 57: Rezultati testov za preverjanje, ali je podatkom o zadovoljstvu zaposlenih z delovnimi pogoji dopustno prilagoditi normalno porazdelitev

	oddelek v podjetju	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
zadovoljstvo zaposlenih z delovnimi pogoji	Informatika	0,207	15	0,084	0,917	15	0,175
	Prodaja	0,231	23	0,003	0,880	23	0,010

a. Lilliefors Significance Correction

b)

Tabela 58: Izpis rangov, dobljenih z Mann-Whitney U testom za zadovoljstvo zaposlenih z delovnimi pogoji

	oddelek v podjetju	N	Mean Rank	Sum of Ranks
zadovoljstvo zaposlenih z delovnimi pogoji	Informatika	15	25,53	383,00
	Prodaja	23	15,57	358,00
	Total	38		

Tabela 59: Rezultati Mann-Whitney U testa za zadovoljstvo zaposlenih z delovnimi pogoji

	zadovoljstvo zaposlenih z delovnimi pogoji
Mann-Whitney U	82,000
Wilcoxon W	358,000
Z	-2,798
Asymp. Sig. (2-tailed)	0,005
Exact Sig. [2*(1-tailed Sig.)]	0,006 ^b

a. Grouping Variable: oddelek v podjetju

b. Not corrected for ties.

Iz tabele 59 je razvidno, da je $|Z| > 1,96$, $p < 0,01$, zato zavrnilo ničelno domnevo in sprejmemo raziskovalno domnevo: H1: Povprečna ranga o zadovoljstvu zaposlenih z delovnimi pogoji v oddelkih Informatika in Prodaja se statistično značilno razlikujeta.

Naloga 3.

V poglavju o normalni porazdelitvi smo pri nalogi 3 že ugotovili, da vrednostim te spremenljivke na določenem tržišču ni dopustno prilagoditi normalne porazdelitve. V datoteki *Mann-Whitney U test_Vrednost storitev.sav* so poleg podatkov o vrednosti 70 opravljenih storitev na prvem tržišču tudi podatki o vrednostih 40 opravljenih storitev na drugem tržišču (Bastič in drugi, 2008).

- Za opisani primer zapišite ustrezni domnevi.
- Preverite ničelno domnevo, zapisano pri nalogi a).
- Pojasnite pri nalogi b) dobljene rezultate.

Odgovori in izpisi rezultatov:

a)

Ker spremenljivki *vrednost storitev* ni dopustno prilagoditi normalne porazdelitve, bomo za primerjanje dveh neodvisnih vzorcev uporabili neparametrični test, tj. Mann-Whitney U test. Vrednostim te spremenljivke se bodo pri izvedbi tega testa priredili rangi. Zato bomo preverjali naslednjo ničelno domnevo:

H0: Povprečni rang vrednosti opravljenih storitev na prvem tržišču je enak kot povprečni rang vrednosti opravljenih storitev na drugem tržišču.

Zapisali smo tudi raziskovalno domnevo:

H1: Povprečna ranga vrednosti opravljenih storitev na prvem in drugem tržišču se razlikujeta.

b)

Tabela 60: Rangi za vrednost opravljenih storitev za prvo in drugo tržišče

	tržišče	N	Mean Rank	Sum of Ranks
vrednost storitev	prvo tržišče	70	61,16	4281,00
	drugo tržišče	40	45,60	1824,00
	Total	110		

Tabela 61: Rezultati Mann-Whitney U testa za neodvisna vzorca za primer vrednosti storitev

	vrednost storitev
Mann-Whitney U	1004,000
Wilcoxon W	1824,000
Z	-2,461
Asymp. Sig. (2-tailed)	0,014

a. Grouping Variable: tržišče

c)

Iz tabele 60 je razvidno, da je povprečni rang vrednosti opravljenih storitev na prvem tržišču višji kot povprečni rang vrednosti opravljenih storitev na drugem tržišču. Iz tabele 61 je razvidno, da je $p < 0,05$, zato ničelno domnevo zavrnemo in sprejmemo raziskovalno domnevo: H1: Povprečna ranga vrednosti opravljenih storitev na prvem in drugem tržišču se razlikujeta.

4 FAKTORSKA ANALIZA

Faktorsko analizo uporabimo, kadar je mogoče iz večjega števila medsebojno povezanih merjenih spremenljivk, oblikovati manjše število med seboj neodvisnih faktorjev (kadar so izpolnjeni pogoji za izvedbo faktorske analize). Faktorji predstavljajo linearno kombinacijo merjenih spremenljivk (Tabachnick & Fidell, 2013).

Naloga 1.

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 oziroma kakšen je njihov odnos do sprejema mobilnih oglasov.

S tem namenom je bila zasnovana anketna raziskava, ki je temeljila na vprašalniku s 5-stopenjsko Likertovo lestvico, s katero podjetje meri oziroma ugotavlja:

- 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.

Na osnovi podatkov v datoteki *Faktorska analiza_Mobilno oglaševanje.sav* izvedite faktorsko analizo za večdimenzionalno spremenljivko *odnos do sprejema mobilnih oglasov* in pojasnite rezultate.

Potek za izvedbo faktorske analize: Kliknemo *Analyze*, izberemo *Dimension Reduction* ter nato *Factor*. V desno okence prenesemo tiste spremenljivke, za katere želimo izvesti faktorsko analizo. V našem primeru spremenljivke *V2.1* do *V2.4* prenesemo v desno okence. S klikom na gumb *Descriptives* se odpre novo pogovorno okno, v katerem pri *Statistics* izberemo *Univariate descriptives* in *Initial Solution*. V okencu *Correlation Matrix* izberemo *Coefficients*, *Significance levels* in *KMO and Bartlett's Test of sphericity* ter kliknemo na *Continue*.

Kliknemo gumb *Extraction*. Med ponujenimi metodami izberemo *Principal Components*, v okencu *Analyze* izberemo *Correlation matrix*, v okencu *Extract* pa pri *Eigenvalues over* vtipkamo *1*. S klikom na *Continue* se vrnemo v osnovno pogovorno okno.

Kliknemo gumb *Rotation*. Med ponujenimi metodami izberemo *Varimax* (je najbolj poznana metoda, pri kateri skušamo doseči, da ima vsaka opazovana spremenljivka visoko faktorsko utež le pri enem faktorju, pri drugih faktorjih pa so uteži čim manjše – izrazitejše faktorske uteži omogočajo lažjo interpretacijo), v okencu *Display* izberemo *Rotated Solution*. S klikom na *Continue* se vrnemo v osnovno pogovorno okno. Kliknemo *Scores* in izberemo *Save as variables*. S klikom na *Continue* se vrnemo v osnovno pogovorno okno in kliknemo *OK*.

Odgovori in izpisi rezultatov:

Tabela 62: Korelacijska matrika

		V2.1 Zdi se mi da je ideja o uporabljanju mobilnih oglasov dobra	V2.2 Imam pozitiven odnos do sprejema mobilnih oglasov	V2.3 Za mobilne oglase mi je vseeno	V2.4 Mobilno oglaševanje je zanimivo
Correlation	V2.1 Zdi se mi da je ideja o uporabljanju mobilnih oglasov dobra	1,000	0,366	-0,265	0,436
	V2.2 Imam pozitiven odnos do sprejema mobilnih oglasov	0,366	1,000	-0,417	0,581
	V2.3 Za mobilne oglase mi je vseeno	-0,265	-0,417	1,000	-0,385
	V2.4 Mobilno oglaševanje je zanimivo	0,436	0,581	-0,385	1,000
Sig. (1-tailed)	V2.1 Zdi se mi da je ideja o uporabljanju mobilnih oglasov dobra		0,013	0,057	0,004
	V2.2 Imam pozitiven odnos do sprejema mobilnih oglasov	0,013		0,005	0,000
	V2.3 Za mobilne oglase mi je vseeno	0,057	0,005		0,009
	V2.4 Mobilno oglaševanje je zanimivo	0,004	0,000	0,009	

Iz korelacijske matrike v tabeli 62 izhaja, da so spremenljivke med seboj večinoma statistično značilno in srednje močno pozitivno povezane, razen spremenljivke V2.3, kjer je povezanost z ostalimi spremenljivkami negativna. Za lažjo interpretacijo dobljenih rezultatov faktorске analize bomo zato vrednosti spremenljivke V2.3 najprej prekodirali v obratnem vrstnem redu (Vrednost 5 preide v vrednost 1, vrednost 2 v vrednost 4 itd.).

V nadaljevanju so prikazani rezultati faktorске analize, kjer so vrednosti spremenljivke V2.3 prekodirane.

Tabela 63: Kaiser-Meyer-Olkinova statistika in Bartlettov test sferičnosti za večdimenzionalno spremenljivko odnos do sprejema mobilnih oglasov

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0,734
Bartlett's Test of Sphericity	Approx. Chi-Square	29,817
	df	6
	Sig.	0,000

Smiselnost uporabe faktorjske analize preizkušamo z Bartlettovim testom sferičnosti. Z njim preizkušamo ničelno domnevo, da je osnovna korelacijska matrika enaka matriki enote, kar pomeni, da ne obstaja povezanost med opazovanimi (merjenimi) spremenljivkami. Poleg Bartlettovega testa sferičnosti se uporablja še Kaiser-Meyer-Olkinova statistika (KMO), za katero je značilno, da je uporaba faktorjske analize smiselna pri vrednosti, ki je večja od 0,5 (Tabachnick in Fidell, 2013). V našem primeru KMO (0,734) in stopnja značilnosti Bartlettovega testa ($p < 0,001$) kažeta na smiselnost uporabe faktorjske analize (tabela 63).

Tabela 64: Komunalitete za večdimenzionalno spremenljivko odnos do sprejema mobilnih oglasov

	Initial	Extraction
V2.1 Zdi se mi da je ideja o uporabljanju mobilnih oglasov dobra	1,000	0,452
V2.2 Imam pozitiven odnos do sprejema mobilnih oglasov	1,000	0,657
V2.3 Za mobilne oglase mi je vseeno	1,000	0,452
V2.4 Mobilno oglaševanje je zanimivo	1,000	0,680

Posamezne trditve je potrebno proučiti glede na vrednost komunalitet, ki morajo biti višje od 0,40 (ibid). Vrednosti vseh komunalitet v tabeli 64 za večdimenzionalno spremenljivko *odnos do sprejema mobilnih oglasov* so višje od 0,40, zato nismo izločili nobene spremenljivke.

Tabela 65: Celotna pojasnjena varianca za večdimenzionalno spremenljivko odnos do sprejema mobilnih oglasov

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,240	55,997	55,997	2,240	55,997	55,997
2	0,743	18,586	74,583			
3	0,607	15,164	89,747			
4	0,410	10,253	100,000			

Iz tabele 65 je razvidno, da smo iz štirih merjenih spremenljivk pridobili tudi 4 glavne komponente, a le pri prvi glavni komponenti je lastna vrednost večja od 1. V nadaljevanju je ob upoštevanju tega kriterija (lastna vrednost večja od 1), zato upoštevan oziroma izločen samo en, to je prvi faktor. Ta faktor pojasnjuje nekoliko manj kot 56 % variabilnosti merjenih spremenljivk skupaj. Glede na tak dobljeni rezultat, bi bilo smiselno razmisliti o upoštevanju dvofaktorske rešitve, saj prva dva faktorja skupaj pojasnjujeta skoraj 75 % variabilnost merjenih spremenljivk skupaj. V nadaljevanju prikazujemo enofaktorsko rešitev; rotacija faktorjev ni bila izvedena, saj je le-to smiselno izvesti le, kadar sta izločena vsaj dva faktorja in dobljenih faktorjev ni mogoče smiselno interpretirati.

Tabela 66: Faktorske uteži za večdimenzionalno spremenljivko odnos do sprejema mobilnih oglasov

	Component
	1
V2.1 Zdi se mi da je ideja o uporabljanju mobilnih oglasov dobra	0,672
V2.2 Imam pozitiven odnos do sprejema mobilnih oglasov	0,810
V2.3 Za mobilne oglase mi je vseeno (prekodirana)	0,672
V2.4 Mobilno oglaševanje je zanimivo	0,825

Tabela 66 prikazuje, da so vse faktorske uteži, a_{ij} , višje od 0,60. Kvadrat faktorske uteži pri i -ti spremenljivki in j -tem faktorju označuje delež pojasnjene variance i -te spremenljivke z j -tim faktorjem (ibid) (delež pojasnjene variance i -te spremenljivke z j -tim faktorjem je prikazan v tabeli 64). Pri *odnosu do sprejema mobilnih oglasov* ima tako najpomembnejšo vlogo spremenljivka *Mobilno oglaševanje je zanimivo*, pri kateri je vrednost faktorske uteži najvišja.

Naloga 2.

Na osnovi podatkov v datoteki *Faktorska analiza_Mobilno oglaševanje.sav* izvedite faktorsko analizo za večdimenzionalno spremenljivko *zaznana uporabnost mobilnih oglasov* in pojasnite rezultate.

Odgovori in izpisi rezultatov:

Tabela 67: Kaiser-Meyer-Olkinova statistika in Bartlettov test sferičnosti za večdimenzionalno spremenljivko zaznana uporabnost mobilnih oglasov

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0,786
Bartlett's Test of Sphericity	Approx. Chi-Square	154,727
	df	28
	Sig.	0,000

V tabeli 67 KMO (0,786) in stopnja značilnosti Bartlettovega testa ($p < 0,001$) kažeta na smiselnost uporabe faktorske analize.

Tabela 68: Komunalitete za večdimenzionalno spremenljivko zaznana uporabnost mobilnih oglasov

	Initial	Extraction
V1.1 Uporaba mobilnega oglasa mi pomaga pri odločitvi nakupa dodatnih storitev/izdelka	1,000	0,645
V1.2 Skozi mobilno oglaševanje dobim točne informacije	1,000	0,525
V1.3 Mobilno oglaševanje varuje naravno okolje (npr. zmanjševanje tiskovin)	1,000	0,907
V1.4 Uporaba mobilnih oglasov se mi zdi koristna	1,000	0,783
V1.5 Mobilni oglasi me obveščajo o novostih	1,000	0,775
V1.6 Preko mobilnih oglasov lahko dobim koristne informacije	1,000	0,821
V1.7 Preko mobilnih oglasov dobim boljši opis oglaševanih izdelkov	1,000	0,681
V1.8 Mobilno oglaševanje je prilagojeno mojim potrebam	1,000	0,332

Vrednosti vseh komunalitet v tabeli 68 za večdimenzionalno spremenljivko *zaznana uporabnost mobilnih oglasov* so višje od 0,40, razen pri spremenljivki *V1.8 Mobilno oglaševanje je prilagojeno mojim potrebam* je vrednost komunalitete 0,332, zato smo navedeno spremenljivko izločili.

Zaradi izločitve spremenljivke *V1.8 Mobilno oglaševanje je prilagojeno mojim potrebam* ponovno opravimo preizkušanje glede ustreznosti faktorske analize:

Tabela 69: Kaiser-Meyer-Olkinova statistika in Bartlettov test sferičnosti brez spremenljivke V1.8 Mobilno oglaševanje je prilagojeno mojim potrebam

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0,781
Bartlett's Test of Sphericity	Approx. Chi-Square	143,985
	df	21
	Sig.	0,000

KMO (0,781) in stopnja značilnosti Bartlettovega testa ($p < 0,001$) kažeta na smiselnost uporabe faktorske analize (tabela 69).

Tabela 70: Komunalitete brez spremenljivke V1.8 Mobilno oglaševanje je prilagojeno mojim potrebam

	Initial	Extraction
V1.1 Uporaba mobilnega oglasa mi pomaga pri odločitvi nakupa dodatnih storitev/izdelka	1,000	0,654
V1.2 Skozi mobilno oglaševanje dobim točne informacije	1,000	0,511
V1.3 Mobilno oglaševanje varuje naravno okolje (npr. zmanjševanje tiskovin)	1,000	0,899
V1.4 Uporaba mobilnih oglasov se mi zdi koristna	1,000	0,813
V1.5 Mobilni oglasi me obveščajo o novostih	1,000	0,808
V1.6 Preko mobilnih oglasov vlažje dobim koristne informacije	1,000	0,821
V1.7 Preko mobilnih oglasov dobim boljši opis oglaševanih izdelkov	1,000	0,690

Vrednosti vseh komunalitet v tabeli 70 za večdimenzionalno spremenljivko *zaznana uporabnost mobilnih oglasov* so višje od 0,40, zato nismo dodatno izločili nobene spremenljivke.

Tabela 71: Celotna pojasnjena varianca za večdimenzionalno spremenljivko zaznana uporabnost mobilnih oglasov

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,068	58,120	58,120	4,068	58,120	58,120	3,591	51,297	51,297
2	1,127	16,106	74,226	1,127	16,106	74,226	1,605	22,929	74,226
3	0,700	9,998	84,224						
4	0,444	6,349	90,573						
5	0,329	4,694	95,267						
6	0,195	2,787	98,054						
7	0,136	1,946	100,000						

Iz tabele 71 je razvidno, da se iz sedmih osnovnih spremenljivk oblikujeta dva faktorja, pri katerih je lastna vrednosti večja od 1 in ki skupaj pojasnjujeta 74,23 % variance vseh osnovnih sedmih spremenljivk skupaj, in sicer prvi faktor 58,12 %, drugi faktor 16,11 %.

Tabela 72: Rotirane faktorske uteži za večdimenzionalno spremenljivko zaznana uporabnost mobilnih oglasov

	Component	
	1	2
V1.1 Uporaba mobilnega oglasa mi pomaga pri odločitvi nakupa dodatnih storitev/izdelka	0,798	0,130
V1.2 Skozi mobilno oglaševanje dobim točne informacije	0,712	0,065
V1.3 Mobilno oglaševanje varuje naravno okolje (npr. zmanjševanje tiskovin)	-0,030	0,948
V1.4 Uporaba mobilnih oglasov se mi zdi koristna	0,709	0,557
V1.5 Mobilni oglasi me obveščajo o novostih	0,674	0,595
V1.6 Preko mobilnih oglasov vlažje dobim koristne informacije	0,897	0,130
V1.7 Preko mobilnih oglasov dobim boljši opis oglaševanih izdelkov	0,828	0,061

Za izboljšanje strukture oblikovanih faktorjev smo vrednosti uteži rotirali z metodo Varimax. Oba faktorja se poimenujeta glede na spremenljivke, ki so v vsakem faktorju vključene in ga zastopajo. Tako bi lahko prvi faktor poimenovali *zaznana uporabnost – splošno*, drugi faktor pa *zaznana uporabnost in varovanje okolja* (tabela 72).

Naloga 3.

V podjetju so želeli izmeriti zadovoljstvo svojih zaposlenih. Zaposleni so na 5-stopenjski Likertovi lestvici izrazili stopnjo strinjanja (kjer pomeni 1-sploš se ne strinjam in 5-popolnoma se strinjam) z navedenimi trditvami o zadovoljstvu na delovnem mestu. Podatki so v datoteki *Faktorska analiza_Zadovoljstvo zaposlenih.sav*

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

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

Odgovori in izpisi rezultatov:

Tabela 73: Kaiser-Meyer-Olkinova statistika in Bartlettov test sferičnosti za spremenljivko zadovoljstvo zaposlenih

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0,958
Bartlett's Test of Sphericity	Approx. Chi-Square	16180,237
	df	66
	Sig.	0,000

Tabela 74: Komunalitete za spremenljivko zadovoljstvo zaposlenih

	Initial	Extraction
V podjetju sem zadovoljen/na z delovnim časom in razporeditvijo delovnih obveznosti.	1,000	0,813
V podjetju sem zadovoljen/na z organiziranjem delovnih nalog po meri starejših.	1,000	0,821
V podjetju sem zadovoljen/na z omogočanjem fleksibilnega delovnega časa.	1,000	0,842
V podjetju sem zadovoljen/na z ravnovesjem med zasebnim in poklicnim življenjem.	1,000	0,553
V podjetju sem zadovoljen/na z omogočanjem samouravnavanja hitrosti opravljenega dela.	1,000	0,833
V podjetju sem zadovoljen/na s količino programov v okviru aktivnega staranja in zdravega načina življenja.	1,000	0,784
V podjetju sem zadovoljen/na z medgeneracijskim sodelovanjem ter s tem porazdelitev dela.	1,000	0,840
V podjetju sem zadovoljen/na z delovnimi pogoji - kot so boljša svetloba, klima, večji napisi.	1,000	0,599
V podjetju sem zadovoljen/na z omogočanjem prožnega delovnika.	1,000	0,714
V podjetju sem zadovoljen/na z omogočanjem delitve delovnega mesta in s tem zmanjšanje obremenjenosti na delovnem mestu.	1,000	0,839
Zadovoljen/na sem z medsebojnimi odnosi.	1,000	0,752
Zadovoljen/na sem z vodstvom v podjetju.	1,000	0,831

Iz dvanajstih merjenih spremenljivk smo v postopku faktorске analize izločili en faktor, ki pojasnjuje 76,8 % variabilnosti osnovnih dvanajstih spremenljivk (tabela 74).

Tabela 75: Celotna pojasnjena varianca za spremenljivko zadovoljstvo zaposlenih

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	9,221	76,844	76,844	9,221	76,844	76,844
2	0,711	5,924	82,768			
3	0,481	4,005	86,773			
4	0,280	2,336	89,109			
5	0,249	2,076	91,185			
6	0,231	1,922	93,107			
7	0,197	1,642	94,748			
8	0,157	1,305	96,053			
9	0,129	1,073	97,126			
10	0,125	1,044	98,170			
11	0,115	0,955	99,126			
12	0,105	0,874	100,000			

Tabela 76: Faktorske uteži za spremenljivko zadovoljstvo zaposlenih

	Component
	1
V podjetju sem zadovoljen/na z delovnim časom in razporeditvijo delovnih obveznosti.	0,902
V podjetju sem zadovoljen/na z organiziranjem delovnih nalog po meri starejših.	0,906
V podjetju sem zadovoljen/na z omogočanjem fleksibilnega delovnega časa.	0,918
V podjetju sem zadovoljen/na z ravnovesjem med zasebnim in poklicnim življenjem.	0,744
V podjetju sem zadovoljen/na z omogočanjem samouravnavanja hitrosti opravljenega dela.	0,913
V podjetju sem zadovoljen/na s količino programov v okviru aktivnega staranja in zdravega načina življenja.	0,886
V podjetju sem zadovoljen/na z medgeneracijskim sodelovanjem ter s tem porazdelitev dela.	0,917
V podjetju sem zadovoljen/na z delovnimi pogoji - kot so boljša svetloba, klima, večji napisi.	0,774
V podjetju sem zadovoljen/na z omogočanjem prožnega delovnika.	0,845
V podjetju sem zadovoljen/na z omogočanjem delitve delovnega mesta in s tem zmanjšanje obremenjenosti na delovnem mestu.	0,916
Zadovoljen/na sem z medsebojnimi odnosi.	0,867
Zadovoljen/na sem z vodstvom v podjetju.	0,911

Tabela 77: Cronbachova alfa za faktor zadovoljstvo zaposlenih

Cronbach's Alpha	N of Items
0,971	12

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 $\alpha \geq 0,80$, označimo kot zgledno, če je koeficient na intervalu $0,70 \leq \alpha < 0,80$, kot zelo dobro, na intervalu $0,60 \leq \alpha < 0,70$ kot zmerno, če je koeficient α manjši od 0,60, pa kot komaj sprejemljivo (Tabachnick in Fidell, 2013).

Tabela 77 kaže, da je vrednost Cronbachove alfe zgledna (0,971), zato lahko trdimo, da je faktor *zadovoljstvo zaposlenih* zgledne zanesljivosti.

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 76.

Tabela 78: 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	Slaba korelacija
0,51–0,79	0,26–0,64	Srednje močna korelacija
0,80–0,99	0,65–0,99	Močna korelacija
1	1	Popolna korelacija

Vir: Artenjak, 2003

- Standardna napaka ocene odvisne spremenljivke (σ_y): pokaže, ali na variabilnost spremenljivke y , razen spremenljivke x_1 , vplivajo še druge spremenljivke in slučajni vplivi (Artenjak, 2003).

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

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

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

Statistično značilen vpliv pojasnjevalne spremenljivke x_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 1.

Podjetje želi ugotoviti ali *vodenje v podjetju* (neodvisna spremenljivka) vpliva na *delovno zavzetost njihovih zaposlenih* (odvisna spremenljivka). Zaposleni so na 5-stopenjski Likertovi lestvici označili stopnjo strinjanja (1 – sploh se ne strinjam, 2 – ne strinjam se, 3 – delno se strinjam, 4 – strinjam se, 5 – popolnoma se strinjam) z navedenimi trditvami, ki so se nanašala na vodenje v podjetju in njihovo delovno zavzetost (trditve so navedene v datoteki *Regresijska analiza_zaposleni.sav*). Predhodno je bila izvedena faktorska analiza, s katero smo izločili en faktor – *delovna zavzetost zaposlenih* (y) in en faktor – *vodenje* (x_1).

Odprite datoteko *Regresijska analiza_zaposleni.sav*. Preverite rezultate faktorke analize za oba dobljena faktorja (za *delovno zavzetost zaposlenih* in za *vodenje*) in jih vsebinsko pojasnite. Vrednosti obeh dobljenih faktorjev shranite v datoteko s podatki.

Nato izvedite enostavno linearno regresijo in pojasnite rezultate.

Postopek – izvedba faktorke analize: Kliknemo *Analyze*, nato *Dimension Reduction* in *Factor*. Odpre se pogovorno okno in prenesemo spremenljivke, ki sodijo pod večdimenzionalno spremenljivko *delovna zavzetost zaposlenih*. Da se bodo vrednosti dobljenih faktorjev shranile v datoteko s podatki, kliknemo *Scores* in obkljukamo *Save as variables* ter pod *Method* označimo *Regression*. Enak postopek izvedemo še za večdimenzionalno spremenljivko *vodenje*.

Postopek – izvedba enostavne linearne regresije: Kliknemo na *Analyze*, nato *Regression* in izberemo *Linear*. Odpre se pogovorno okno, v katerem spremenljivko *delovna zavzetost zaposlenih* prenesemo v desno okence, v polje *Dependent*, spremenljivko *vodenje* prenesemo v desno okence, v polje *Independent*. Kliknemo na *Method* in izberemo *Enter*. Kliknemo *Statistics* in v okencu *Regression Coefficient* izberemo *Estimates* ter *Model Fit*. Nato kliknemo *Continue* in *OK*.

Odgovori in izpisi rezultatov:

Tabela 79: Rezultat - korelacijski in determinacijski koeficient

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0,957 ^a	0,916	0,915	0,29245091

a. Predictors: (Constant), Vodenje

Vrednost korelacijskega koeficienta znaša 0,957, kar kaže na to, da med spremenljivkama *vodenje* in *delovna zavzetost zaposlenih* obstaja močna linerna povezanost. Vrednost determinacijskega koeficienta znaša 0,916. Determinacijski koeficient pojasnjuje, da smo 91,6 % celotne variance delovne zavzetosti zaposlenih uspeli pojasniti z variabilnostjo neodvisne spremenljivke (*vodenje*) (tabela 79).

Tabela 80: Rezultat – F-test (ANOVA)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	117,798	1	117,798	1377,314	0,000 ^b
	Residual	10,862	127	0,086		
	Total	128,660	128			

a. Dependent Variable: Delovna zavzetost zaposlenih

b. Predictors: (Constant), Vodenje

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 neodvisno spremenljivko - vodenje in odvisno spremenljivko - delovna zavzetost zaposlenih (ničelno domnevo: $H_0: r^2_{xy} = 0$ zavrnemo in sprejmemo raziskovalno domnevo: $H_1: r^2_{xy} > 0$) (tabela 80).

Tabela 81: Rezultat – regresijski koeficienti in t-test

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-0,005	0,026		-0,175	0,862
	Vodenje	0,959	0,026	0,957	37,112	0,000

a. Dependent Variable: Delovna zavzetost zaposlenih

Vrednost t-testa in stopnja značilnosti ($p < 0,001$) kažeta, da je regresijski koeficient β_1 različen od nič (potrdimo ničelno domnevo: $H_1: \beta_1 \neq 0$), kar pomeni, da pojasnjevalna spremenljivka *vodenje* statistično značilno vpliva na *delovno zavzetost zaposlenih* (kar sicer kaže tudi rezultat pri preverjanju kakovosti regresijskega modela, saj je v model vključena le ta ena pojasnjevalna spremenljivka), regresijska konstanta pa ni statistično značilno različna od 0 (tabela 81).

Naloga 2.

Zanima nas ali je vrednost prodaje nekega trgovskega podjetja odvisna od števila izvedenih propagandnih akcij (vir podatkov: Bastič in drugi, 2008). Izvedite enostavno linearno regresijsko analizo in vsebinsko pojasnite dobljene rezultate. Uporabite podatke *Regresijska analiza_podatki regresija.sav*

Odgovori in izpisi rezultatov:

Tabela 82: Rezultat - korelacijski in determinacijski koeficient

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0,880 ^a	0,775	0,769	595,602

a. Predictors: (Constant), propaganda

Tabela 83: Rezultat – F-test (ANOVA)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	46345100,888	1	46345100,888	130,644	0,000 ^b
	Residual	13480201,487	38	354742,144		
	Total	59825302,375	39			

a. Dependent Variable: prodaja

b. Predictors: (Constant), propaganda

Tabela 84: Rezultat – regresijski koeficienti in t-test

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1354,336	259,065		5,228	0,000
	propaganda	253,077	22,142	0,880	11,430	0,000

a. Dependent Variable: prodaja

5.2 Multipla regresijska analiza

Multiplo regresijsko uporabimo takrat, kadar na odvisno spremenljivko (y) vpliva več neodvisnih spremenljivk x_i ($i = 1, 2, \dots, k$) (in so izpolnjeni pogoji za izvedbo multiple regresijske analize) (Tabachnick & Fidell, 2013).

Multipli korelacijski koeficient R kaže jakost povezanosti med odvisno in k neodvisnimi spremenljivkami in je prikazan v absolutni vrednosti. Prilagojeni determinacijski koeficient R^2 pa predstavlja delež variance odvisne spremenljivke, ki je pojasnjena z variabilnostjo v model vključenih neodvisnih spremenljivk (ibid).

Naloga 1.

Uporabite podatke iz naloge 1 pri poglavju Enostavna regresija (datoteka: *Regresijska analiza_zaposleni.sav*) in ugotovite ali je *delovna zavzetost zaposlenih* odvisna od njihovega *zadovoljstva* in *motiviranosti* na delovnem mestu.

- Izvedite faktorsko analizo in shranite vrednosti dobljenega faktorja (enofaktorske rešitve), za izvedbo regresijske analize.
- Pojasnite multipli korelacijski koeficient in prilagojeni determinacijski koeficient.
- Preverite ustreznost modela kot celote in zapišite ustrezno postavljeno domnevo. Pojasnite rezultate.
- Zapišite ustrezno postavljene domneve za preverjanje statistične značilnosti regresijskih koeficientov v regresijski funkciji ter pojasnite rezultate.
- Zapišite enačbo regresijske funkcije z ocenjenimi regresijskimi koeficienti.

Odgovori in izpisi rezultatov:

Tabela 85: Rezultat – multipli korelacijski in prilagojeni determinacijski koeficient

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0,970 ^a	0,942	0,941	0,24394278

a. Predictors: (Constant), Motiviranost zaposlenih, Zadovoljstvo zaposlenih

Vrednost multiplega korelacijskega koeficienta znaša 0,970, kar kaže na to, da med odvisno spremenljivko *delovna zavzetost zaposlenih* in neodvisnima spremenljivkama *motiviranost zaposlenih* in *zadovoljstvo zaposlenih* obstaja močna povezanost. Vrednost prilagojenega multiplega determinacijskega koeficienta znaša 0,941. Prilagojen determinacijski koeficient pojasnjuje, da je z regresijskim modelom (s spremenljivkama *motiviranost zaposlenih* (x_1) in *zadovoljstvo zaposlenih* (x_2)), pojasnjene 94,2 % celotne variance *delovne zavzetosti zaposlenih* (tabela 85).

Tabela 86: Rezultat – F-test (ANOVA)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	121,337	2	60,669	1019,503	0,000 ^b
	Residual	7,498	126	0,060		
	Total	128,835	128			

a. Dependent Variable: Delovna zavzetost zaposlenih

b. Predictors: (Constant), Motiviranost zaposlenih, Zadovoljstvo zaposlenih

Kakovost regresijskega modela kot celote smo preverjali z F-testom, pri čemer lahko glede na vrednost p ($p < 0,001$) trdimo, da je model kakovosten. Navedeno pomeni, da obstaja odvisnost med odvisno spremenljivko (*delovna zavzetost zaposlenih*) in vsaj eno neodvisno spremenljivko. Z F-testom preizkušamo domnevo:

H0: Prilagojeni determinacijski koeficient je enak 0 ($R^2 = 0$)

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

Na osnovi rezultatov lahko zavrremo ničelno domnevo, da je $R^2 = 0$ (tabela 86).

Tabela 87: Rezultat – regresijski koeficienti in t-test

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-0,001	0,021		-0,026	0,979
	Zadovoljstvo zaposlenih	0,741	0,078	0,741	9,537	0,000
	Motiviranost zaposlenih	0,237	0,078	0,237	3,047	0,003

a. Dependent Variable: Delovna zavzetost zaposlenih

Pri tabeli 87 za preverjanje statistične značilnosti regresijskih koeficientov v regresijski funkciji napišemo naslednje domneve:

H0: $\beta_i = 0$

H1: $\beta_i \neq 0$

($i = 1,2$).

Podobno lahko preverimo tudi statistično značilnost regresijske konstante.

T-test in stopnja značilnosti za oba regresijska koeficienta ($p < 0,001$ za prvi regresijski koeficient, oziroma $p = 0,003$ za drugega kažeta, da obe pojasnjevalni spremenljivki statistično značilno vplivata na odvisno spremenljivko, regresijska konstanta pa ni statistično značilno različna od 0. Dobljena enačba regresijske funkcije, z ocenjenimi vrednostmi regresijskih koeficientov na osnovi uporabljenih vzorčnih podatkov, se glasi (tabela 87):

$\hat{y} = -0,001 + 0,741 x_1 + 0,237 x_2$ (kjer pomeni x_1 zadovoljstvo zaposlenih in x_2 motiviranost zaposlenih).

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 (ne obstaja multikolineranost med pojasnjevalnima spremenljivkama).

Naloga 2.

Zanima nas ali je vrednost prodaje nekega trgovskega podjetja odvisna od števila izvedenih propagandnih akcij (x_1) ter od števila trgovskih potnikov (x_2) (vir podatkov: Bastič in drugi, 2008). Izvedite multiplo linearno regresijsko analizo in vsebinsko pojasnite dobljene rezultate. Uporabite podatke *Regresijska analiza_podatki regresija.sav*.

- Pojasnite multipli korelacijski koeficient in prilagojeni determinacijski koeficient.
- Preverite ustreznost modela kot celote in zapišite ustrezno postavljeno domnevo. Pojasnite rezultate.
- Zapišite ustrezno postavljene domneve za preverjanje statistične značilnosti regresijskih koeficientov v regresijski funkciji ter pojasnite rezultate.
- Zapišite in pojasnite enačbo regresijske funkcije.
- Z ustreznimi kazalci preverite morebitno prisotnost multikolineranosti. Vsebinsko pojasnite rezultate.

Odgovori in izpisi rezultatov:

Tabela 88: Rezultat – multipli korelacijski in prilagojeni determinacijski koeficient

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0,935 ^a	0,874	0,867	451,647

a. Predictors: (Constant), število trgovskih potnikov, propaganda

Tabela 89: Rezultat – F-test (ANOVA)

Model	Sum of Squares	df	Mean Square	F	Sig.	
1	Regression	52277845,899	2	26138922,949	128,141	0,000 ^b
	Residual	7547456,476	37	203985,310		
	Total	59825302,375	39			

a. Dependent Variable: prodaja

b. Predictors: (Constant), število trgovskih potnikov, propaganda

Tabela 90: Rezultat – regresijski koeficienti, t-test in kazalnika kolinearnosti

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	693,285	231,555		2,994	0,005		
	propaganda	141,562	26,636	0,492	5,315	0,000	0,397	2,517
	Št. trg. potnikov	375,313	69,593	0,500	5,393	0,000	0,397	2,517

a. Dependent Variable: prodaja

Enačba regresijske funkcije (tabela 90): $\hat{y} = 693,285 + 141,562x_1 + 375,313x_2$ (kjer pomeni x_1 število izvedenih propagandnih akcij ter x_2 število trgovskih potnikov).

Ocenjeni regresijski koeficient pri x_1 pove, da se vrednost prodaje v povprečju poveča za 141,562 d.e., če se število propagandnih akcij poveča za 1 enoto in pri tem vrednost spremenljivke število trgovskih potnikov ostane nespremenjena.

Kazalnika Tolerance in VIF kažeta, da med spremenljivkama ne obstaja pomembna stopnja odvisnosti (multikolineranost ni prisotna), saj je ocenjeno, da je multikolineranost prisotna, če je $VIF > 10$ (Kutner, Nachtsheim, & Neter, 2004).

6 DISKRIMINANTNA ANALIZA

Diskriminatno analizo uporabljamo, kadar želimo oceniti pripadnost statistične enote določeni skupini statističnih enot (skupine se medsebojno izključujejo), glede na vrednosti statističnih neodvisnih spremenljivk (številске spremenljivke, t.i. discriminating variables), oziroma nas zanima, ali je mogoče statistične enote glede na vrednosti teh neodvisnih spremenljivk, razvrstiti v dve vnaprej znani skupini - diskriminantna analiza z dvema skupinama, ali v več skupin – multipla diskriminantna analiza (in so izpolnjeni pogoji za izvedbo diskriminatne analize) (Tabachnick & Fidell, 2013).

Kazalniki v okviru diskriminantne analize z dvema skupinama (ibid):

- **Lastna vrednost diskriminantne funkcije:** delež pojasnjene variance (vsota kvadratov odklonov med skupinami, je deljena z vsoto kvadratov odklonov znotraj skupin). Večja lastna vrednost kaže na večjo kakovost diskriminantne funkcije.
- **Koeficient kanonične korelacije:** korelacijski koeficient med ocenjenimi vrednostmi in dejanskimi vrednostmi odvisne spremenljivke. Vrednost blizu 1 označuje večjo kakovost diskriminantne funkcije,
- **Wilk's lambda in Hi-kvadrat test:** Če je vrednost kazalnika Wilk's lambda blizu 1, to kaže na to, da sta povprečni vrednosti diskriminantne funkcije v obeh skupinah enaki (kar pomeni slaba kakovost diskriminantne funkcije). Vrednost blizu 0 pa kaže na to, da je variabilnost znotraj obeh skupin majhna v primerjavi s skupno variabilnostjo spremenljivke (oziroma dobra kakovost diskriminantne funkcije). S Hi-kvadrat testom preverjamo domnevo, da sta povprečni vrednosti diskriminantne funkcije v obeh skupinah enaki. Če domnevo lahko zavrnemo (pri nizkem tveganju) to kaže na ustrezno kakovost diskriminantne funkcije.

- **Centroid:** je povprečna vrednost diskriminantne funkcije za statistične enote v pripadajoči skupini.
- **Klasifikacijska matrika:** kaže število in odstotek pravilno in nepravilno razvrščenih statističnih enot.

Naloga 1.

Razpolagamo s podatki slučajnega vzorca izvozno usmerjenih podjetij v neki gospodarski panogi (datoteka *diskriminantna analiza.sav*). Podjetja smo razdelili v dve skupini (spremenljivka *EU*) – na tista, pri katerih je pretežni del izvoza ustvarjen na trgih Evropske unije ($EU = 1$) in na tista, ki pretežni del izvoza ustvarijo na globalnih trgih izven Evropske unije ($EU = 0$).

Zanima nas, ali so naslednje spremenljivke tiste, ki so pomembne za razvrstitev podjetij v ti dve skupini:

- Kolikšen delež izdelkov na tujih trgih prodate pod svojo blagovno znamko? (spremenljivka *vp2*: podjetja so podala delež, od več kot 0 do 100 %).

Pri naslednjih spremenljivkah pa so podjetja podala strinjanje s trditvami glede tržnih in tehnoloških sprememb na trgu, kjer podjetje deluje (na lestvici od 1 – nikakor se ne strinjam, do 7 – povsem se strinjam), spremenljivke *vp8a*, *vp8b*, *vp8g*, *vp8h*, *vp8i*:

- Spremembe v tehnologiji na tem trgu so hitre.
- Spremembe v tehnologiji na tem trgu predstavljajo velike priložnosti.
- Odjemalci na tem trgu so zelo dovzetni za nove izdelke (storitve).
- Novi odjemalci imajo glede izdelkov (storitev) potrebe, ki so drugačne od potreb obstoječih odjemalcev.
- Poslujemo na trgu, na katerem se želje odjemalcev zelo počasi spreminjajo.

Z diskriminantno analizo preverite, ali so zgoraj opisane neodvisne spremenljivke pomembne za razvrščanje podjetij v dve skupini - na tista, pri katerih je pretežni del izvoza ustvarjen na trgih Evropske unije ($EU = 1$) in na tista, ki pretežni del izvoza ustvarijo na globalnih trgih izven Evropske unije ($EU = 0$).

Postopek – izvedba diskriminantne analize: Kliknemo *Analyze*, nato *Classify* in *Discriminant*. Odpre se pogovorno okno, kjer v polje *Grouping variable* prenesemo spremenljivko *EU*, pod gumbom *Define Range* vnesemo vrednosti 0 in 1. V polje *Independents* prenesemo vseh 6 neodvisnih spremenljivk. Izberemo gumbek *Statistics* in izberemo v polje *Descriptive* gumbek *Means*. Za grafični prikaz centroidov izberemo gumbek *Classify* in v okvirju *Plots* izberemo polje *Separate-Groups*. Za prikaz *Klasifikacijske tabele* oziroma matrike izberemo v okvirju *Display* polje *Summary table*. Stisnemo gumb *OK*.

Odgovori in izpisi rezultatov:

Tabela 91: Izpis rezultata – Lastna vrednost diskriminantne funkcije (angl. Eigenvalues)

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	2,729 ^a	100,0	100,0	0,855

a. First 1 canonical discriminant functions were used in the analysis.

Lastna vrednost 2,729, ki je visoka ter vrednost koeficienta kanonične korelacije, ki je 0,855 nakazujeta, da je kakovost diskriminantne funkcije visoka ter da so vključene neodvisne spremenljivke pomembne za razlikovanje podjetij v obeh skupinah (tabela 91).

To potrjuje tudi vrednost Wilks' lambda ter hi-kvadrat test, s katerim preverjamo trditev, da je povprečna vrednost diskriminantne funkcije v obeh skupinah enaka (kar bi pomenilo, da vključene neodvisne spremenljivke niso pomembne za razlikovanje med skupinama). Z manj kot 5 % tveganjem lahko zavrnamo domnevo, da je povprečna vrednost diskriminantne funkcije v obeh skupinah enaka in potrdimo kakovost diskriminantne funkcije (tabela 92).

Tabela 92: Wilks' Lambda in hi-kvadrat test

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	0,268	35,533	6	0,000

Zapišemo lahko tudi enačbo diskriminantne funkcije (tabela 93):

$$D = -0.829 \text{ } vp2 + 0,014 \text{ } vp8a + 0,507 \text{ } vp8b + 0,446 \text{ } vp8g - 0,611 \text{ } vp8h + 0,597 \text{ } vp8i$$

K razlikovanju med skupinama tako največ pripomore spremenljivka *vp2*, rezultati v tabeli 94 deskriptivne statistike pa nam tudi povedo, da je povprečna vrednost te spremenljivke v skupini podjetij, ki pretežni del izvoza opravijo na trgih EU, v povprečju enaka 32,4 %, v skupini podjetij, ki pretežni del izvoza opravijo na globalnih trgih izven EU, pa je v povprečju enaka 63,9 %.

V tabeli 95 - *Classification results* vidimo, da je 93.8 % pravilno razvrščenih statističnih enot (izvozno usmerjena podjetja), kar tudi potrjuje kakovost diskriminantne funkcije.

Tabela 93: Koeficienti diskriminantne funkcije (angl. Standardized Canonical Discriminant - Function Coefficients)

	Function
	1
Kolikšen delež izdelkov na tujih trgih prodate pod svojo blagovno znamko? (<i>vp2</i>)	-0,829
Spremembe v tehnologiji na tem trgu so hitre. (<i>vp8a</i>)	0,014
Spremembe v tehnologiji na tem trgu predstavljajo velike priložnosti. (<i>vp8b</i>)	0,507
Odjemalci na tem trgu so zelo dovzetni za nove izdelke (storitve). (<i>vp8g</i>)	0,446
Novi odjemalci imajo glede izdelkov (storitev) potrebe, ki so drugačne od potreb obstoječih odjemalcev. (<i>vp8h</i>)	-0,611
Poslujemo na trgu, na katerem se želje odjemalcev zelo počasi spreminjajo. (<i>vp8i</i>)	0,597

Tabela 94: Opisna statistika

Evropska unija		Mean	Std. Deviation	Valid N (listwise)	
				Unweighted	Weighted
ne	Kolikšen delež izdelkov na tujih trgih prodate pod svojo blagovno znamko?	63,8571	17,28431	14	14,000
	Spremembe v tehnologiji na tem trgu so hitre.	4,9286	1,81720	14	14,000
	Spremembe v tehnologiji na tem trgu predstavljajo velike priložnosti.	4,2143	1,80506	14	14,000
	Odjemalci na tem trgu so zelo dovzetni za nove izdelke (storitve).	4,5000	1,74312	14	14,000
	Novi odjemalci imajo glede izdelkov (storitev) potrebe, ki so drugačne od potreb obstoječih odjemalcev.	4,2857	1,20439	14	14,000
	Poslujemo na trgu, na katerem se želje odjemalcev zelo počasi spreminjajo.	2,5000	0,75955	14	14,000
da	Kolikšen delež izdelkov na tujih trgih prodate pod svojo blagovno znamko?	32,4278	13,37468	18	18,000
	Spremembe v tehnologiji na tem trgu so hitre.	4,5556	1,78958	18	18,000
	Spremembe v tehnologiji na tem trgu predstavljajo velike priložnosti.	4,7778	1,51679	18	18,000
	Odjemalci na tem trgu so zelo dovzetni za nove izdelke (storitve).	4,3333	1,41421	18	18,000
	Novi odjemalci imajo glede izdelkov (storitev) potrebe, ki so drugačne od potreb obstoječih odjemalcev.	3,2222	1,35280	18	18,000
	Poslujemo na trgu, na katerem se želje odjemalcev zelo počasi spreminjajo.	4,4444	1,61690	18	18,000
To- tal	Kolikšen delež izdelkov na tujih trgih prodate pod svojo blagovno znamko?	46,1781	21,77871	32	32,000
	Spremembe v tehnologiji na tem trgu so hitre.	4,7188	1,78225	32	32,000
	Spremembe v tehnologiji na tem trgu predstavljajo velike priložnosti.	4,5313	1,64580	32	32,000
	Odjemalci na tem trgu so zelo dovzetni za nove izdelke (storitve).	4,4063	1,54208	32	32,000
	Novi odjemalci imajo glede izdelkov (storitev) potrebe, ki so drugačne od potreb obstoječih odjemalcev.	3,6875	1,37811	32	32,000
	Poslujemo na trgu, na katerem se želje odjemalcev zelo počasi spreminjajo.	3,5938	1,62360	32	32,000

Tabela 95: Klasifikacijska tabela (angl. Classification Results)

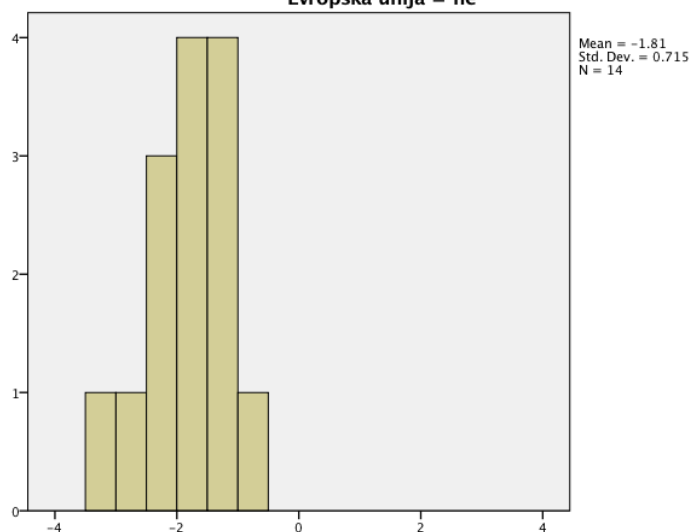
		Evropska unija	Predicted Group Membership		Total
			ne	da	
Original	Count	ne	14	0	14
		da	2	16	18
	%	ne	100,0	0,0	100,0
		da	11,1	88,9	100,0

a. 93.8 % of original grouped cases correctly classified.

Na spodnjih dveh grafih je prikazana frekvenčna porazdelitev vrednosti diskriminantne funkcije v obeh skupinah, v zadnji tabeli pa še oba centroida – povprečni vrednosti diskriminantne funkcije v obeh skupinah.

Canonical Discriminant Function 1

Evropska unija = ne



Canonical Discriminant Function 1

Evropska unija = da

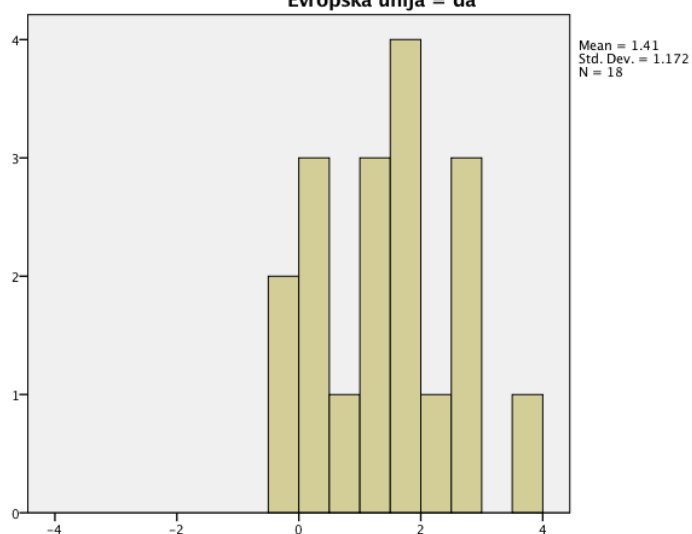


Tabela 96: Centroida (angl. Functions at Group Centroids)

	Function
Evropska unija	1
ne	-1,814
da	1,411
Unstandardized canonical discriminant functions evaluated at group means	

LITERATURA IN VIRI

- Agresti, A., Finlay, B. (2009). *Statistical methods for the Social sciences*. Pearson: Prentice Hall.
- Artenjak, J. (2003), *Poslovna statistika*, Prenovljena in dopolnjena izdaja. Maribor: UM Ekonomsko-poslovna fakulteta.
- Bastič, M. (2006). *Metode raziskovanja*. Maribor: UM Ekonomsko-poslovna fakulteta.
- Bastič, M., Leskovar-Špacapan, G., Čančer, V. (2008). *Gradivo za vaje pri predmetu Metode raziskovanja bolonjskega magistrskega študijskega programa »Ekonomske in poslovne vede« na Ekonomsko-poslovni fakulteti Univerze v Mariboru*. Maribor: UM EPF.
- Corder, G. W., Foreman, D. I. (2014). *Nonparametric statistics for non-statisticians: A step-by-step approach*. 2. izdaja. New Jersey, CA: Wiley.
- Goos, P., Meintrup, D. (2015). *Statistics with JMP: Graphs, Descriptive Statistics and Probability*. New Jersey, CA: Wiley.
- Kožuh, B. (2005). *Statistika II*. Ljubljana: Filozofska fakulteta.
- Kutner, M. H., Nachtsheim, C. J., Neter, J. (2004). *Applied Linear Regression Models* (4th ed.). McGraw-Hill Irwin.
- Tabachnick, B.G., Fidell, L.S. (2013). *Using multivariate statistics*. Pearson: Boston, Columbus etc.
- Tominc, P., Kramberger, T. (2007). *Statistične metode v logistiki*. Celje: UM Fakulteta za logistiko.



Univerza v Mariboru

Ekonomsko-poslovna fakulteta

