Univerzitetna založba
Univerze v Mariboru

POSLOVNA STATISTIKA

Polona
TOMINC

Maja
ROŽMAN

Univerza v Mariboru

Ekonomsko-poslovna fakulteta

POSLOVNA STATISTIKA

Avtorici

Polona Tominc

Maja Rožman

Maj 2023

Naslov *Title*	**Poslovna statistika** *Business Statistics*
Avtorici *Authors*	Polona Tominc (Univerza v Mariboru, Ekonomsko-poslovna fakulteta) Maja Rožman (Univerza v Mariboru, Ekonomsko-poslovna fakulteta)
Recenzija *Review*	Vesna Čančer (Univerza v Mariboru, Ekonomsko-poslovna fakulteta) Dijana Oreški (Univerza v Zagrebu, Fakulteta organizacije in informatike v Varaždinu) Blaž Frešer (Univerza v Mariboru, Ekonomsko-poslovna fakulteta)
Jezikovni pregled *Language edeting*	Alenka Plos (Univerza v Mariboru, Ekonomsko-poslovna fakulteta)
Tehnična urednika *Technical editors*	Dunja Legat (Univerza v Mariboru, Univerzitetna založba) Jan Perša (Univerza v Mariboru, Univerzitetna založba)
Oblikovanje ovitka *Cover designer*	Jan Perša (Univerza v Mariboru, Univerzitetna založba)
Grafične priloge *Graphic material*	Tominc, Rožman, 2023
Grtafika na ovitku *Cover graphics*	Analytics Statistics, avtor: madartzgraphics, CC0, Pixabay.com, 2023

Založnik *Published by*	**Univerza v Mariboru** **Univerzitetna založba** Slomškov trg 15, 2000Maribor, Slovenija https://press.um.si, zalozba@um.si
Izdajatelj *Issued by*	**Univerza v Mariboru** **Ekonomsko-poslovna fakulteta** Razlagova ulica 14, 2000 Maribor, Slovenija https://www.epf.um.si, epf@um.si
Izdaja *Edition*	Prva izdaja
Izdano *Published at*	Maribor, maj 2023
Vrsta publikacije *Publication type*	E-knjiga
Dostopno na *Available at*	https://press.um.si/index.php/ump/catalog/book/782

```
CIP - Kataložni zapis o publikaciji
Univerzitetna knjižnica Maribor

311.42(075.8)(076.5)(0.034.2)

TOMINC, Polona
    Poslovna statistika [Elektronski vir] /
avtorici Polona Tominc, Maja Rožman. - 1. izd.
- E-knjiga. - Maribor : Univerza v Mariboru,
Univerzitetna založba, 2023

Način dostopa (URL):
https://press.um.si/index.php/ump/catalog/
book/782
ISBN 978-961-286-741-6 (PDF)
doi: 10.18690/um.epf.6.2023
COBISS.SI-ID 152596995
```


ISBN	978-961-286-741-6 (pdf)
DOI	https://doi.org/10.18690/um.epf.6.2023
Cena *Price*	Brezplačni izvod
Odgovorna oseba založnika *For publisher*	prof. dr. Zdravko Kačič, rektor Univerze v Mariboru
Citiranje *Attribution*	Tominc, P., Rožman, M. (2023). *Poslovna statistika.* Univerza v Mariboru, Univerzitetna založba. doi: 10.18690/um.epf.6.2023

Kazalo

Uvod **1**

1 UREJANJE IN PRIKAZOVANJE PODATKOV **3**
Naloga 1 7
Naloga 2 9
Naloga 3 10
Naloga 4 11
Naloga 5 13
Naloga 6 15
Naloga 7 16
Naloga 8 17
Naloga 9 18
Naloga 10 20
Naloga 11 22

2 DESKRIPTIVNA STATISTIKA **23**
Naloga 12 28
Naloga 13 29
Naloga 14 31
Naloga 15 33
Naloga 16 34
Naloga 17 35
Naloga 18 35
Naloga 19 37
Naloga 20 38
Naloga 21 39

3 TEORETIČNA NORMALNA PORAZDELITEV **41**
Naloga 22 44
Naloga 23 46
Naloga 24 47
Naloga 25 48
Naloga 26 48
Naloga 27 49
Naloga 28 50
Naloga 29 52

4 ENOSTAVNA REGRESIJSKA ANALIZA **53**
Naloga 30 57
Naloga 31 60
Naloga 32 61
Naloga 33 62
Naloga 34 63
Naloga 35 64
Naloga 36 65
Naloga 37 66
Naloga 38 67

5 OSNOVE VZORČENJA ... 69
Naloga 39 ... 72
Naloga 40 ... 73
Naloga 41 ... 73
Naloga 42 ... 74
Naloga 43 ... 74
Naloga 44 ... 75
Naloga 45 ... 76
Naloga 46 ... 77
Naloga 47 ... 77
Naloga 48 ... 78
Naloga 49 ... 79
Naloga 50 ... 79
Naloga 51 ... 80
Naloga 52 ... 81
Naloga 53 ... 81

6 ČASOVNE VRSTE ... 83
Naloga 54 ... 86
Naloga 55 ... 87
Naloga 56 ... 88
Naloga 57 ... 89
Naloga 58 ... 90
Naloga 59 ... 91
Naloga 60 ... 92
Naloga 61 ... 93
Naloga 62 ... 94
Naloga 63 ... 95
Naloga 64 ... 96
Naloga 65 ... 97
Naloga 66 ... 98

7 REŠITVE NALOG ... 99
7.1 Urejanje in prikazovanje podatkov ... 99
7.2 Deskriptivna statistika ... 107
7.3 Teoretične porazdelitve ... 120
7.4 Enostavna regresijska analiza ... 125
7.5 Osnove vzorčenja ... 138
7.6 Časovne vrste ... 148

OBRAZCI ... 159
1. del: Urejanje, prikazovanje in analiza podatkov ... 159
2. del: Korelacija in regresija ... 163
3. del: Časovne vrste ... 164
4. del: Osnovni pojmi statističnega sklepanja ... 165

LITERATURA ... 169

Uvod

Pri izvedbi statističnega raziskovanja se običajno srečamo z množico podatkov različnega izvora in različnih tipov. Ti podatki, pretvorjeni v ustrezne statistične rezultate, naj bi čim natančneje opisovali proučevani pojav (na primer, ekonomski, družbeni itd.) ali ponujali uporabniku čim natančnejšo statistično sliko tega pojava. Tako je statistika znanstvena disciplina, ki se ukvarja z zbiranjem podatkov in njihovim prikazovanjem, z obdelavo podatkov ter analizo dobljenih rezultatov na vseh področjih raziskovanja množičnih pojavov.

Pri vsaki statistični raziskavi moramo v skladu s postavljenim ciljem in namenom raziskave pridobiti ustrezne podatke. Podatke delimo na primarne, to so tisti, ki jih moramo pridobiti posebej za raziskavo, ter sekundarne, to so podatki, ki so že zbrani in jih lahko uporabimo brez dodatnega zbiranja (že zbrane podatke na primer lahko dobimo na Statističnem uradu RS).

O *deskriptivni ali opisni statistiki* govorimo takrat, kadar opisujemo dani podatkovni niz in predstavljamo npr. velikost vzorca (N), odstotke števila statističnih enot z določeno lastnostjo (%), frekvence (f_k), minimalne (Min) in maksimalne (Max) vrednosti spremenljivk, srednje vrednosti, standardni odklon, varianco itd. *Inferenčna statistika* se nanaša na ocenjevanje parametrov in preverjanje domnev z ustreznimi statističnimi metodami (na primer regresijska analiza, t-test, ANOVA itd). *Parameter* je številska ali opisna vrednost, ki opisuje neko značilnost statistične množice ali populacije. *Statistika* pa je številska ali opisna vrednost, ki ocenjuje neko značilnost statistične množice in jo dobimo iz vzorca.

Zato je predmet **Poslovna statistika**, za katerega je namenjena ta zbirka vaj, pomemben za razvijanje sposobnosti razumevanja informacij v podatkih. Predmet Poslovna statistika vsebuje naslednje vsebinske sklope:

a) prikazovanje podatkov v tabelah in grafih,
b) relativna števila,
c) srednje vrednosti, mere variabilnosti, asimetrije in sploščenosti,
d) intervalno ocenjevanje vrednosti statističnih parametrov in osnove preizkušanja domnev o statističnih parametrih,
e) osnove enostavne regresije,
f) osnove analize in napovedovanja vrednosti v časovnih vrstah.

Študenti v okviru predmeta spoznajo uporabnost statističnih metod pri reševanju poslovnih problemov ter utrdijo in nadgradijo teoretično znanje na področju statističnih tehnik in metod, ki omogočajo spremeniti različne podatke v uporabne informacije za poslovno odločanje. Študenti usvojijo analitičen matematično statističen pristop k preučevanju poslovnih problemov, ki se sestoji iz naslednjih korakov: (1) formulacija problema na statističen način, (2) izbira ustrezne statistične metode, (3) reševanje problema in (4) interpretacija rezultatov v smislu možnih rešitev problema.

1 UREJANJE IN PRIKAZOVANJE PODATKOV

Statistika je veda, ki kvantitativno proučuje pojave v naravi in družbi ter tako z različnimi statističnimi metodami odkriva zakonitosti teh pojavov (Moore idr., 2016; Selvamuthu in Das, 2018, Ghauri idr., 2020).

Statistično raziskavo opravimo na statistični množici. Statistična množica je možica statističnih enot, ki izpolnjujejo določene opredeljujoče lastnosti. Statistično množico lahko imenujemo *populacija*. Vsak posamezni element statistične množice imenujemo *statistična enota* (na primer en študent, eno podjetje). Če je statistična množica ali populacija prevelika, raziskavo opravimo na *vzorcu* (Moore idr., 2016; Holmes idr., 2018). Vzorec je del celotne populacije, na osnovi katerega izvedemo sklepanje o celotni populaciji, in mora imeti določene lastnosti, ki jih bomo opisali v nadaljevanju. Pri tem je pomembno zagotoviti reprezentativnost vzorca. Vzorec je reprezentativen, če omogoča, da lahko rezultate, ki smo jih dobili na osnovi vzorca, posplošimo na celotno statistično množico. Temeljno načelo pa je, da mora biti vzorec slučajen. Za *slučajni vzorec* velja, da ima vsaka statistična enota v statistični množici ali populaciji znano in neničelno verjetnost (ta verjetnost je vnaprej znana), da je izbrana v slučajni vzorec (Tominc in Kramberger, 2007; Lind idr., 2021). Število statističnih enot, ki jih zajamemo v raziskavi, označujemo s črko N (število statističnih enot v statistični množici označimo z N in število statističnih enot v vzorcu označimo z n).

Lastnost, ki jo preučujemo pri posamezni statistični enoti, je *statistična spremenljivka*. Statistične spremenljivke so lahko opisne (atributne) ali številske (numerične). Za opisne spremenljivke velja, da lahko njihove vrednosti izražamo le z besedami (na primer spol, rojstni kraj, izobrazba). Številske spremenljivke so tiste spremenljivke, katerih vrednosti lahko izražamo s števili (na primer starost, prihodek). Med številskimi spremenljivkami pa ločimo zvezne in nezvezne spremenljivke. Zvezne spremenljivke so številske spremenljivke, ki lahko zavzamejo katerokoli vrednost na intervalu (na primer teža izdelka, dolžina, čas). Nezvezne spremenljivke (diskretne spremenljivke) pa so tiste številske spremenljivke, ki lahko zavzamejo le določene končne, največkrat celoštevilčne vrednosti (na primer število podjetij, število članov v gospodinjstvu).

Primer razvrščanja statističnih spremenljivk:

Statistična spremenljivka	Vrednost statistične spremenljivke	Vrsta statistične spremenljivke
Število članov v gospodinjstvu	2; 3; 4; 5	numerična, nezvezna
Število članov v gospodinjstvu	manj kot 3; 3 ali več	Opisna
Mesečna neto plača v d.e.	900; 1.200,80; 2.100; 2.530,55	numerična, zvezna
Dolžina izdelka v cm	8,5; 10; 15,8; 13,4; 16	numerična, zvezna
Spol	moški, ženski	Opisna
Velikost podjetja	malo, srednje veliko, veliko	Opisna

Najosnovnejšo obliko statističnih podatkov predstavljajo statistične vrste, ki jih delimo na tri skupine: *časovne, krajevne* in *stvarne statistične vrste* (Tominc in Kramberger, 2007). Tabelarično urejene podatke statistične vrste lahko grafično prikažemo z ustreznim grafom, ki še bolj nazorno prikazuje značilnosti statistične vrste. Tako je najpomembnejša značilnost časovne statistične vrste, ki jo želimo običajno spoznati, *dinamika,* kar pomeni spreminjanje vrednosti opazovane spremenljivke skozi čas. Za prikaz dinamike v časovni vrsti je tako zelo pogosto uporabljan grafični prikaz, ki ga imenujemo *linijski grafikon.*

V raziskavah so podatki o vrednosti številske statistične spremenljivke za posamične statistične enote, kadar jih enostavno zapisujemo enega za drugim, nepregledni ali neurejeni, zato jih je smiselno urediti po velikosti v ranžirno vrsto ali jih združiti v skupine v frekvenčne razrede (vse podatke, s katerimi razpolagamo, razdelimo v določeno število frekvenčnih razredov, ki so lahko enako ali različno široki – širino k-tega frekvenčnega razreda označimo z i_k in predstavlja pri zvezno opredeljenih mejah razredov razliko med zgornjo in spodnjo mejo k-tega razreda: $y_{k,\,max}$ in $y_{k,\,min}$). Vrednosti številske spremenljivke lahko uredimo v razrede *frekvenčne porazdelitve*, ki jih grafično prikazujemo s *frekvenčnimi histogrami ali poligoni.* Če analiziramo vrednosti opisne spremenljivke, pa le-te uredimo v skupine na osnovi možnih vrednosti opisne spremenljivke (Tominc in Kramberger, 2007). *Frekvenca* *(f_k)* nam pove, kako pogosto se pojavlja vrednost spremenljivke, ki je po vrednosti znotraj mej k-tega razreda (med spodnjo in zgornjo mejo k-tega razreda). Na primer, 8 študentov je doseglo na izpitu iz Poslovne statistike med 90 in 100 točk, 10 študentov je doseglo med 85 in manj kot 90 točk, 15 študentov je doseglo med 70 in manj kot 85 točk ipd. V navedenem primeru je število študentov *frekvenca*. *Kumulativna frekvenca (F_k)* k-tega frekvenčnega razreda pove, pri koliko statističnih enotah je vrednost statistične spremenljivke enaka ali manjša od zgornje meje tega k-tega frekvenčnega razreda. Torej če seštejemo frekvenco k-tega razreda in frekvence vseh razredov pred tem, dobimo kumulativno frekvenco k-tega razreda. *Relativna frekvenca* je razmerje (ali delež) med številom statističnih enot z določeno vrednostjo spremenljivke (ali vrednostjo spremenljivke v opredeljenem intervalu za vrednosti spremenljivke) in skupnim številom statističnih enot v podatkovnem nizu. Relativne frekvence so lahko izražene v deležu ali v odstotku.

Relativno strukturo podatkovnega niza (vzorca ali statistične množice), bodisi glede na število statističnih enot v posameznih skupinah ali razredih frekvenčne porazdelitve v primerjavi z vsemi statističnimi enotami podatkovnega niza bodisi glede na vsoto vrednosti spremenljivke v posameznih skupinah v primerjavi s celotno vsoto vseh vrednosti spremenljivke (*total),* grafično najpogosteje prikazujemo s *strukturnim stolpcem* (Tominc in Kramberger, 2007), v katerem prikažemo *strukturne odstotke*. Drugi grafični prikazi so še *strukturni krog, strukturni polkrog, strukturni kvadrat.*

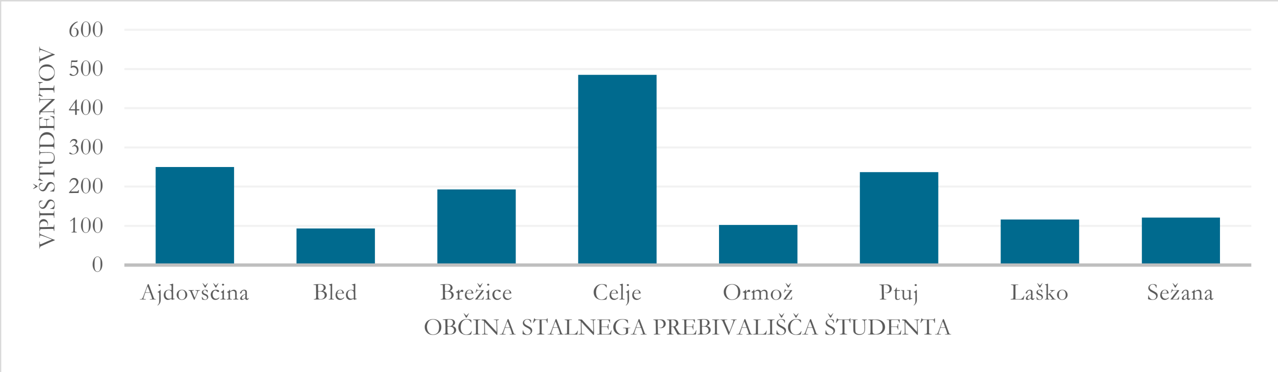
Primer rešene naloge:

V preglednici so podatki o študentih terciarnega izobraževanja, ki so vpisani v visokošolski univerzitetni študij po občini stalnega prebivališča v študijskem letu 2021/2022.

Občina	Ajdovščina	Bled	Brežice	Celje	Ormož	Ptuj	Laško	Sežana
Vpis študentov	250	93	193	485	102	237	116	121

Statistično vrsto za število vpisa študentov, ki so vpisani v visokošolski univerzitetni študij v študijskem letu 2021/2022, prikažite grafično.

Grafični prikaz:



Prikazana je krajevna statistična vrsta, saj so vrednosti spremenljivke urejene po geografskih enotah (občinah).

Primer rešene naloge:

Študenti podiplomskega študijskega programa na Ekonomsko-poslovni fakulteti v Mariboru pri določenem predmetu na izpitu za rešitev testa porabijo različno količino časa (čas je merjen v minutah). Podatki so podani v preglednici:

Čas reševanja testa v minutah	Število študentov
Od 10 do pod 20	5
Od 20 do pod 30	14
Od 30 do pod 40	23
Od 40 do pod 50	38
Od 50 do pod 60	60
Skupaj	140

a) Opredelite statistično enoto ter spremenljivko.

Statistična enota je vsak posamezen element statistične množice, kar pomeni, da je v našem primeru statistična enota en študent podiplomskega študijskega programa na Ekonomsko-poslovni fakulteti v Mariboru. Statistična spremenljivka opisuje lastnost statistične enote in je v našem primeru čas reševanja testa v minutah (številska, zvezna spremenljivka).

b) Izračunajte kumulativne člene frekvenčne porazdelitve.

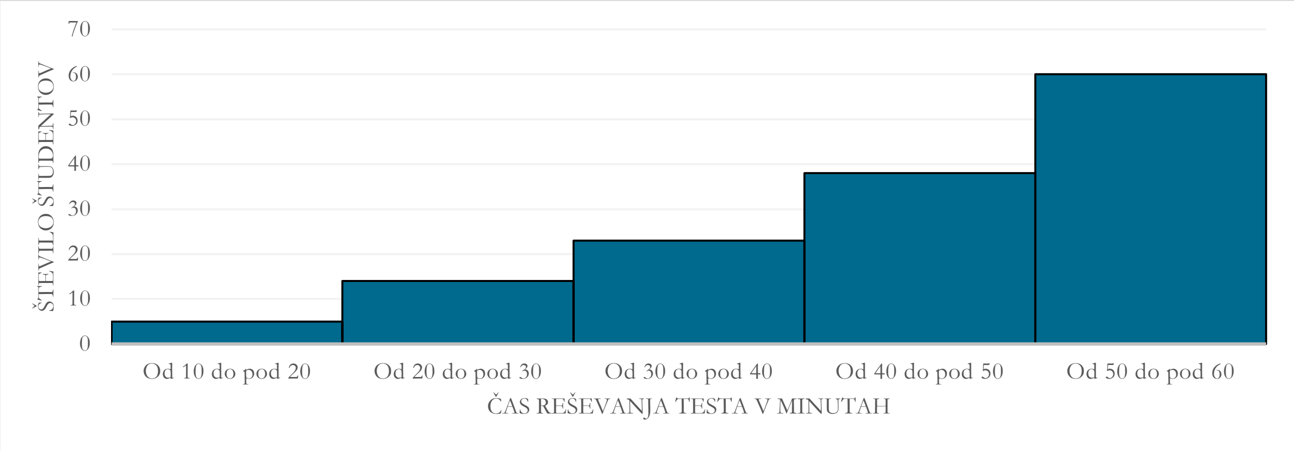
Kumulativni členi frekvenčne porazdelitve

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

Čas reševanja testa v minutah	Število študentov (f_k)	Kumulativna frekvenčna porazdelitev (F_k)
Od 10 do pod 20	5	5
Od 20 do pod 30	14	5 + 14 = 19
Od 30 do pod 40	23	19 + 23 = 42
Od 40 do pod 50	38	42 + 38 = 80
Od 50 do pod 60	60	80 + 60 = 140
Skupaj	140	

c) Frekvenčno porazdelitev prikažite grafično.

Grafični prikaz:



Naloga 1

V preglednici so podatki o številu trgovin neke trgovske verige, v šestih slovenskih regijah, v nekem časovnem obdobju ter podatki o višini investicij v trgovske objekte te trgovske verige (v d.e.):

Regija	Gorenjska	Goriška	Primorska	Koroška	Pomurska	Posavska	Zasavska
Št. trgovin	21	25	14	17	25	12	18
Investicije (v d.e.)	314	100	100	300	260	175	250

a) Kako imenujemo statistični vrsti v preglednici?

b) Statistično vrsto za število trgovin prikažite grafično.

c) Za podatke o višini investicij po regijah prikažite relativno strukturo skupno porabljenih sredstev po regijah.

Naloga 2

Analizirati želimo najvišjo dnevno temperaturo (°C) v Mariboru od januarja do decembra v določenem letu. Podatki so podani v preglednici:

Mesec	Jan.	Feb.	Mar.	Apr.	Maj	Jun.	Jul.	Avg.	Sep.	Nov.	Okt.	Dec.
Temp.	3,6	4,6	11,2	15,4	20,6	23.6	25,7	25,4	21,2	15,5	8,6	4,5

a) Kako imenujemo statistični vrsti v preglednici?

b) Statistično vrsto za število trgovin prikažite grafično.

Naloga 3

V preglednici so podani podatki o skupni količini padavin (v mm) po vremenskih postajah, ki so bile izmerjene 17. septembra 2022:

Vremenska postaja	Količina padavin (v mm)
Leskovec pri Krškem	72,6
Krško	75,9
Lisca	126,6
Bučerca	58,6
Cerklje ob Krki	91,4
Birna vas	96,6
Sevnica	140,5
Pišece	89,4

a) Kako imenujemo statistični vrsti v preglednici?

b) Statistično vrsto prikažite grafično.

Naloga 4

V tovarni porabijo zaposleni za izdelavo določenega izdelka različno količino časa (čas je merjen v minutah). Podatki so podani v preglednici:

Izdelava določenega izdelka v minutah	Število zaposlenih
Od 11 do 20	12
Od 21 do 30	16
Od 31 do 40	10
Od 41 do 50	8
Od 51 do 60	4
Skupaj	50

a) Opredelite statistično množico, statistično enoto ter spremenljivko.
b) Izračunajte kumulativne člene frekvenčne porazdelitve.
c) Frekvenčno porazdelitev ter kumulativno frekvenčno porazdelitev prikažite grafično.

Naloga 5

V vzorec smo zajeli naključno izbranih 190 malih podjetij v Sloveniji, ki smo jih razvrstili glede na število zaposlenih. Podatki za 190 malih podjetij glede na število zaposlenih:

Število zaposlenih	Število podjetij
Od 1 do 10	16
Od 11 do 20	27
Od 21 do 30	42
Od 31 do 40	55
Od 41 do 50	50

a) Opredelite statistično množico, statistično enoto ter spremenljivko ter njene značilnosti.
b) Določite spodnje in zgornje meje razredov ter določite širino razredov.
c) Prikažite strukturo podjetij glede na število zaposlenih v strukturnem stolpcu in strukturnem krogu.
d) Frekvenčno porazdelitev prikažite grafično.

Naloga 6

Razpolagamo s podatki o oceni zadovoljstva zaposlenih (na lestvici od 1 do 100) za 50 zaposlenih:

15	17	18	18	18	18	19	19	19	20
23	24	25	25	26	27	29	29	29	29
30	30	35	37	38	36	39	39	39	39
41	45	42	46	46	48	48	48	46	49
55	62	65	72	72	85	91	95	91	95

a) Opredelite statistično množico, statistično enoto ter spremenljivko, njene značilnosti in zalogo vrednosti spremenjivke.

b) Sestavite frekvenčno porazdelitev pri pogojih: $y_{1,min} = y_{min}$; $r = 8$, $i_1 = 10$, $i_{2\text{-}8} = 9$; meje razredov so podane nezvezno.

c) Frekvenčno porazdelitev ter kumulativno frekvenčno porazdelitev prikažite grafično.

Naloga 7

Za 300 študentov, ki smo jih opazovali glede na število ur študija (zvezna spremenljivka) za izpit iz predmeta Statistika, imate na razpolago naslednje podatke:

y_{min} = 41 ur y_{max} = 97 ur $y_{1,min}$ = 40 ur $y_{6,max}$ = 100 ur

$r = 6$ $i_k = i = 10$ za $k = 1,2,\ldots,r$

$F_1 = 25$ $F_2 = 75$ $F_3 = 175$ $F_4 = 250$ $F_5 = 290$ $F_6 = 300$

a) Na osnovi danih podatkov sestavite frekvenčno porazdelitev in jo grafično prikažite.

b) Koliko odstotkov študentov je porabilo od 60 do 70 ur študija?

c) Koliko odstotkov študentov je porabilo od 60 do 70 ur študija?

Naloga 8

Podatki za 500 kupcev glede na porabljen znesek za nakup v neki trgovini, na dan 31.12. preteklega leta, v eni od slovenskih regij so:

Znesek za nakup	Število kupcev
Od 1 do pod 50	80
Od 50 do pod 100	125
Od 100 do pod 150	148
Od 150 do pod 200	112
Od 200 do pod 250	35

a) Opredelite statistično enoto, statistično množico, statistično spremenljivko ter njene vrednosti.
b) Prikažite relativno strukturo statistične množice s strukturnim stolpcem.

Naloga 9

V trgovinah so pričeli s prodajo novega izdelka. Število kosov prodanega izdelka v enem tednu je podano v preglednici.

Število trgovin	Število kosov prodanega izdelka
Od 0 do 5	7
Od 6 do 11	14
Od 12 do 17	19
Od 18 do 23	26
Od 24 do 29	35
Od 30 do 35	42

a) Pojasnite statistično enoto, statistično množico, statistično spremenljivko in njene vrednosti.
b) Izračunajte in pojasnite člene kumulativne frekvenčne porazdelitve in jo prikažite grafično.
c) Izračunajte in pojasnite relativne frekvence ter grafično prikažite strukturo statistične množice glede na skupno število kosov prodanega izdelka v enem dnevu v posameznih razredih.

Naloga 10

Izbruh novega virusa COVID-19 je spremenil način poslovanja številnih podjetij. V preglednici so podatki za 137 podjetij in številu zaposlenih, ki so prešli na delo na daljavo v letu 2021.

Število zaposlenih	Število podjetij
Od 5 do 15	38
Od 16 do 26	35
Od 27 do 37	27
Od 38 do 48	20
Od 49 do 59	17
Skupaj	137

a) Pojasnite statistično enoto, statistično množico, statistično spremenljivko in njene vrednosti.
b) Frekvenčno porazdelitev prikažite grafično.
c) Izračunajte in pojasnite člene kumulativne frekvenčne porazdelitve.
d) Izračunajte in pojasnite relativne frekvence ter grafično prikažite relativno strukturo statistične množice.

Naloga 11

Za 450 zaposlenih, ki smo jih opazovali glede na število ur bolniških izostankov (zvezna spremenljivka) v podjetju X, so na razpolago naslednji podatki:

$y_{1,min} = 10$ ur $y_{7,max} = 73$ ur $r = 7$ $i = 9$

$F_1 = 13$ $F_2 = 38$ $F_3 = 154$ $F_4 = 188$ $F_5 = 244$ $F_6 = 282$ $F_7 = 450$

a) Opredelite statistično enoto in spremenljivko ter njene značilnosti.
b) Na osnovi danih podatkov sestavite frekvenčno porazdelitev in jo grafično prikažite.

c) Koliko odstotkov zaposlenih je imelo od 37 do 46 ur bolniških izostankov?

d) Koliko odstotkov zaposlenih je imelo do 64 ur bolniških izostankov?

Univerzitetna založba
Univerze v Mariboru

2 DESKRIPTIVNA STATISTIKA

V okviru deskriptivne ali opisne statistike običajno opredeljujemo kvantile, mere osrednje tendence, mere variabilnosti ter mere asimetrije in sploščenosti.

Kvantili so mere, ki opredeljujejo položaj posamezne statistične enote glede na ostale po vrednosti obravnavane statistične spremenljivke. Najpomembnejši kvantili so kvartili in decili, ki razdelijo vse vrednosti spremenljivke, urejene po velikosti, na četrtine oziroma na deset enakih delov, ter mediana, ki je tista vrednost spremenljivke, ki razdeli vse statistične enote na dve enaki skupini. Med najpomembnejše *mere osrednje tendence* uvrščamo aritmetično sredino, modus (najpogostejša vrednost) in mediano, ki smo jo že omenili. Najpomembnejše *mere variabilnosti* zajemajo variacijski razmik, kvartilni in decilni razmik, varianco in standardni odklon ter koeficient variabilnosti. Med *mere asimetrije in sploščenosti* uvrščamo koeficient asimetrije in koeficient sploščenosti (Holmes idr., 2018; Heumann idr., 2016).

Za izračun kvantilov je treba statistične enote razvrstiti po vrednosti obravnavane spremenljivke od najmanjše do največje (*ranžirna vrsta* pri manjšem številu statističnih enot ali frekvenčna porazdelitev pri večjem številu statističnih enot). Pri ranžirni vrsti rang R_i izraža i-to mesto, ki pripada določeni vrednosti obravnavane spremenljivke (y_i). Kvantilni rang P_i pa nam pove (v deležu), kolikšen delež (ali preračunano v odstotek) statističnih enot ima vrednost spremenljivke – manjšo ali največ enako obravnavani vrednosti spremenljivke y_i, ki ji pripada rang R_i.

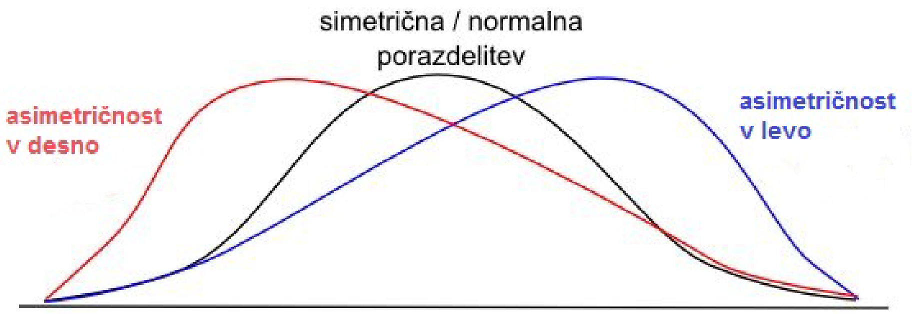
Kvartili (Q) razdelijo urejene podatke na četrtine (po 25 %), *decili (D)* pa na desetine (po 10 %). *Mediana (Me)* je enaka drugemu kvartilu, pa tudi petemu decilu, kot smo omenili že zgoraj.

Aritmetična sredina je srednja vrednost, ki jo izračunamo tako, da vsoto vseh vrednosti obravnavane spremenljivke delimo s številom statističnih enot v podatkovnem nizu. Aritmetično sredino označimo kot $\bar{y}$, če obravnavamo podatke statistične množice (*N*), in kot $\bar{Y}$, če obravnavamo podatke iz vzorca (*n*). *Modus* je enak tisti vrednosti spremenljivke, ki se najpogosteje pojavlja (označimo ga z *Mo*). *Mediana* je srednja vrednost, od katere ima 50 % enot manjše ali enake vrednosti, 50 % enot pa večje vrednosti (označimo jo z *Me*), in sodi v skupino kvantilov, kot je bilo že omenjeno.

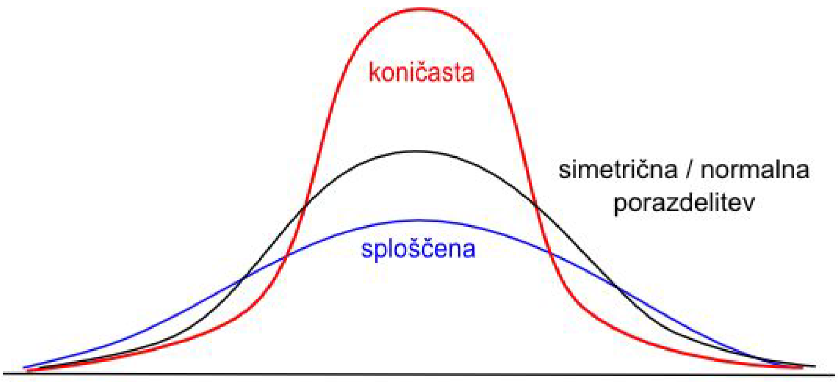
Variacijski razmik (*VR*) je razlika med največjo (y_{max}) in najmanjšo (y_{min}) vrednostjo spremenljivke. *Kvartilni razmik* je mera variabilnosti, ki označuje razpon srednje velikih vrednosti (50 % podatkov na sredini ranžirne vrste ali frekvenčne porazdelitve. To je razlika med *tretjim kvartilom* (Q_3) in *prvim kvartilom* (Q_1). *Decilni razmik* pa je razlika med *devetim* (D_9) *in prvim decilom* (D_1) (srednjih 80 % vrednosti). *Varianca* meri povprečno kvadratno odstopanje posameznih vrednosti spremenljivke od aritmetične sredine. Definirana je kot povprečje kvadratov odklonov posameznih vrednosti od aritmetične sredine. Varianco pri podatkih statistične množice označimo kot σ^2, pri podatki iz vzorca pa jo označimo kot s^2 (*nepristranska ocena variance*). *Standardni odklon* je definiran kot kvadratni koren iz variance. S standardnim odklonom lahko merimo, kako so razpršene vrednosti okoli aritmetične sredine zbranih podatkov (standardni odklon pri podatkih statistične množice označimo kot σ, pri podatkih iz vzorca pa kot označimo s) (Tominc in Kramberger, 2007). *Koeficient variabilnosti, izražen v % (KV %),* pove, koliko odstotkov aritmetične sredine predstavlja standardni odklon za dani podatkovni niz.

Vsako porazdelitev je mogoče standardizirati oziroma je mogoče vsaki vrednosti spremenljivke poiskati njeno standardizirano vrednost, kar bomo podrobneje obravnavali v naslednjem poglavju o teoretični normalni porazdelitvi.

Med *mere asimetrije in sploščenosti* uvrščamo koeficient asimetrije in koeficient sploščenosti. *Asimetrične porazdelitve* (slika 1) so lahko *asimetrične v desno* (pozitivna asimetrična porazdelitev), zanje je značilna večja gostitev pri manjših vrednostih spremenljivke, ali *asimetrične v levo* (negativna asimetrična porazdelitev) in je zanje značilna večja gostitev vrednosti pri večjih vrednostih spremenljivke. Koeficient asimetrije je manjši od 0, če je za porazdelitev spremenljivke značilna asimetrija v levo, pri asimetriji v desno je koeficient asimetrije večji od 0. 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* (slika 2) 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 teoretični normalni porazdelitvi, ki jo bomo obravnavali v nadaljevanju, sta tako koeficienta asimetričnosti in sploščenosti enaka 0 (Freedman idr., 2007; Evans idr., 2010).



Slika 1: Asimetričnost porazdelitve
(Vir: Freedman idr., 2007; Evans idr., 2010)



Slika 2: Sploščenost porazdelitve
(Vir: Freedman idr., 2007; Evans idr., 2010)

Primer rešene naloge:

Mesečno število izposojenih knjig v knjižnici X za šest študentov je bilo sledeče: 4, 6, 8, 2, 5, 7.

a) Izračunajte in pojasnite povprečno vrednost za mesečno število izposojenih knjig v knjižnici.
b) Izračunajte in pojasnite standardni odklon za mesečno število izposojenih knjig v knjižnici.
c) Pojasnite mero variabilnosti, ki upošteva variabilnost za 80 % študentov, ki se glede na mesečno število izposojenih knjig razvrščajo na sredino ranžirne vrste.

a) Za izračun povprečnega števila izposojenih knjig uporabimo enačbo *(aritmetična sredina iz nerazvrščenih vrednosti – v obrazcih št. 1.24)*:

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

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

Odg.: Povprečna vrednost za mesečno število izposojenih knjig v knjižnici šestih študentov znaša 5,3 knjig.

b) Izračunamo varianco po enačbi (*varianca iz nerazvrščenih vrednosti – v obrazcih št. 1.34*):

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

$$\sigma^2 = \frac{1}{6} \cdot [(4-5{,}3)^2 + (6-5{,}3)^2 + (8-5{,}3)^2 + (2-5{,}3)^2 + (5-5{,}3)^2 + (7-5{,}3)^2] = 3{,}89 \text{ knjig}^2$$

Izračunamo standardni odklon (v obrazcih št. 1.38):

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

$$\sigma = \sqrt{3{,}89} = 1{,}97 \text{ knjig}$$

c) Podatke uredimo v ranžirno vrsto:

R_i	1	2	3	4	5	6
y_i	2	4	5	6	7	8

Decilni razmik (v obrazcih št. 1.33): $D = D_9 - D_1$

D_1: 10 %

Uporabimo skupino enačb z naslovom *kvantili iz nerazvrščenih vrednosti* (v obrazcih št. 1.18, 1.19 in 1.20):

$P_i = 0{,}1$

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

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

$R_0 \leq R_i < R_1$

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

$y_0 \leq y_i < y_1$

$y_0 = 2 \leq y_i < y_1 = 4$

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

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

10 % študentov si je v knjižnici izposodilo 2,2 knjig ali manj.

D_9: 90 %

$P_i = 0{,}9$

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

$R_0 \leq R_i < R_1$

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

$y_0 \leq y_i < y_1$

$y_0 = 7 \leq y_i < y_1 = 8$

$y_i = 7 + \frac{5{,}9-5}{6-5} \cdot (8 - 7) = 7{,}9$ knjig

90 % študentov si je v knjižnici izposodilo 7,9 knjig ali manj.

D = 7,9–2,2 = 5,7 knjig

80 % študentov, ki se glede na mesečno število izposojenih knjig razvrščajo na sredino ranžirne vrste, se med seboj razlikujejo za največ 5,7 izposojenih knjig.

Naloga 12

Definirajte oziroma pojasnite naslednje pojme:

ranžirna vrsta, kvartili, decili, mediana, modus, aritmetična sredina, variacijski razmik, kvartilni oziroma decilni razmik, varianca, standardni odklon, koeficient variabilnosti, koeficient asimetrije, koeficient sploščenosti.

Naloga 13

V 20 poslovalnicah trgovskega podjetja X je bilo v opazovanem letu naslednje število pritožb kupcev:

16, 8, 32, 18, 4, 26, 10, 12, 34, 14, 25, 19, 5, 9, 17, 33, 27, 22, 7, 20

a) Izračunajte, koliko odstotkov poslovalnic je imelo število pritožb manjše od 13.
b) Izračunajte, koliko odstotkov poslovalnic je imelo število pritožb večje od 23.
c) Izračunajte kvartilni razmik.
d) Izračunajte povprečno število pritožb.
e) Izračunajte število pritožb v 50 % poslovalnic z največjim številom pritožb.
f) Izračunajte mero variabilnosti, ki upošteva število pritožb v 80 % poslovalnic, ki se glede na število pritožb razvrščajo na sredino ranžirne vrste.
g) V konkurenčnem trgovskem podjetju Y je povprečno število pritožb v njihovih poslovalnicah osem, standardni odklon pa dve pritožbi. V katerem trgovskem podjetju se poslovalnice glede na število pritožb kupcev med seboj bolj razlikujejo?

Naloga 14

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

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

Organizacija B: N = 730

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

Izračunajte in pojasnite:

a) kvartilni razmik in mediano (za organizacijo A),

b) povprečno porabljeni čas za en izdelek v organizaciji A in v organizaciji B,

c) variacijski razmik za podatke organizacije A,

d) varianco in standardni odklon za podatke organizacij A in B,

e) koeficient variabilnosti za podatke organizacij A in B.

Naloga 15

Na nekem področju smo opazovali 260 sodnikov okrajnih sodišč po številu obravnavanih zadev v določenem časovnem razdobju. Podatki so v preglednici:

Število obravnavanih zadev	Število sodnikov
Od 31 do 60	35
Od 61 do 90	52
Od 91 do 120	74
Od 121 do 150	41
Od 151 do 180	32
Od 181 do 210	26
Skupaj	**260**

a) Navedite statistično množico, enoto, spremenljivko in vrednosti spremenljivke.
b) Grafično ocenite asimetrijo gornje porazdelitve.
c) Izračunajte delež standardnega odklona v aritmetični sredini.

Naloga 16

Na nekem cestnem odseku so v 45 zaporednih dneh našteli takšno število osebnih vozil:

Število osebnih vozil	Število dni
Od 1 do 10	6
Od 11 do 20	14
Od 21 do 30	15
Od 31 do 40	5
Od 41 do 50	5
SKUPAJ	45

a) Izračunajte in pojasnite variacijski razmik za število osebnih vozil.

b) Grafično prikažite frekvenčno porazdelitev.

c) Izračunajte in pojasnite aritmetično sredino za število osebnih vozil v opazovanih 45 dnevih.

Naloga 17

Donosi desetih delnic (v d.e.) so bili v opazovanem obdobju sledeči:

10	40	50	50	71	82	800	850	1000	1100

a) Opredelite statistično enoto, statistično spremenljivko in statistično množico.

b) Kolikšne donose je doseglo 25 % najmanj donosnih delnic in kolikšne 25 % najbolj donosnih delnic?

Naloga 18

V banki "X" so opazovali 1451 varčevalcev glede na višino vloženih sredstev (v 10^2 evrov). Podatki so v preglednici:

Vložena sredstva v 10^2 evrov	Število vlagateljev
Od 1 do manj kot 10	223
Od 10 do manj kot 20	356
Od 20 do manj kot 50	439
Od 50 do manj kot 100	245
Od 100 do manj kot 500	188

a) V banki "Y" je višina povprečno vloženih sredstev 456 vlagateljev 7530 evrov, varianca pa 7.022.500 evrov^2. V kateri banki se vlagatelji glede na vložena sredstva med seboj bolj razlikujejo?

b) Porazdelitev vlagateljev banke "X" prikažite grafično.

Naloga 19

Tedensko število nadur za osem zaposlenih v podjetju X je sledeče: 5, 6, 9, 11, 4, 8, 7, 3.

a) Izračunajte povprečno tedensko število nadur zaposlenih.
b) Izračunajte delež standardnega odklona v aritmetični sredini.
c) Kolikšno je tedensko število nadur 50 % zaposlenih, ki so imeli najmanj nadur?
d) Pojasnite mero variabilnosti, ki upošteva variabilnost za 50 % zaposlenih, ki se glede na tedensko število nadur razvrščajo na sredino ranžirne vrste.

Naloga 20

V podjetju X so za deset zaposlenih zabeležili število dni, ko so bili odsotni zaradi bolniškega staleža:

Zaposlen	1	2	3	4	5	6	7	8	9	10
Število bolniških dni	3	5	8	8	12	15	20	24	24	30

a) Za spremenljivko število bolniških dni izpišite modus.
b) Pojasnite mero variabilnosti, ki upošteva variabilnost za 80 % zaposlenih, ki se glede na število bolniških dni razvrščajo na sredino ranžirne vrste.
c) Kolikšno število bolniških dni so imeli tisti zaposleni, ki spadajo med 30 % zaposlenih z najmanj bolniških dni v podjetju?
d) Kolikšno število bolniških dni so imeli tisti zaposleni, ki spadajo med 30 % zaposlenih z največ bolniških dni v podjetju?

Naloga 21

Devet študentov je za nakup študijskega gradiva pri različnih predmetih porabilo določene zneske (v d.e.):

160	220	75	246	98	180	290	260	195

a) Opredelite statistično enoto in statistično spremenljivko.

b) Izračunajte in pojasnite koeficient asimetrije na podlagi mediane ter vse parametre, ki ste jih pri tem izračunali.

Univerzitetna založba
Univerze v Mariboru

3 TEORETIČNA NORMALNA PORAZDELITEV

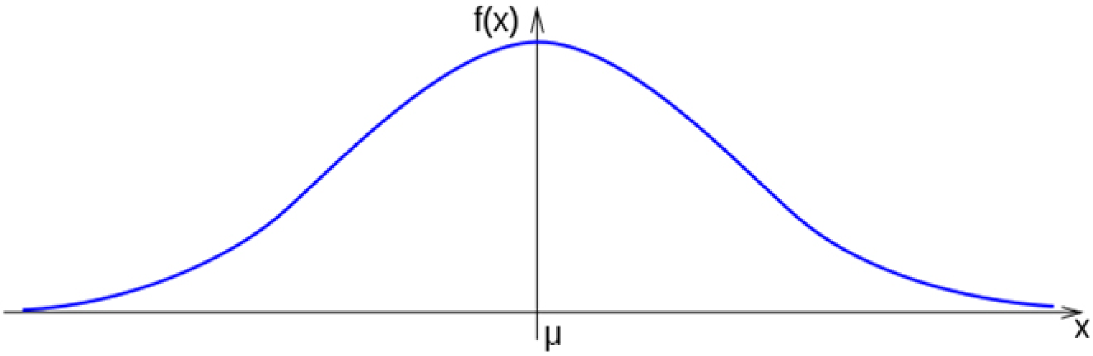
Normalna porazdelitev je definirana z dvema parametroma, aritmetično sredino ($\bar{y}$, velikokrat pa se za aritmetično sredino v statistični množici uporablja tudi oznaka μ), in standardnim odklonom (σ) (Holmes idr., 2018). Normalna porazdelitev je simetrična glede na aritmetično sredino, aritmetična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

Krivulja normalne porazdelitve je unimodalna, zvonasta, simetrična in se asimptotično približuje osi x. Večina lastnosti se v naravi razporeja v obliki normalne porazdelitve (Evans idr., 2010).

Sprememba standardnega odklona σ povzroči spremembo oblike normalne krivulje, krivulja postane bolj sploščena ali pa bolj koničasta, odvisno od σ. Sprememba aritmetične sredine pa povzroči, da se graf premakne v levo ali desno. To pomeni, da obstaja neskončno število normalnih porazdelitev. Pri uporabi v statistiki je pomembna porazdelitev *standardizirana normalna porazdelitev.*

Standardizirana normalna porazdelitev je normalna porazdelitev standardiziranih vrednosti, imenovanih tudi z-vrednosti, vsaki vrednosti y spremenljivke, porazdeljeni po poljubni normalni porazdelitvi, je mogoče izračunati njeno standardizirano z-vrednost (Holmes idr., 2018) z upoštevanjem *transformacijske enačbe* (v obrazcih).

Povprečje za standardizirano normalno porazdelitev je enako 0, standardni odklon in varianca pa sta enaka 1. Slika 3 prikazuje normalno porazdelitev.



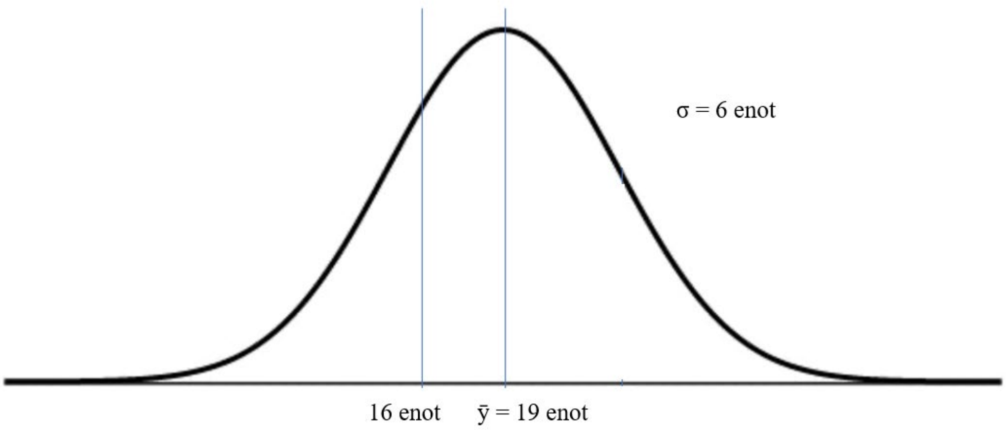
Slika 3: Normalna porazdelitev
(Vir: Balakrishnan in Nevzorov, 2003)

Kasneje se bomo srečali še z drugo teoretično porazdelitvijo – *t-porazdelitev*, v poglavju o vzorčenju. *Studentova t-porazdelitev* je zelo podobna normalni porazdelitvi. Glavna razlika je v tem, da nima enotne oblike. Njeno obliko določa število stopinj prostosti (n–1). Pri manjšem številu stopinj prostosti je t-porazdelitev bolj razvlečena od normalne (sploščena), z naraščanjem števila stopinj prostosti pa se po obliki vse bolj približuje normalni porazdelitvi.

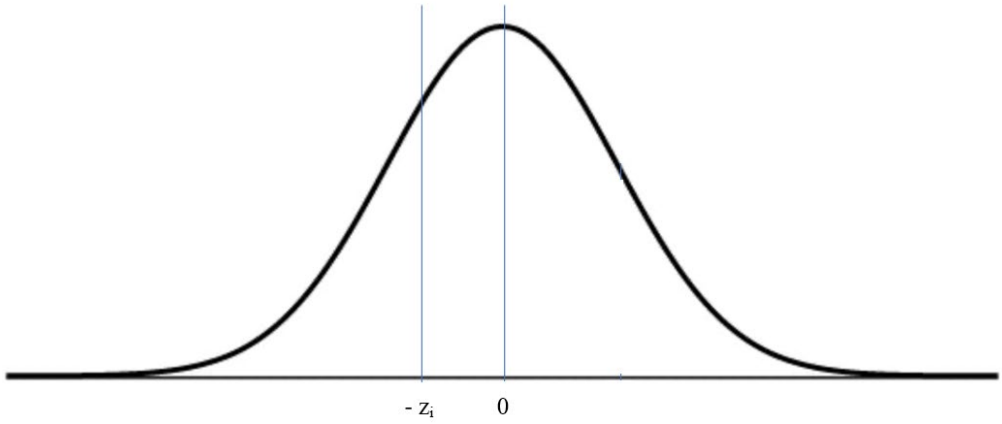
<u>Primer rešene naloge:</u>

Izračunajte verjetnost, da normalno porazdeljena spremenljivka y zavzame vrednost, ki je večja od 16 enot. Upoštevajte, da aritmetična sredina znaša 19 enot in standardni odklon 6 enot.

Prikaz normalne porazdelitve:



Standardizirana normalna porazdelitev:



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

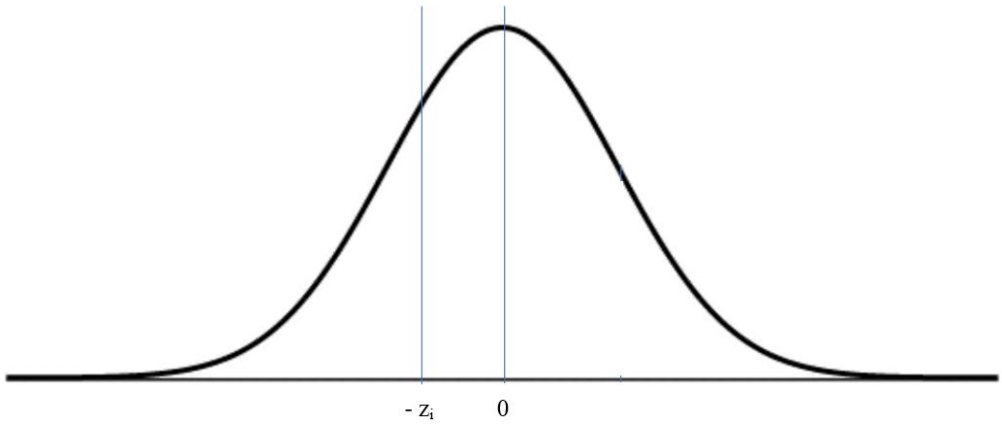
P (y > 16 enot) = 0,5 + H (–0,5) = 0,5 + 0,1915 = 0,6915 = 69,15 %

Vrednost H (z_i) odčitamo iz tabele *ploščine H(z) za standardizirano normalno porazdelitev.* Verjetnost, da normalno porazdeljena spremenljivka y zavzame vrednost, ki je večja od 16 enot, je 69,15 %.

Primer rešene naloge:

Dolžina minut pisanja testa pri določenem predmetu pri študentih se porazdeljuje po normalni porazdelitvi, z aritmetično sredino 40 minut in standardnim odklonom 15 minut. Izračunajte, koliko minut porabijo tisti študenti za pisanje testa pri določenem predmetu, ki spadajo med 20 % študentov, ki so pisali test najmanj časa.

Standardizirana normalna porazdelitev:



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

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

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

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

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

Študenti, ki spadajo med 20 % tistih, ki so za test porabili najmanj časa, so za pisanje porabili 27,4 minut.

Naloga 22

V gospodinjstvih v neki družbi je čas uporabe družinskega računalnika v gospodinjstvu za igranje igric porazdeljen po normalni porazdelitvi, z aritmetično sredino dve uri in standardnim odklonom 0,5 ure.

a) Izračunajte verjetnost, da je v naključno izbranem gospodinjstvu družinski računalnik v uporabi za igranje igric med 1,8 in 2,75 urami na dan.
b) V koliko odstotkih gospodinjstev se družinski računalnik uporablja več kot pet ur dnevno za igranje igric?
c) Koliko časa je družinski računalnik v uporabi za igranje igric pri tistih 25 odstotkih gospodinjstev, kjer je ta čas najkrajši?

Naloga 23

Število minut učenja en dan pred izpitom pri študentih pri predmetu Poslovna statistika se porazdeljuje po normalni porazdelitvi, z aritmetično sredino 240 minut in standardnim odklonom 80 minut. Izračunajte, kolikšna je verjetnost, da bo naključno izbran študent študiral en dan pred izpitom več kot 210 minut.

Naloga 24

Izračunajte verjetnost, da normalno porazdeljena spremenljivka y zavzame vrednost, ki je večja od 14 enot. Upoštevajte, da aritmetična sredina znaša 17 enot in standardni odklon 5 enot.

Naloga 25

Podjetje preučuje tri alternativne investicijske možnosti. Pričakovan donos v vseh primerih je porazdeljen po normalni porazdelitvi. Pri prvi alternativni investicijski možnosti povprečna vrednost donosa znaša 1780 € in standardni odklon 115 €. Pri drugi alternativni investicijski možnosti povprečna vrednost donosa znaša 2168 € in standardni odklon 425 €. Pri tretji alternativni investicijski možnosti povprečna vrednost donosa znaša 3000 € in standardni odklon 515 €.

Pri kateri investicijski možnosti je manjša verjetnost, da bo donos večji od 1500 €?

Naloga 26

V tovarni je 600 zaposlenih, ki jih proučujemo po odstotku dosežene norme (spremenljivka je normalno porazdeljena). Aritmetična sredina za odstotek dosežene norme v zadnjem delovnem mesecu je bila 102 %, standardni odklon pa 4 %.

a) Kolikšna je verjetnost, da bo naključno izbran zaposleni dosegel normo največ 98 %?
b) Kolikšna je verjetnost, da bo naključno izbran zaposleni dosegel normo med 97 in 105 %?
c) Določite mejne vrednosti za razmik, v katerem se nahaja 80 % zaposlenih, ki po doseženi normi ležijo simetrično na povprečno vrednost. Kako se imenuje izračunani razmik?

Naloga 27

Višina vloženih denarnih sredstev varčevalcev opazovane banke se porazdeljuje po normalni porazdelitvi z aritmetično sredino 1500 evrov in standardnim odklonom 400 evrov. Na osnovi standardizirane normalne porazdelitve izračunajte verjetnost, da je vloženi znesek naključno izbranega varčevalca od 600 do 1200 evrov.

Naloga 28

Dolžina minut izdelave enega izdelka pri zaposlenih v podjetju X se porazdeljuje po normalni porazdelitvi, z aritmetično sredino 90 minut in standardnim odklonom 35 minut.

a) Izračunajte, koliko minut porabijo za izdelavo enega izdelka tisti zaposleni, ki spadajo med 40 % tistih, pri katerih je poraba minut za izdelavo izdelka najvišja?

b) Izračunajte, koliko minut porabijo za izdelavo enega izdelka tisti zaposleni, ki spadajo med 30 % tistih, pri katerih je poraba minut za izdelavo izdelka najnižja?

c) Izračunajte, koliko odstotkov zaposlenih je porabilo za izdelavo enega izdelka med 50 in 105 minut za izdelavo enega izdelka?

d) Izračunajte, koliko odstotkov zaposlenih je porabilo za izdelavo enega izdelka med 100 in 130 minut za izdelavo enega izdelka?

Naloga 29

Ocena študentov na izpitu pri predmetu Poslovna statistika se porazdeljuje po normalni porazdelitvi, z aritmetično sredino 5,5 in varianco 2. Izračunajte $P(D3 < y < Q3) = \gamma$.

Univerzitetna založba
Univerze v Mariboru

4 ENOSTAVNA REGRESIJSKA ANALIZA

S pomočjo enostavne linearne regresije analiziramo odvisnost med odvisno (*y*) in eno neodvisno ali pojasnjevalno spremenljivko (*x*) oziroma utemeljeno domnevamo, da neodvisna spremenljivka *(x)* vpliva na odvisno spremenljivko *(y)*. Grafični prikaz, ki ga uporabljamo pri enostavni regresijski analizi, se imenuje *razsevni grafikon*. Razsevni grafikon je grafični prikaz povezanosti med dvema spremenljivkama (slika 4).

Slika 4: Povezanost med odvisno in neodvisno spremenljivko
(Vir: Montgomery idr., 2021)

Kazalci enostavne linearne regresije

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

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

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

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

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

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

Standardna napaka ocene odvisne spremenljivke (σ_{ey}) pokaže, ali na variabilnost spremenljivke y, razen spremenljivke x, vplivajo še druge spremenljivke in slučajni vplivi (Seber in Lee, 2003).

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

Primer rešene naloge:

V preglednici so podatki o povprečni porabi časa za točenje goriva v minutah (x) in povprečnem znesku za gorivo v d.e. (y):

Povprečen čas v minutah (x)	5,6	7,4	6,8	9,3	10,4	8,5
Povprečen znesek v d.e. (y)	12,4	33,8	20,8	45,8	55,7	50,3

a) Izračunajte in pojasnite vse kazalce linearne korelacije in regresije.
b) Z verjetnostjo 95,4 % ocenite povprečen znesek za gorivo pri povprečni porabi časa 10 minut.

a) *Izračun kazalcev linearne korelacije in regresije (v obrazcih št. 2.2. ali 2.3, 2.4, 2.5, 2.6, 2.8 ali 2.9, 2.10):*

Obe oceni regresijskih koeficientov izračunamo na osnovi obrazcev *sistema normalnih enačb za izračun regresijskih koeficientov* oziroma enačb, ki so iz tega sistema izpeljane.

Preden se lotimo izračuna kazalcev linearne korelacije in regresije, izračunamo:

$\Sigma\ x_i = 5{,}6 + 7{,}4 + 6{,}8 + 9{,}3 + 10{,}4 + 8{,}5 = 48$

$\Sigma\ x_i^2 = 5{,}6^2 + 7{,}4^2 + 6{,}8^2 + 9{,}3^2 + 10{,}4^2 + 8{,}5^2 = 399{,}26$

$\Sigma\ y_i = 12{,}4 + 33{,}8 + 20{,}8 + 45{,}8 + 55{,}7 + 50{,}3 = 218{,}8$

$\Sigma\ y_i^2 = 12{,}4^2 + 33{,}8^2 + 20{,}8^2 + 45{,}8^2 + 55{,}7^2 + 50{,}3^2 = 9459{,}06$

$\Sigma\ x_i \cdot y_i = (5{,}6 \cdot 12{,}4) + (7{,}4 \cdot 33{,}8) + (6{,}8 \cdot 20{,}8) + \ldots + (8{,}5 \cdot 50{,}3) = 1893{,}77$

$\bar{x} = \frac{48}{6} = 8$

$\bar{y} = \frac{218{,}8}{6} = 36{,}47$

Izračun kovariance (v obrazcih št. 3.5):

$$c_{xy} = \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \bar{x} \cdot \bar{y}$$

$C_{xy} = \frac{1}{6} \cdot 1893{,}77 - 8 \cdot 36{,}47 = 23{,}868$

Oba regresijska koeficienta a in b izračunamo tako (v obrazcih št. 2.2 ali 2.3 in 2.4):

$$b = \frac{c_{xy}}{\sigma_x^2}$$

$$b = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} = \frac{\frac{1}{N}\sum_{i=1}^{N} x_i y_i - \bar{x} \cdot \bar{y}}{\frac{1}{N}\sum_{i=1}^{N} x_i^2 - \bar{x}^2}$$

$$b = \frac{\frac{1}{6} \cdot 1893{,}77 - 8 \cdot 36{,}47}{\frac{1}{6} \cdot 399{,}26 - 8^2} = 9{,}38$$

$$a = \bar{y} - b\bar{x}$$

a = 36,47 – 9,38 · 8 = –38,57

Izračunana regresijska koeficienta vstavimo v enačbo regresijske premice:

$$\hat{y} = a + b \cdot x$$
$$\hat{y} = -38{,}57 + 9{,}38\,x$$

Pomen regresijskega koeficienta a: Regresijska konstanta a nima vedno smiselnega vsebinskega pomena. Tako je tudi v tem primeru, saj pri času točenja 0 minut povprečni znesek za plačilo goriva ne bo negativna vrednost, pač pa bo enak 0.
Pomen regresijskega koeficienta b: Če se povprečna poraba časa za točenje goriva *(x)* poveča za eno enoto (v minutah), se povprečen znesek goriva *(y)* v povprečju poveča za 9,38 d.e.

Korelacijski koeficient izračunamo po obrazcu (v obrazcih št. 2.9):

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y}$$

Za izračun σ_x in σ_y uporabimo enačbo za *varianco iz nerazvrščenih vrednosti* (v obrazcih št. 1.34):

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

$$\sigma_x^2 = \frac{1}{6} \cdot [(5{,}6 - 8)^2 + (7{,}4 - 8)^2 + \ldots + (8{,}5 - 8)^2] = 15{,}26$$

$$\sigma_x = \sqrt{15{,}26} = 3{,}91$$

$$\sigma_y^2 = \frac{1}{6} \cdot [(12{,}4 - 36{,}47)^2 + (33{,}8 - 36{,}47)^2 + \ldots + (50{,}3 - 36{,}47)^2] = 1480{,}1534$$

$$\sigma_y = \sqrt{1480{,}1534} = 38{,}47$$

$$r_{xy} = 9{,}38 \cdot \frac{3{,}91}{38{,}47} = 0{,}953$$

Na osnovi rezultata (r_{xy} = 0,953) vidimo, da obstaja močna povezanost med odvisno (povprečen znesek za gorivo) in neodvisno spremenljivko (povprečna poraba časa za točenje goriva). Smer povezanosti je pozitivna.

Izračun determinacijskega koeficienta:

$r_{xy}^2 = 0{,}953^2 = 0{,}908$ oz. 90,8 %

Delež pojasnjene variance v skupni varianci za odvisno spremenljivko znaša 90,8 %. Standardno napako ocene odvisne spremenljivke izračunamo po obrazcu (v obrazcih št. 2.10):

$$\sigma_{ey} = \sigma_y \sqrt{1 - r_{xy}^2}$$

$$\sigma_{ey} = 38{,}47 \cdot \sqrt{1 - 0{,}908} = 11{,}67$$

Standardna napaka ocene odvisne spremenljivke je različna od 0, kar pomeni, da na povprečen znesek za gorivo (odvisna spremenljivka) poleg povprečne porabe časa za točenje goriva (neodvisna spremenljivka) vplivajo še druge spremenljivke in slučajni vplivi.
b)

Izračun povprečnega zneska za gorivo pri povprečni porabi časa 10 minut z verjetnostjo 95,4 %:

$x = 10$

$\hat{y} = -38{,}57 + 9{,}38\,x$

$\hat{y}_{x=10} = -38{,}57 + 9{,}38 \cdot 10$

$\hat{y}_{x=10} = 55{,}29$ d.e.

$P(55{,}29 - 2 \cdot 11{,}67 < y_{x=10} < 55{,}29 + 2 \cdot 11{,}67) = 95{,}4\ \%$

> Pri domnevi o normalni porazdelitvi na primer velja, da se v razmiku $\hat{y}_x \pm 2 \cdot \sigma_{ey}$ nahaja **95,4 %** vseh vrednosti.
>
> $\mathbf{P(\hat{y}_x - 2 \cdot \sigma_{ey} < y_x < \hat{y}_x + 2 \cdot \sigma_{ey}) = 0{,}954 = 95{,}4\ \%}$
>
> <u>Vsebinski pomen:</u> S 95,4-% verjetnostjo lahko trdimo, da spremenljivka y pri dani vrednosti spremenljivke $x = x_0$ zavzame vrednost v mejah zapisanega intervala.
>
> Na primer, pri domnevi o normalni porazdelitvi ocen se v razmiku $\hat{y}_x \pm 1 \cdot \sigma_{ey}$ nahaja **68,3 %** vseh vrednosti. Medtem ko se pri domnevi o normalni porazdelitvi v razmiku $\hat{y}_x \pm 3 \cdot \sigma_{ey}$ nahaja **99,7 %** vseh vrednosti.

$P(31{,}95 < y_{x=10} < 78{,}63) = 95{,}4\ \%$

Pri povprečni porabi časa za točenje goriva x = 10 (v minutah) bo povprečen znesek za gorivo med 31,95 d.e. in 78,63 d.e., kar trdimo s 95,4-% verjetnostjo.

Naloga 30

Za šest podjetij so podatki o investicijah v novo tehnologijo (v 10^6 evrov) ter ustvarjenemu dobičku (v 10^3 evrov) v letu X naslednji:

Podjetje	A	B	C	D	E	F
Investicije v tehnologijo	115	130	140	149	160	171
Ustvarjen dobiček	328	330	390	361	421	400

a) S prikazom dvojic vrednosti opazovanih spremenljivk v razsevnem grafikonu določite obliko, smer in jakost odvisnosti med spremenljivkama.

b) Izračunajte parametre regresijske premice, izračunano regresijsko premico vrišite v razsevni grafikon.

c) Ocenite višino dobička za podjetje, ki bi investiralo v tehnologijo $x = 180$ (v 10^6 evrov), točkovna ocena.

d) Izračunajte parameter, na osnovi katerega določite smer in jakost linearne korelacijske odvisnosti.

e) Izračunajte delež pojasnjene variance v skupni varianci za odvisno spremenljivko.

f) Izračunajte standardno napako ocene odvisne spremenljivke.

g) Ob upoštevanju linearne korelacijske odvisnosti ocenite z verjetnostjo 95,4-% višino dobička pri $x = 180$ (v 10^6 evrov).

Naloga 31

V banki "X" so ob različni obrestni meri (x) zabeležili naslednje zneske kratkoročnih oblik varčevanja (y):

Obrestna mera v %	(x)	3,4	4,2	5,6	6,4	8
Varčevanje (v d.e.)	(y)	123	165	197	234	258

Ocenite z zanesljivostjo 95,4 % znesek kratkoročnih oblik varčevanja pri obrestni meri $x = 7$ % ter pojasnite vse kazalce linearne regresije in korelacije, ki ste jih izračunali.

Naloga 32

Opazovali smo osem sodnikov glede na višino osebnih dohodkov (y) in dolžino delovne dobe (x). Ob upoštevanju linearne korelacijske odvisnosti in podatkov, ki so podani spodaj, ocenite osebni dohodek tistega sodnika, ki ima 20 let delovne dobe. Oceno napravite z verjetnostjo 0,954.

$C_{xy} = 270$ $\quad$ $VAR(x) = 26{,}9$ $\quad$ $VAR(y) = 3063{,}7$ $\quad$ $\Sigma xi = 130$ $\quad$ $\Sigma yi = 1680$

Naloga 33

Pojasnite:

a) smer in jakost povezanosti med odvisno in neodvisno spremenljivko, če je $r_{xy} = -0,8$;

b) determinacijski koeficient, če je $r_{xy}^2 = 0,90$;

c) velikost standardne napake ocene (σ_{ey}), če je $r_{xy} = -1$.

Naloga 34

Domnevamo, da je znesek stroškov na delavca (v d.e.), ki ga podjetja namenjajo za okoljevarstvene dejavnosti, odvisen od velikosti podjetja (merjeno v številu zaposlenih). Za pet podjetij so podatki v preglednici.

Število zaposlenih	25	46	120	91	37
Znesek stroškov/delavca	15	17	35	30	25

a) Narišite razsevni grafikon in ga pojasnite.
b) Izračunajte in pojasnite vse kazalce linearne korelacije in regresije.
c) Ocenite z zanesljivostjo 95,4 % znesek stroškov za okoljevarstvene dejavnosti podjetja, če ima podjetje 180 zaposlenih.

Naloga 35

V preglednici so podatki o zasedenosti sob v hotelih ter njihovih cenah:

Povprečni % zasedenosti	Cena v dolarjih
78,70	80,99
68,70	82,48
63,90	76,50
71,70	101,76
70,80	100,25
80,50	169,19
66,90	78,74

Izračunajte koeficiente linearne regresijske funkcije in ocenite, kolikšno ceno lahko pričakujemo za sobo, ki je povprečno zasedena 95 %. Intervalno oceno napravite s 95,4-odstotno verjetnostjo.

Naloga 36

V preglednici so podatki o času (v minutah), ki ga študenti porabijo za učenje zadnji dan pred izpitom Poslovna statistika (x) in njihovi oceni izpita (y):

Poraba časa (v min)	Ocena
180	8
35	4
95	7
140	10
125	9
60	7
75	6
12	5

a) Ocenite osnovne značilnosti povezanosti med spremenljivkama ter izračunajte in pojasnite vse kazalce linearne korelacije in regresije.
b) Izpišite enačbo regresijske premice.

Naloga 37

V preglednici so podatki o tedenskem številu potrjenih primerov s SARS-CoV-2 glede na dve različni lokaciji prenosa okužbe, in sicer delovno mesto (spremenljivka y) in trgovina (spremenljivka x):

Teden	Delovno mesto	Trgovina
1. (2020–47)	257	186
2. (2020–48)	260	216
3. (2020–49)	274	221
4. (2020–50)	269	266
5. (2020–51)	233	227
6. (2020–52)	178	156
7. (2020–53)	127	174

a) Narišite razsevni grafikon in ga pojasnite.
b) Izračunajte regresijski koeficient b in pojasnite rezultat.
c) Izračunajte korelacijski koeficient in pojasnite rezultat.

Naloga 38

Razpolagamo s podatki o oceni motiviranosti zaposlenih (na lestvici od 1 do 100) in različnem izplačilu denarne nagrade (v €) za šest zaposlenih v podjetju X. Zanima nas, ali višina izplačila denarne nagrade vpliva na povečanje motiviranosti zaposlenih.

Zaposlen	A	B	C	D	E	F
Ocena motiviranosti zaposlenega	26	35	63	78	88	95
Izplačilo denarne nagrade (v €)	150	180	225	260	285	300

a) prikazom dvojic vrednosti opazovanih spremenljivk v razsevnem grafikonu določite obliko, smer in jakost odvisnosti med spremenljivkama.
b) Izračunajte parametre regresijske premice, izračunano regresijsko premico vrišite v razsevni grafikon.
c) Ocenite višino motivacije zaposlenih pri izplačilu denarne nagrade x = 310 €, točkovna ocena.
d) Izračunajte parameter, na osnovi katerega določite smer in jakost linearne korelacijske odvisnosti.
e) Izračunajte delež pojasnjene variance v skupni varianci za odvisno spremenljivko.
f) Izračunajte standardno napako ocene odvisne spremenljivke.
g) Ob upoštevanju linearne korelacijske odvisnosti ocenite z verjetnostjo 95,4 % višino motivacije zaposlenih pri izplačilu denarne nagrade x = 295 €.

5 OSNOVE VZORČENJA

Na podlagi vzorčnih podatkov verjetnostnega vzorca lahko opravimo postopek statističnega sklepanja, ko rezultat iz vzorca posplošimo na statistično množico. Ta del statistike se imenuje *inferenčna statistika*. Vzorčni podatki nam služijo za izračun ocene statističnega parametra v statistični množici. Zavedamo se, da ta *vzorčna oziroma točkovna ocena* najverjetneje ni natančna vrednost parametra v statistični množici, ampak je le njegova ocena. Po izračunu vzorčne ocene statističnega parametra zato izračunamo *intervalno oceno,* imenovano *interval zaupanja*, z določeno verjetnostjo, ki ji rečemo *stopnja zaupanja* (Heumann idr., 2016).

Postopek, ki ga uporabimo za izračun intervala zaupanja, je odvisen od želene stopnje zaupanja, in od tega, ali razpolagamo z informacijami o porazdelitvi obravnavane spremenljivke v statistični množici (na primer znan standardni odklon v statistični množici) ter od vzorca in njegove velikost (Holmes idr., 2018).

V okviru predmeta Poslovna statistika se ukvarjamo z *intervali zaupanja za aritmetično sredino, strukturni odstotek in total* v velikih in malih vzorcih.

Poleg intervalov zaupanja za oceno vrednosti statističnega parametra v statistični množici se v tem vsebinskem sklopu ukvarjamo tudi z *preverjanjem hipoteze* o vrednosti določenega statističnega parametra. Testiranje hipoteze vključuje zbiranje podatkov iz slučajnega vzorca in vrednotenje zbranih podatkov. Na podlagi analize podatkov se odločimo, ali obstaja dovolj informacij in ustrezne okoliščine za zavrnitev ničelne hipoteze *(H0)* ali ne.

Ničelna hipoteza, H0, je trditev, da je vrednost aritmetične sredino (ali strukturnega odstotka ali totala) statistične množice enaka neki vnaprej določeni vrednosti (v okviru predmeta Poslovna statistika se ukvarjamo samo s preverjanjem hipotez takega tipa).

Alternativna ali raziskovalna hipoteza, H1, pa je trditev o vrednosti statističnega parametra v statistični množici, ki je nasprotna trditvi H0, in menimo, da drži takrat, ko H0 lahko zavrnemo. Za to, da zavrnemo ničelno hipotezo, običajno potrebujemo 90 ali več odstotno verjetnost, da je to pravilna odločitev, oziroma 10 ali manj odstotno verjetnost, da se zmotimo, če H0 zavrnemo, kar imenujemo *stopnja tveganja ali stopnja značilnosti preizkusa*. Običajno pri preizkušanju hipoteze H0 uporabljamo kot najvišjo še dopustno petodstotno stopnjo tveganja (maksimalna dopustna stopnja tveganja).

Primer rešene naloge:

Želeli smo ugotoviti, koliko časa na teden porabijo delodajalci za ocenjevanje inovativnih predlogov svojih zaposlenih. V vzorec smo vključili 50 delodajalcev, ki so izjavili, da povprečen porabljen čas na teden za ocenjevanje inovativnih predlogov znaša šest ur. Vrednost vzorčnega standardnega odklona pa znaša 2,5 ur. S 95-% verjetnostjo ocenite povprečno tedensko porabo časa delodajalcev za ocenjevanje inovativnih predlogov svojih zaposlenih.

Imamo velik vzorec, ker je n = 50, ter želimo izvesti dvostransko intervalno ocenjevanje aritmetične sredine:

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

Kritične vrednosti za spremenljivko z:

Stopnja tveganja α	Dvostransko ocenjevanje
10 %	± 1,645
5 %	± 1,96
1 %	± 2,58

Izračunamo standardno napako ocene aritmetične sredine (v obrazcih št. 4.4):

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

$se_{\bar{y}} = \frac{2,5}{\sqrt{50}} = 0,354$ ur

Uporabimo obrazec za dvostransko ocenjevanje aritmetične sredine (v obrazcih št. 4.1):

$$\bar{Y} - z \cdot SE_{\bar{y}} < \bar{y} < \bar{Y} + z \cdot SE_{\bar{y}}$$

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

$$P(6 - 1,96 \cdot 0,354 < \bar{y} < 6 + 1,96 \cdot 0,354) = 95\ \%$$

$$P(5,306 < \bar{y} < 6,694) = 95\ \%$$

S 95-odstotno verjetnostjo ocenjujemo, da znaša v statistični množici povprečen tedenski čas delodajalcev za ocenjevanje inovativnih predlogov zaposlenih med 5,306 in 6,694 ur.

Primer rešene naloge:

Kot smo že v prejšnji nalogi omenili, smo v vzorec vključili 50 delodajalcev, ki so izjavili, da povprečen porabljen čas na teden za ocenjevanje inovativnih predlogov znaša šest ur. Vrednost standardnega odklona pa znaša 2,5 ur. Z desetodstotnim tveganjem preizkusite domnevo, da je povprečen porabljen čas na teden za ocenjevanje inovativnih predlogov enak 4,5 ure.

H0: $\bar{y}_D = 4,5$ ur

H1: $\bar{y}_D \neq 4,5$ ur

$\bar{y} = 6$ ur

$s = 2,5$ ur

$se_{\bar{y}} = \frac{2,5}{\sqrt{50}} = 0,354$ ur

$\alpha = 10\ \%$ (oziroma 0,10), kar pomeni, da je kritična vrednost spremenljivke $z = \pm 1,645$
Izračunamo z po enačbi (izračun testne vrednosti pri preizkušanju domneve o aritmetični sredini – z-test, v obrazcih št. 4.22):

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

$z = \frac{6-4,5}{0,354} = 4,23$ (dobljena standardizirana vrednost vzorčne vrednosti ne pade v interval $\pm$ 1,645, pač pa izven intervala)

Na osnovi rezultata vidimo, da povprečen porabljen čas na teden za ocenjevanje inovativnih predlogov ni enak 4,5 ure, tako da zavrnemo domnevo H0 in sprejmemo raziskovalno domnevo H1 z manj kot desetodstotnim tveganjem oziroma verjetnostjo, da smo se pri tem našem zaključku zmotili.

Naloga 39

V slučajnem vzorcu n = 200 kupcev je bila povprečna poraba izdelka 20,8 kosov izdelka v časovni enoti, nepristranska ocena variance pa 38,44 kosov^2. Izračunajte 95-odstotni interval zaupanja za povprečno porabo izdelka v osnovni statistični množici.

Naloga 40

V šestih naključno izbranih podjetjih neke regije je med zaposlenimi tak odstotek žensk: 23 %, 47 %, 5 %, 45 %, 65 % in 12 %. Izračunajte in pojasnite interval zaupanja za povprečen odstotek žensk med zaposlenimi v statistični množici. Oceno napravite s petodstotnim tveganjem.

Naloga 41

Na zavodu za zaposlovanje, na katerem je prijavljenih 10.000 nezaposlenih oseb, so želeli ugotoviti odstotek (%) nezaposlenih, ki prejemajo plačila na osnovi »dela na črno«. Iz slučajnega vzorca 625 nezaposlenih so ugotovili, da 125 ljudi dela na črno. Upoštevajte stopnjo tveganja 0,20.

Naloga 42

Na ravni značilnosti α = 0,10 preizkusite domnevo, da je povprečna višina denarnih sredstev 120 vlagateljev 5000 EUR, če smo iz te skupine naključno izbrali 36 vlagateljev in zanje ugotovili, da je:

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

Naloga 43

S strojem proizvajamo nek izdelek. Stroj naj bi bil nastavljen tako, da je povprečna dolžina izdelkov 20 cm. Na osnovi slučajnega vzorca n = 5 izdelkov smo ugotovili, da je vzorčna aritmetična sredina 20,3 cm in s = 0,2 cm. Upoštevajte, da je α = 1 %. Ali rezultati kažejo, da stroj ni pravilno nastavljen?

Naloga 44

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

Znesek v 10^2 EUR	Število oseb
od 10 do pod 20	45
od 20 do pod 40	132
od 40 do pod 80	92
od 80 do pod 120	41
od 120 do pod 200	22
od 200 do pod 500	8
Skupaj	340

a) S 95-% verjetnostjo določite interval zaupanja za povprečni znesek, ki so ga v vrednostne papirje vložili imetniki vrednostnih papirjev pri tej banki.

b) Če je v opazovani banki registriranih 3648 lastnikov vrednostnih papirjev, ocenite z 90-% verjetnostjo skupni znesek vlog v vrednostne papirje pri tej banki.

c) Na ravni značilnosti $\alpha = 0{,}10$ preverite domnevo, da je povprečen znesek, ki so ga lastniki vložili v vrednostne papirje, enak $75 \cdot 10^2$ EUR.

d) Z 99-% verjetnostjo preizkusite domnevo, da je % imetnikov vrednostnih papirjev, ki imajo zneske vrednostnih papirjev od $40 \cdot 10^2$ do $120 \cdot 10^2$ EUR, 35 % in nato še izračunajte, da jih je največ 35 %.

e) Z 80-% verjetnostjo preizkusite domnevo, da je skupni znesek vseh imetnikov vrednostnih papirjev skupaj enak 20.000.000 EUR.

Naloga 45

V knjižnem klubu so za devet naključno izbranih oseb zbrali podatke o višini letnih zneskov za nakup knjig. Zneski so:

v EUR	600	800	1200	900	400	800	700	500	400

a) Z 80-% verjetnostjo ocenite povprečen letni znesek za nakup knjig.

b) Z 10-% tveganjem preizkusite domnevo, da je skupni znesek za nakup knjig v opazovanem knjižnem klubu, ki ima skupno 2860 članov, 1.800.000 EUR.

Naloga 46

Ali lahko trdimo, da delavci izdelajo povprečno 100 izdelkov v časovni enoti, če smo za 50 slučajno izbranih delavcev ugotovili, da izdelajo povprečno 102 izdelka v časovni enoti in je nepristranska ocena variance 6,25? Upoštevajte stopnjo tveganja 0,05.

Naloga 47

V preglednici so podatki o številu delavcev v podjetjih neke panoge, $n = 20$, ki smo jih zajeli v slučajni vzorec.

Število delavcev	Število podjetij
od 1 do 5	3
od 6 do 10	4
od 11 do 15	2
od 16 do 20	1
od 21 do 25	1
od 26 do 30	5
od 31 do 35	2
od 36 do 40	2

Z 99-odstotno verjetnostjo določite povprečno število zaposlenih delavcev v vseh podjetjih te gospodarske panoge.

Naloga 48

V nekem podjetju so na osnovi naključnega vzorca 150 ljudi ugotovili, da je njihova povprečna zamuda na delo 12,5 minut, nepristranska ocena standardnega odklona pa znaša 5 minut.

a) Izračunajte in pojasnite 85-odstotni interval zaupanja za povprečno zamujen čas vseh delavcev.
b) Pri enoodstotni stopnji značilnosti preizkusa preverite domnevo, da je povprečna zamuda delavcev v podjetju enaka 10 minut.

Naloga 49

V slučajnem vzorcu smo zajeli deset študentov neke fakultete, za katere velja, da v povprečju porabijo 45 % svojega prostega časa za športne aktivnosti, standardni odklon pa je enak 25 %. Ali lahko trdimo, da študenti te fakultete porabijo v povprečju polovico svojega prostega časa za športne aktivnosti? Upoštevajte enoodstotno tveganje.

Naloga 50

Naključno smo anketirali 40 gospodinjstev. Skupni dohodek vseh 40 gospodinjstev skupaj je bil 2.750.000 d.e., nepristranska ocena standardnega odklona pa 460 d.e. S 95-% verjetnostjo ocenite povprečni dohodek na gospodinjstvo v statistični množici.

Naloga 51

Raziskovalci so v slučajni vzorec zajeli n = 7 okuženih oseb s SARS-CoV-2. Zanimalo jih je, koliko dni so bile prisotne posledice slabega počutja s SARS-CoV-2 naključno izbranih oseb.

Podatki o slabem počutju v številu dni okuženih oseb s SARS-CoV-2 so v preglednici:

Oseba	Število dni
1	25
2	30
3	21
4	60
5	17
6	14
7	24

a) Določite 80-% interval zaupanja za povprečno število dni slabega počutja pri okuženih osebah s SARS-CoV-2.

b) Pri stopnji tveganja 5 % preizkusite domnevo, da je povprečno število dni slabega počutja pri okuženih osebah s SARS-CoV-2 največ 30 dni.

Naloga 52

Vsebnost nikotina v štirih cigaretah neke znamke, merjena v miligramih, je 21, 19, 23 in 19. Pri stopnji tveganja 5 % preizkusite domnevo, da je povprečna vsebnost nikotina v cigaretah opazovane znamke enaka 22 miligramom.

Naloga 53

V trgovini X so za naključno izbranih pet oseb zbrali podatke o višini celotnega nakupa na določen dan. Podatki so naslednji:

Kupec	Znesek v EUR
1.	186
2.	216
3.	221
4.	266
5.	227
Skupaj	1116

a) Določite 95-% interval zaupanja za povprečno zapravljen znesek nakupa kupcev v trgovini X.
b) Pri stopnji tveganja 10 % preizkusite domnevo, da je povprečno zapravljen znesek kupcev v trgovini X več kot 220 EUR.

6 ČASOVNE VRSTE

Časovna vrsta je niz istovrstnih podatkov, ki se nanašajo na zaporedne časovne razmike ali trenutke. Dinamiko v časovnih vrstah lahko analiziramo s pomočjo relativnih števil, ki jih imenujemo *indeksi*. Indeksi so v statistiki relativna števila, s katerimi primerjamo za proučevani pojav medsebojno dvoje ali več istovrstnih podatkov (Mišić, 2019). Primerjani podatki morajo biti izraženi v istih merskih enotah. V primeru, da primerjamo med seboj le dva podatka, govorimo o enostavnih indeksih. Z indeksi dobimo zelo nazorno sliko o velikosti relativnih sprememb pojava v času oz. o velikosti relativnih razlik za pojav v prostoru (Ralph idr., 2015).

Pri predmetu Poslovna statistika obravnavamo enostavne indekse in ločimo:

a) indekse s stalno osnovo,
b) verižne indekse,
c) koeficiente dinamike in stopnje rasti.

Indeksi s stalno osnovo izražajo posredno odstotno spremembo med vrednostmi členov statistične vrste, Y_t, ki jih primerjamo z enim od členov, ki ga izberemo za osnovo, ali bazo, Y_0, medtem ko *koeficient dinamike* izraža razmerje med dvema zaporednima členoma osnovne statistične vrste, *stopnja rasti* pa pokaže neposredno odstotno spremembo med dvema zaporednima členoma osnovne statistične vrste – pove nam, za koliko odstotkov se vrednost člena Y_t razlikuje od vrednosti člena Y_{t-1} (Tominc, 2016). (Tominc, 2016). *Verižni indeks* je koeficient dinamike, izražen v odstotkih.

Povprečno vrednost v časovni statistični vrsti opredelimo s *povprečno stopnjo rasti*, ki se izraža v odstotkih (na primer, povprečna stopnja rasti plač v Sloveniji v zadnjih 10 mesecih, povprečna stopnja rasti košarice izbranih cen v zadnjem letu ipd.). Povprečno stopnjo rasti v časovni statistični vrsti izračunamo bodisi iz *povprečnega koeficienta dinamike* ali *povprečnega verižnega indeksa*. Pri tem izhajamo iz postopka izračuna *geometrijske sredine* (in ne aritmetične sredine). Povprečno stopnjo rasti opazovane spremenljivke v preteklem obdobju lahko uporabimo tudi za ocenjevanje vrednosti spremenljivke v prihodnjih časovnih enotah – *napovedovanje vrednosti spremenljivke* v prihodnjih časovnih enotah.

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

Primer rešene naloge:

V preglednici so podatki o številu nočitev tujih turistov v gorskem kraju v Sloveniji, v hotelu X v petih letih:

Leto	2018	2019	2020	2021	2022
Št. tujih turistov	2500	2300	1570	2650	2910

a) Za koliko odstotkov se je število tujih turistov v letu 2021 razlikovalo od števila v predhodnem letu?

Izračunati je potrebno verižni indeks za leto 2021, v %, upoštevajoč enačbo (v obrazcih št. 1.8):

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

$$V_{2021} = \frac{2650}{1570} \cdot 100 = 168{,}79\ \%$$

Število tujih turistov v letu 2021 je enako 168,79 % števila tujih turistov v predhodnem letu (v letu 2020).

b) Ocenite število nočitev tujih turistov v hotelu X v letu 2025 z upoštevanjem povprečne stopnje rasti.

Uporabimo enačbo za *povprečni koeficient dinamike* (v obrazcih št. 1.26):

$$K = \sqrt[T-1]{\frac{Y_T}{Y_1}}$$

$$\mathrm{K} = \sqrt[4]{\frac{2910}{2500}} = \sqrt[4]{1{,}164} = 1{,}039$$

$$\mathrm{S} = (\mathrm{K} - 1) \cdot 100 = (1{,}039 - 1) \cdot 100 = 3{,}9\ \%$$

Število nočitev tujih turistov v hotelu X se je v petih letih povečevalo povprečno za 3,9 % na leto.

$$Y_{2025} = 2910 \cdot K^3 = 2910 \cdot 1{,}039^3 = 3263{,}92$$

V letu 2025 na osnovi povprečne stopnje rasti ocenjujemo, da bo število turistov enako 3263,92 oziroma 3264.

c) Izračunajte in pojasnite linearno funkcijo trenda ter ocenite število prenočitev tujih turistov v hotelu X v letu 2023.

Linearna funkcija trenda:

Sistem normalnih enačb (v obrazcih št. 3.1):

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

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

Leto	t	Yt	t·Yt	t^2
2018	1	2.500	2.500	1
2019	2	2.300	4.600	4
2020	3	1.570	4.710	9
2021	4	2.650	10.600	16
2022	5	2.910	14.550	25
Skupaj	15	11.930	36.960	55

$$5a + 15b = 11.930 \quad / \cdot (-3)$$
$$15a + 55b = 36.960$$
$$0 + 10b = 1.170$$

$$\boxed{\begin{array}{l}(-3) \cdot 5 + 15 = 0 \\ (-3) \cdot 15 + 55 = 10 \\ (-3) \cdot 11.930 + 36.960 = \end{array}}$$

$$b = \frac{1170}{10} = 117$$

$$5a + 15 \cdot 117 = 11.930$$

$$5a + 1.755 = 119.30$$

$$5a = 10.175$$

$$a = \frac{10175}{5} = 2.035$$

Funkcija trenda: $\hat{Y} = 2.035 + 117 \cdot t$

Napoved za število tujih turistov za 6. leto: $\hat{Y}_t = 2.035 + 117 \cdot 6 = 2.737$

Naloga 54

Število in naravno gibanje prebivalstva na opazovanem področju v obdobju zadnjih sedmih let je opisano v preglednici – (vsi podatki so v 10^4 ljudi):

Leto	Srednje štev. preb. (30.6.)	Živorojeni	Umrli
1	198	1,94	1,93
2	198	1,89	1,89
3	199	1,87	1,86
4	198	1,82	1,89
5	198	1,78	1,90
6	198	1,75	1,88
7	199	1,81	1,85

a) Analizirajte relativne spremembe v številu živorojenih v opazovanem obdobju:
   - glede na leto 3,
   - od leta do leta.

b) Izračunajte koeficiente rodnosti (natalitete – število živorojenih otrok na 1000 prebivalcev) in umrljivosti (mortalitete – število umrlih na 1000 prebivalcev). Komentirajte rezultate.

Naloga 55

V podjetju X so sprejeli študente, ki so opravljali prakso. Podani so podatki o tem, koliko študentov je opravljalo prakso v podjetju X v obdobju šestih let. Podatki v obliki indeksov s stalno osnovo v letu 2017 so podani v preglednici:

Leto	2017	2018	2019	2020	2021	2022
$I_{t/2017}$	100	110	125	117	120	118

a) Izračunajte število študentov po letih, ki so v podjetju X opravljali prakso, če je leta 2017 opravljalo prakso 130 študentov.
b) Izračunajte in vsebinsko pojasnite povprečno letno stopnjo rasti števila študentov, ki so opravljali prakso v teh šestih letih.

Naloga 56

Letne stopnje rasti pridelave mesa v kg na prebivalca na opazovanem področju v obdobju zadnjih devetih let so bile:

Leto	1	2	3	4	5	6	7	8	9
S_t%	-	+18,0	-2,1	+2,7	+2,5	+2,5	-5,0	+7,0	-5,8

a) Pojasnite največjo pozitivno in največjo negativno stopnjo rasti opazovanega pojava.

b) Zapišite in pojasnite relativne spremembe v pridelavi mesa v opazovanih letih:

- z vrsto verižnih indeksov,
- z vrsto letnih koeficientov dinamike,
- z vrsto indeksov s stalno osnovo v letu 4.

c) Če je bila pridelava mesa na prebivalca v letu 4 enaka 84,5 kg, izračunajte pridelavo mesa na prebivalca v preostalih letih.

Naloga 57

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

Mesec	I	II	III	IV	V
Vrednost proizvodnje v 10^6 EUR	652	730	840	752	-
Štev. zaposlenih na začetku meseca	214	240	226	208	200

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

Naloga 58

Za opazovano organizacijo imate na razpolago podatke o povprečni plači zaposlenih v prvih petih mesecih leta ter o mesečni stopnji rasti vrednosti prodaje:

Mesec	I	II	III	IV	V
Povprečna mesečna plača zaposlenih (v d.e.)	2600	2800	3100	2900	3200
Mesečna stopnja rasti vrednosti prodaje S_t %	–	+15	–10	+8	–12

a) Izračunajte in pojasnite, kako so se relativno spreminjale plače glede na zadnji opazovani mesec.
b) Izračunajte in pojasnite, kako se je spreminjala vrednost prodaje glede na prvi opazovani mesec.
c) Ali so se v opazovanem podjetju v povprečju bolj dvigovale plače ali vrednost prodaje?
d) Napovejte povprečno mesečno plačo zaposlenih v oktobru, upoštevajoč povprečno stopnjo rasti plač v prvih petih opazovanih mesecih.

Naloga 59

V preglednici so podatki o številu izdanih gradbenih dovoljenj na nekem območju:

Leto	1	2	3	4	5	6	7
Št. izdanih gradb. dovoljenj	115	224	118	400	350	320	300

a) Za koliko odstotkov se je število izdanih gradbenih dovoljenj v letu 4 razlikovalo od števila v predhodnem letu? Kako imenujemo izračunano vrednost?

b) Koliko odstotkov števila izdanih gradbenih dovoljenj iz leta 1 predstavlja število izdanih dovoljenj v letu 7? Kako imenujemo izračunano vrednost?

c) Statistično vrsto prikažite grafično.

Naloga 60

V preglednici so podatki o stopnji rasti plač zaposlenih oseb v nekem podjetju.

Leto	1	2	3	4
S_t %	/	+18,4	+11,7	–9,2

a) Analizirajte, kako so se relativno spreminjale plače glede na leto 2.

b) Izračunajte, kolikšne so bile povprečne plače zaposlenih v obravnavanih letih, če je bila plača leta 3 enaka 170,5 d.e.

c) Izračunajte in pojasnite povprečno stopnjo rasti plač v obravnavani 4-letni časovni vrsti.

d) Napovejte povprečno raven plač zaposlenih v letu 7, če upoštevate povprečno stopnjo rasti v opazovanem obdobju.

Naloga 61

V preglednici so podatki o številu študentov, ki so se vpisali na podiplomski študij za šest zaporednih let.

Leto	1	2	3	4	5	6
Število študentov	215	260	280	300	310	325

a) Kako imenujemo statistično vrsto v preglednici? Statistično vrsto grafično prikažite.

b) Ocenite število študentov v 9. zaporednem letu z upoštevanjem povprečne stopnje rasti.

Naloga 62

V preglednici so podatki o spreminjanju števila prodanih izdelkov v podjetju X, v obliki indeksnega števila od leta 2018 do leta 2023:

Leto	2018	2019	2020	2021	2022	2023
V_t	/	70	100	85	95	110

a) Kako imenujemo statistično vrsto?

b) Vsebinsko pojasnite indeksno število za leto 2021.

c) Izračunajte število prodanih izdelkov po letih, če je bilo leta 2020 število prodanih izdelkov enako 415. Časovno vrsto prikažite grafično.

Naloga 63

V preglednici so podatki o številu upravnih odločb v obdobju petih let:

Leto	1	2	3	4	5
Število upravnih odločb	100	124	245	300	320

a) Kako imenujemo statistično vrsto v preglednici? Statistično vrsto prikažite grafično.

b) Izračunajte in pojasnite linearno funkcijo trenda ter ocenite število izdanih gradbenih dovoljenj v prvem prihodnjem letu.

Naloga 64

Na opazovanem področju smo zabeležili naslednje število nočitev tujih gostov:

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

a) Letne podatke, ki prikazujejo skupno letno število nočitev, narišite v linijskem grafikonu, s prostoročno metodo določite osnovno smer razvoja pojava; z analitično metodo določite parametre funkcije; ocenite število nočitev v prvem prihodnjem letu.

b) Izračunajte sezonske indekse, upoštevajte oceno za število nočitev v prihodnjem letu na osnovi funkcije trenda (rezultat naloge a) in izračunane sezonske indekse ter predvidite število nočitev po sezonah v prihodnjem letu.

Naloga 65

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

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

a) Ocenite stroške za prvo prihodnje leto.
b) Sezonski indeks za zadnje četrtletje znaša SI4 = 225,5 %. Kaj ta indeks vsebinsko pomeni?
c) Ocenite stroške popravil v zadnjem četrtletju prvega prihodnjega leta.

Naloga 66

V preglednici je podano število vpisa študentov na dodiplomskem študiju fakultete X po letih:

Leto	2016	2017	2018	2019	2000	2021	2022
Število vpisa študentov	340	310	290	270	350	370	390

a) Analizirajte, kako se je spreminjalo število vpisa študentov na dodiplomskem študiju fakultete X glede na leto 2022.

b) Analizirajte, kako se je spreminjalo število vpisa študentov na dodiplomskem študiju fakultete X med leti.

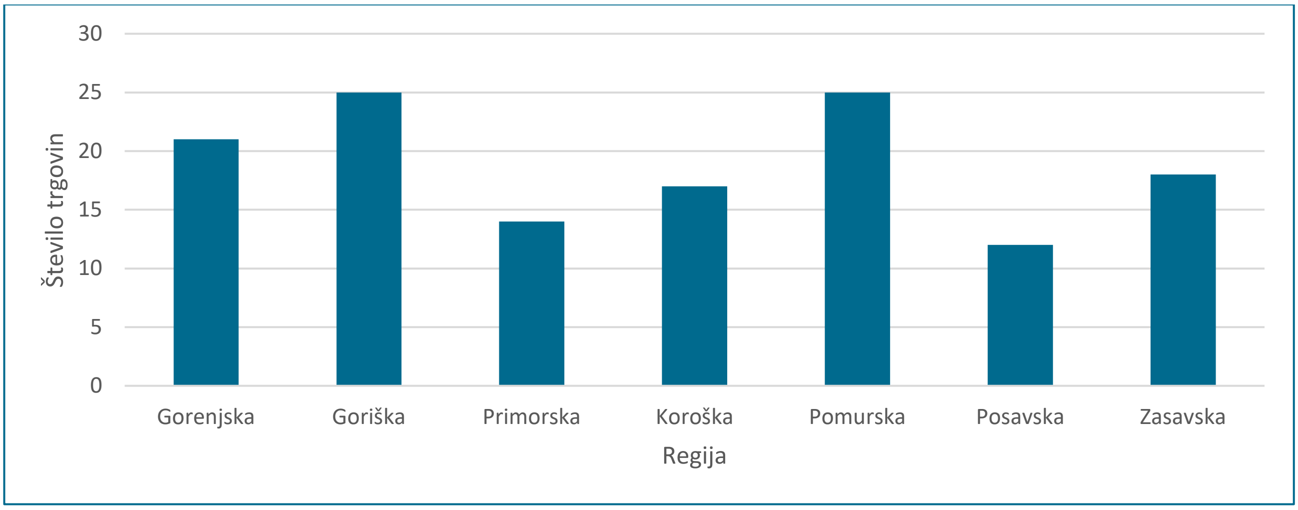
c) Izračunajte in pojasnite linearno funkcijo trenda ter ocenite število vpisa študentov na dodiplomskem študiju fakultete X v letu 2023.

d) Izračunajte in pojasnite sezonska indeksa SI_3 in SI_7.

7 REŠITVE NALOG

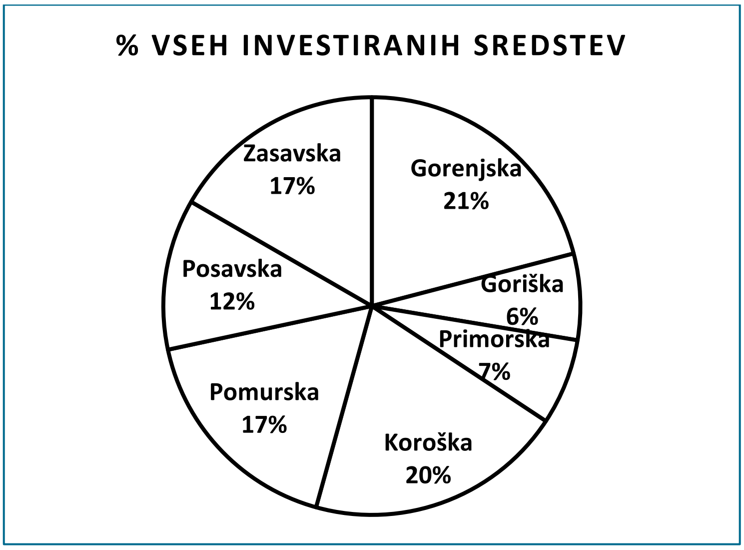
7.1 Urejanje in prikazovanje podatkov

Naloga 1 b)



Naloga 1 c)

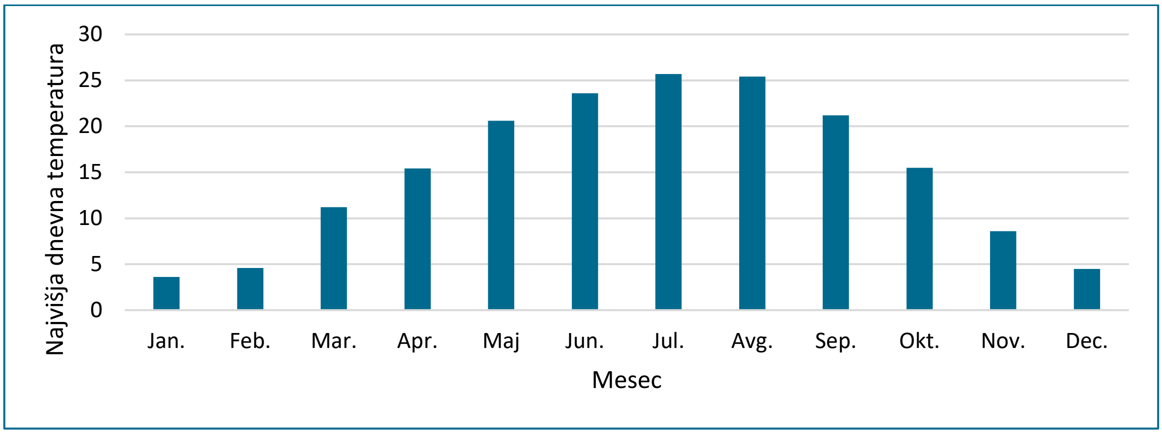
V tem primeru smo narisali strukturni krog ter strukturni pravokotnik, oboje s programom Excel.



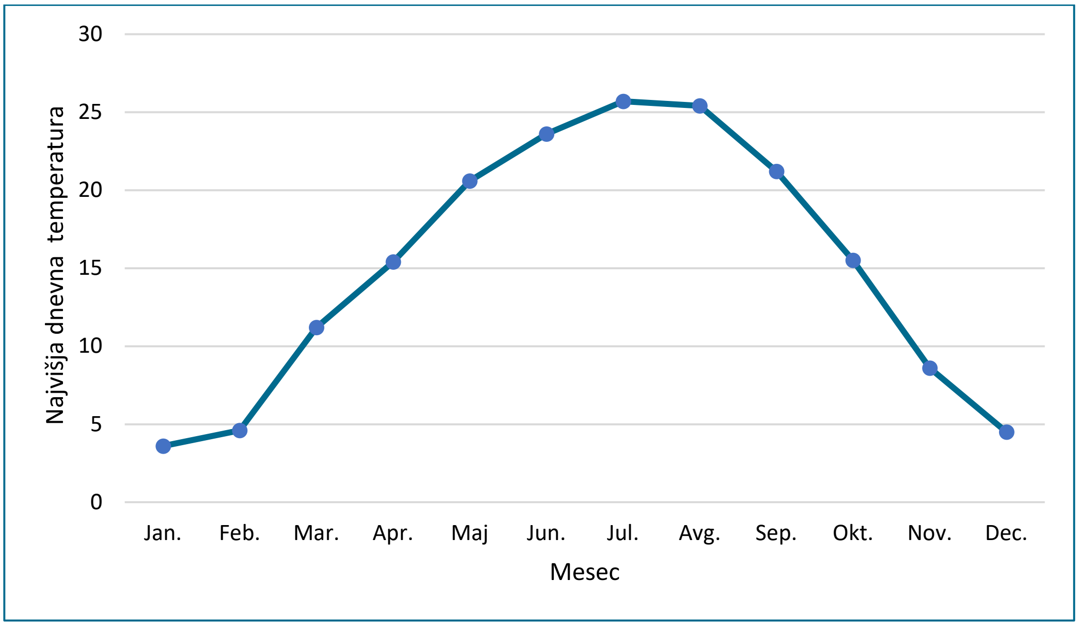
ali

Naloga 2 b)

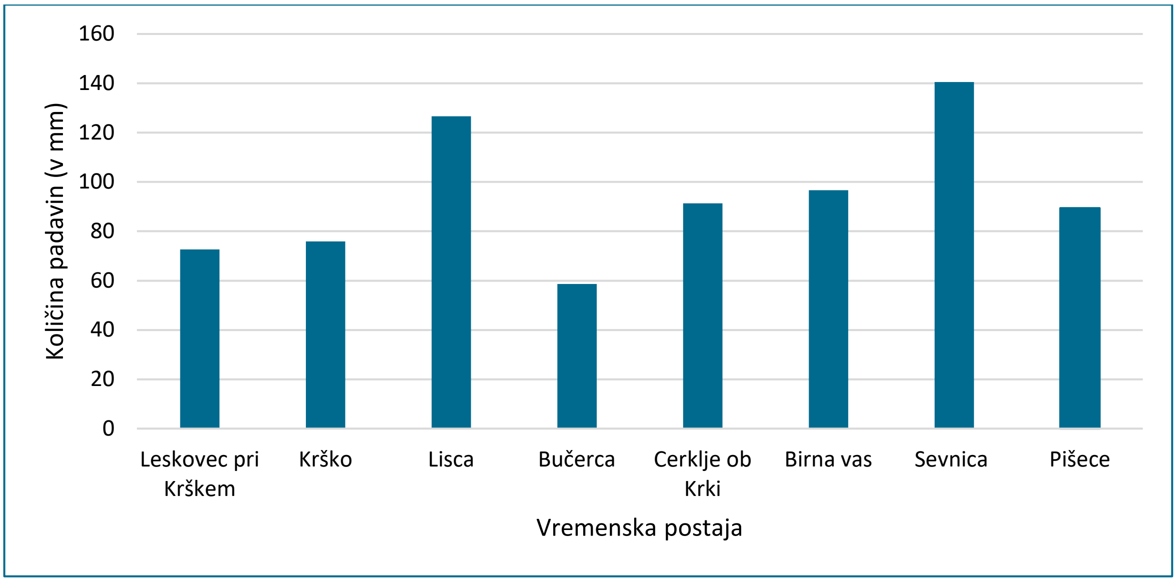
Prikaz s stolpci:



Linijski grafikon:



Naloga 3 b)



Naloga 4 b)

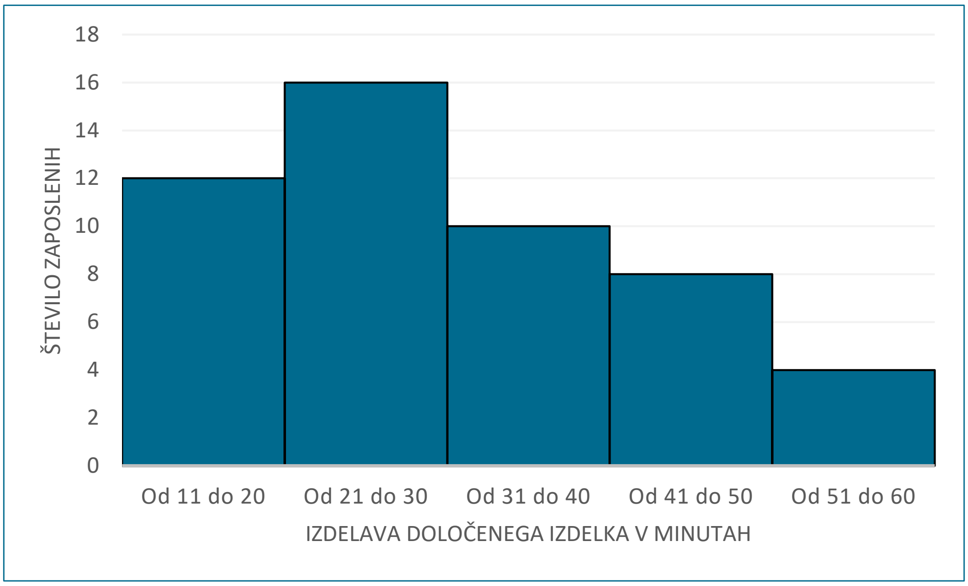
Število zaposlenih	Število podjetij	y_k, min	y_k, min	i_k	$f_{\%k}$
od 1 do 10	16	0,5	10,5	9	8,4
od 11 do 20	27	10,5	20,5	9	14,2
od 21 do 30	42	20,5	30,5	9	22,1
od 31 do 40	55	30,5	40,5	9	28,9
od 41 do 50	50	40,5	50,5	9	26,3
Skupaj	190				100 %

Naloga 5 b)

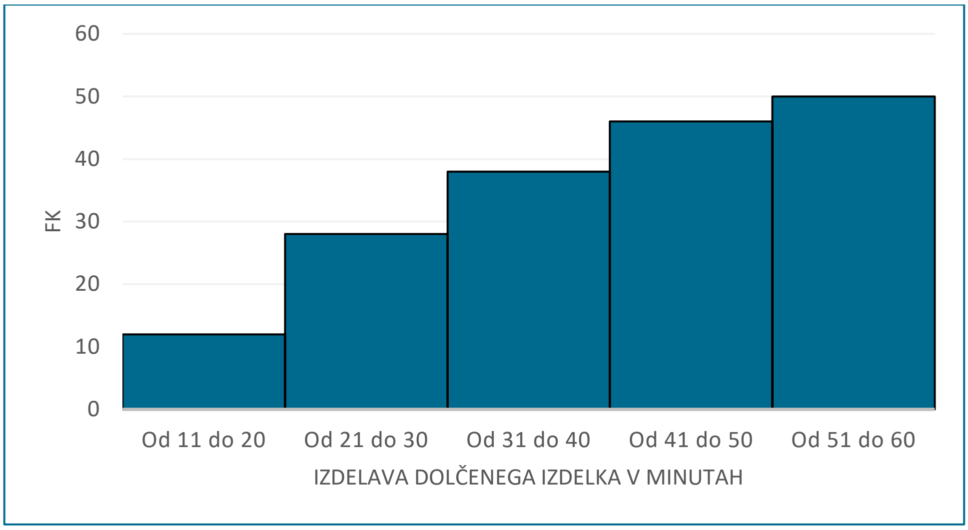
Izdelava določenega izdelka v minutah	f_k	F_k
od 11 do 20	12	12
od 21 do 30	16	28
od 31 do 40	10	38
od 41 do 50	8	46
od 51 do 60	4	50
Skupaj	50	

Naloga 5 c)

Histogram frekvenčne porazdelitve:



Histogram kumulativne frekvenčne porazdelitve:

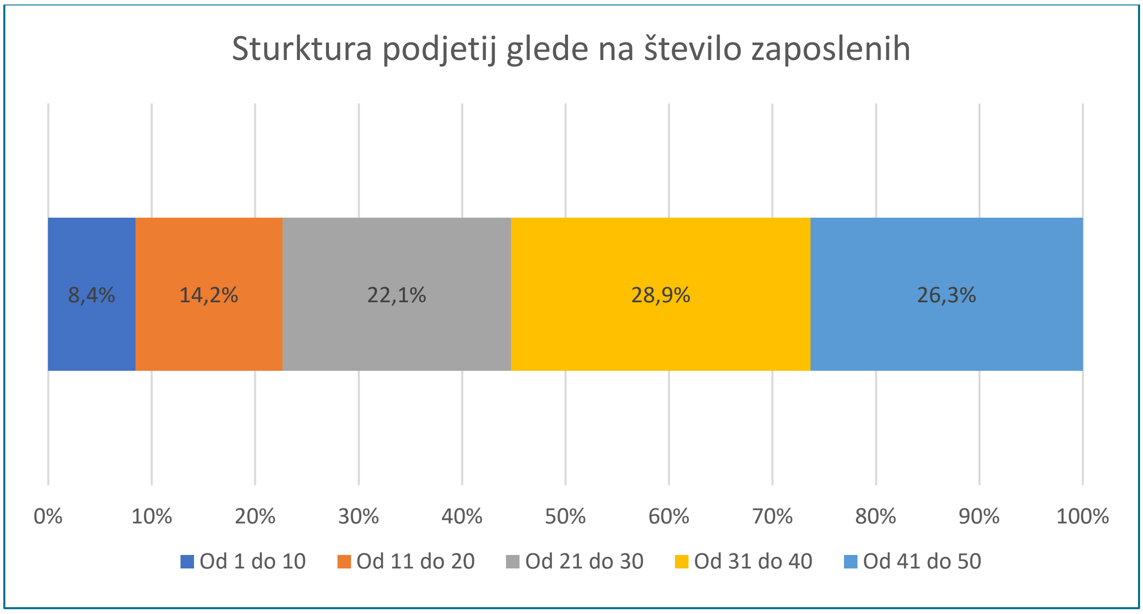


Naloga 5 d)

Struktura podjetij glede na število zaposlenih v strukturnem stolpcu:

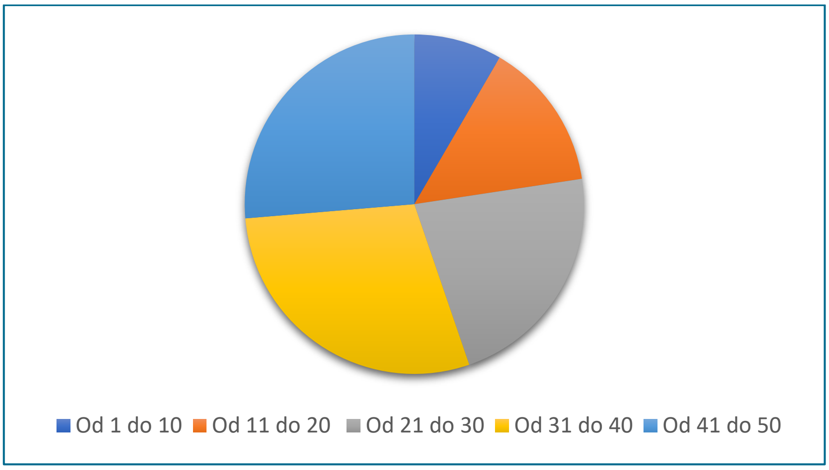
Število zaposlenih	Število podjetij	$f_{\%k}$
od 1 do 10	16	8,4
od 11 do 20	27	14,2
od 21 do 30	42	22,1
od 31 do 40	55	28,9
od 41 do 50	50	26,3
Skupaj	190	100 %

Struktura podjetij glede na število zaposlenih v strukturnem stolpcu:



Struktura podjetij glede na število zaposlenih v strukturnem krogu (v obrazcih št. 1.4):

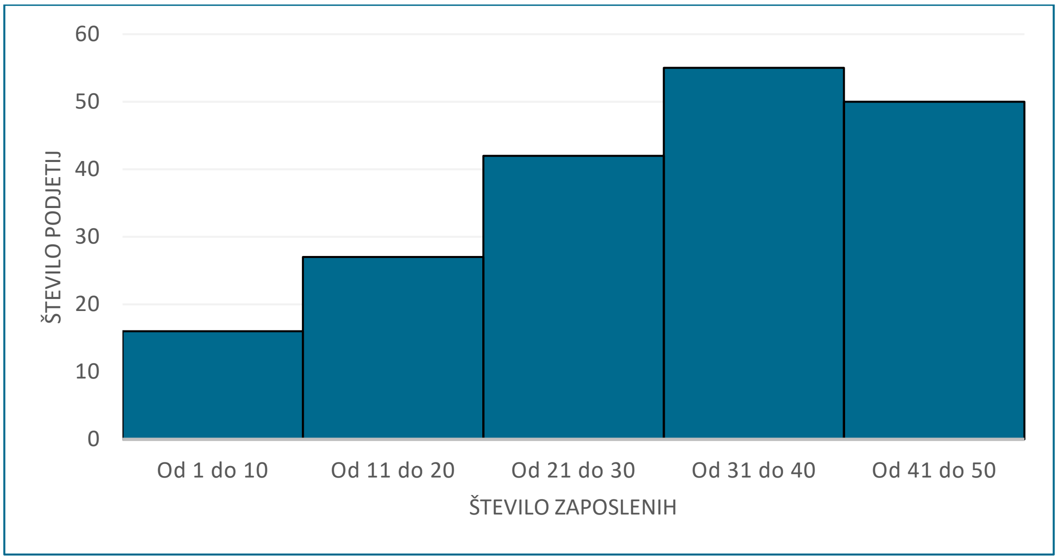
Število zaposlenih	Število podjetij	$f_{\%k}$	$f_{\%k} \cdot 3{,}6$
od 1 do 10	16	8,4	30,2
od 11 do 20	27	14,2	51,1
od 21 do 30	42	22,1	79,6
od 31 do 40	55	28,9	104,0
od 41 do 50	50	26,3	94,7
Skupaj	190	100 %	360 %



Naloga 5 d)

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

Histogram frekvenčne porazdelitve:



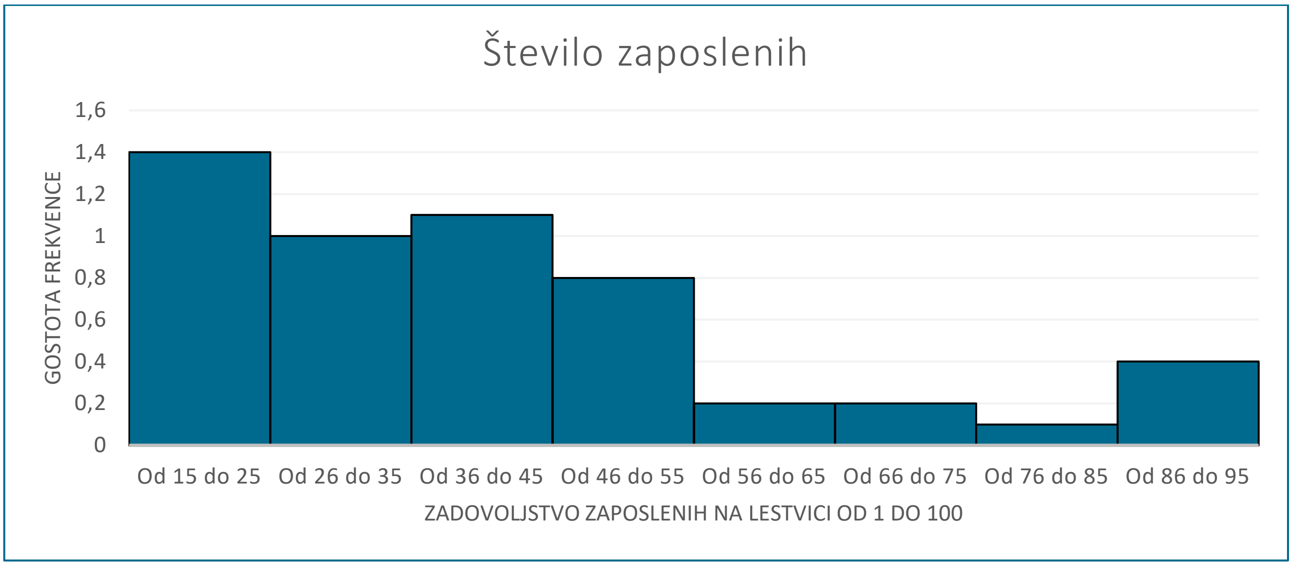
Naloga 6 b)

Zadovoljstvo	f_k	F_k	g_k
od 15 do 25	14	14	1,4
od 26 do 35	9	23	1
od 36 do 45	10	33	1,1
od 46 do 55	8	41	0,8
od 56 do 65	2	43	0,2
od 66 do 75	2	45	0,2
od 76 do 85	1	46	0,1
od 86 do 95	4	50	0,4

Naloga 6 c)

Frekvenčna porazdelitev z različno širokimi razredi: v frekvenčnem histogramu na y osi prikažemo gostoto frekvence (g_k).

Histogram frekvenčne porazdelitve:



Naloga 7 a)

Število ur študija	f_k	F_k
od 40 do pod 50	25	25
od 50 do pod 60	50	75
od 60 do pod 70	100	175
od 70 do pod 80	75	250
od 80 do pod 90	40	290
od 90 do pod 100	10	300
Skupaj	300	

Naloga 7 b), c)

Enačba za strukturni odstotek: $f\%k = \frac{fk}{N} \cdot 100$ (v obrazcih št. 1.3)

Število ur študija	f_k	F_k	$f\%_k$	$F\%_k$
od 40 do pod 50	25	25	8,33	8,33
od 50 do pod 60	50	75	16,67	25
od 60 do pod 70	100	175	33,33	58,33
od 70 do pod 80	75	250	25	83,33
od 80 do pod 90	40	290	13,33	96,66
od 90 do pod 100	10	300	3,33	100
Skupaj	300		100 %	

Naloga 8

Porabljen znesek za nakup	f_k	$f\%_k$
Od 1 do pod 50	80	16
Od 50 do pod 100	125	25
Od 100 do pod 150	148	29,6
Od 150 do pod 200	122	22,4
Od 200 do pod 250	35	7
Skupaj	500	100 %

Naloga 9 b), c)

Število trgovin	f_k	F_k	$f\%_k$
Od 0 do 5	7	7	4,9
Od 6 do 11	14	21	9,8
Od 12 do 17	19	40	13,3
Od 18 do 23	26	66	18,2
Od 24 do 29	35	101	24,5
Od 30 do 35	42	143	29,4
Skupaj	143		100 %

Naloga 10 c), d)

Število zaposlenih	Število podjetij	F_k	$f\%_k$	$F\%k$
Od 5 do 15	38	38	27,7	27,7
Od 16 do 26	35	73	25,5	53,2
Od 27 do 37	27	100	19,7	72,9
Od 38 do 48	20	120	14,6	87,5
Od 49 do 59	17	137	12,4	100
Skupaj	137		100 %	

Naloga 11 b)

Število ur bolniških izostankov	f_k	F_k	$f\%_k$	$F\%_k$
Od 10 do pod 19	13	13	2,89	2,89
Od 19 do pod 28	25	38	5,56	8,45
Od 28 do pod 37	116	154	25,78	34,23
Od 37 do pod 46	34	188	7,56	41,79
Od 46 do pod 55	56	244	12,44	54,23
Od 55 do pod 64	38	282	8,44	62,67
Od 64 do pod 73	168	450	37,33	100
Skupaj	450		100 %	

Naloga 11 c)

7,56 % zaposlenih je imelo od 37 do 46 ur bolniških izostankov.

Naloga 11 d)

62,67 % zaposlenih je imelo do 64 ur bolniških izostankov.

7.2 Deskriptivna statistika

Naloga 13 a)

Podatke uredimo v ranžirno vrsto:

R_i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
y_i	4	5	7	8	9	10	12	14	16	17	18	19	20	22	25	26	27	32	33	34

y_i je znan (y_i = 13), zato pri izračunu uporabimo enačbe (*kvantilni rangi iz nerazvrščenih vrednosti – v obrazcih št. 1.21, 1.22 in 1.23*):

$y_0 \leq y_i < y_1$

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

$P_i = \frac{R_i - 0{,}5}{N}$

V ranžirni vrsti poiščemo položaj dane vrednosti:

$y_i = 13$

$y_0 \leq y_i < y_1$

$y_0 = 12 \leq y_i = 13 < y_1 = 14$

Iz vrednosti členov poiščemo ustrezajoče range:

$R_0 \leq R_i < R_1$

$R_0 = 7 \leq R_i < R_1 = 8$

Izračunamo rang R_i:

$R_i = 7 + \frac{13-12}{14-12} \cdot (8 - 7) = 7{,}5$

Izračunamo relativni rang P_i:

$P_i = \frac{7{,}5-0{,}5}{20} = 0{,}35$

Odg.: 35 % poslovalnic je imelo 13 pritožb ali manj, 65 % poslovalnic pa več kot 13 pritožb.

Naloga 13 b)

V ranžirni vrsti poiščemo položaj dane vrednosti:

$y_i = 23$

$y_0 \leq y_i < y_1$

$y_0 = 22 \leq y_i = 23 < y_1 = 25$

Iz vrednosti členov poiščemo ustrezajoče range:

$R_0 \leq R_i < R_1$

$R_0 = 14 \leq R_i < R_1 = 15$

Izračunamo rang R_i:

$R_i = 14 + \frac{23-22}{25-22} \cdot (15 - 14) = 14{,}33$

Izračunamo relativni rang P_i:

$P_i = \frac{14{,}33-0{,}5}{20} = 0{,}69$

Odg.: 69 % poslovalnic je imelo 23 pritožb ali manj, 31 % poslovalnic pa več kot 23 pritožb.

Naloga 13 c)

Uporabimo enačbo za kvartilni razmik (v obrazcih št. 1.32): $Q = Q_3 - Q_1$ (kjer je Q_3 = 75% in Q_1 = 25 %), vendar je potrebno najprej izračunati:

Relativni rang P_i je znan: $P_i = 0{,}25$ (kjer je Q_1 = 25 %), zato pri izračunu uporabimo enačbe (*kvantili iz nerazvrščenih vrednosti – v obrazcih št. 1.18, 1.19 in 1.20*):

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

$$R_0 \leq R_i < R_1$$

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

Izračunamo rang R_i:

$R_i = 20 \cdot 0{,}25 + 0{,}5 = 5{,}5$

Nato določimo vrednosti:

$R_0 \leq R_i < R_1$

$R_0 = 5 \leq R_i = 5{,}5 < R_1 = 6$

$y_0 \leq y_i < y_1$

$y_0 = 9 \leq yi < y_1 = 10$

Nato izračunamo vrednost y_i:

$y_i = 9 + \frac{5{,}5-5}{6-5} \cdot (10 - 9) = 9{,}5$

Odg.: 25 % poslovalnic je imelo 9,5 pritožb ali manj, 75 % poslovalnic pa več kot 9,5 pritožb.

Pi = 0,75 (kjer je Q_3 = 75 %),

$R_i = 20 \cdot 0{,}75 + 0{,}5 = 15{,}5$

$R_0 \leq R_i < R_1$

$R_0 = 15 \leq R_i = 15{,}5 < R_1 = 16$

$y_0 \leq y_i < y_1$

$y_0 = 25 \leq yi < y_1 = 26$

Nato izračunamo vrednost y_i:

$y_i = 25 + \frac{15{,}5-15}{16-15} \cdot (26 - 25) = 25{,}5$ pritožb

Odg. 75 % poslovalnic je imelo 25,5 pritožb ali manj, 25 % poslovalnic pa več kot 25,5 pritožb.

Kvartilni razmik:

$Q = Q_3 - Q_1 = 25{,}5 - 9{,}5 = 16$ pritožb

Odg.: 50 % poslovalnic, ki glede na število pritožb ležijo na sredini ranžirne vrste, se razlikuje za največ 16 pritožb.

Naloga 13 d)

Za izračun povprečnega števila pritožb uporabimo enačbo (*aritmetična sredina iz nerazvrščenih vrednosti – v obrazcih št. 1.24*):

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

$\bar{y} = \frac{1}{20} \cdot (4 + 5 + 7 + 8 + 9 + 10 + 12 + \ldots + 34) = 17{,}9$ pritožb

Odg.: Povprečno število pritožb za 20 poslovalnic znaša 17,9 pritožb.

Naloga 13 e)

Izračun mediane (Me = 50 %):

$P_i = 0{,}5$

$R_i = 20 \cdot 0{,}5 + 0{,}5 = 10{,}5$

$R_0 \leq R_i < R_1$

$R_0 = 10 \leq R_i = 10{,}5 < R_1 = 11$

$y_0 \leq y_i < y_1$

$y_0 = 17 \leq y_i < y_1 = 18$

Nato izračunamo vrednost y_i:

$y_i = 17 + \frac{10{,}5-10}{11-10} \cdot (18 - 17) = 17{,}5$ pritožb

Odg. 50 % poslovalnic je imelo 17,5 pritožb ali manj, 50 % poslovalnic pa več kot 17,5 pritožb.

Naloga 13 f)

Uporabimo enačbo za decilni razmik (v obrazcih št. 1.33): $D = D_9 - D_1$ (kjer je $D_9 = 90$ % in $D_1 = 10$ %), vendar je potrebno najprej izračunati:

Relativni rang P_i je znan: $P_i = 0{,}9$ (kjer je $D_9 = 90$ %), zato pri izračunu uporabimo enačbe (*kvantili iz nerazvrščenih vrednosti – v obrazcih št. 1.18, 1.19, 1.20*):

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

$R_0 \leq R_i < R_1$

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

Izračunamo rang R_i:

$R_i = 20 \cdot 0{,}9 + 0{,}5 = 18{,}5$

Nato določimo vrednosti:

$R_0 \leq R_i < R_1$

$R_0 = 18 \leq R_i = 18{,}5 < R_1 = 19$

$y_0 \leq y_i < y_1$

$y_0 = 32 \leq yi < y_1 = 33$

Nato izračunamo vrednost y_i:

$y_i = 32 + \frac{18{,}5-18}{19-18} \cdot (33 - 32) = 32{,}5$ pritožb

Odg.: 90 % poslovalnic je imelo 32,5 pritožb ali manj, 10 % poslovalnic pa več kot 32,5 pritožb.

$Pi = 0{,}1$ (kjer je $D_1 = 10$ %)

$R_i = 20 \cdot 0{,}1 + 0{,}5 = 2{,}5$

$R_0 \leq R_i < R_1$

$R_0 = 2 \leq R_i = 2{,}5 < R_1 = 3$

$y_0 \leq y_i < y_1$

$y_0 = 5 \leq yi < y_1 = 7$

Nato izračunamo vrednost y_i:

$y_i = 5 + \frac{2{,}5-2}{3-2} \cdot (7 - 5) = 6$ pritožb

Odg. 10 % poslovalnic je imelo 6 pritožb ali manj, 90 % poslovalnic pa več kot 9 pritožb.

Decilni razmik:

$D = D_9 - D_1 = 32{,}5 - 6 = 26{,}5$ pritožb

Odg.: 80% poslovalnic, ki glede na število pritožb ležijo na sredini ranžirne vrste, se razlikuje za največ 26,5 pritožb.

Naloga 13 g)

Izračunamo koeficient variabilnosti v odstotku za oba podjetja po enačbi (v obrazcih št. 1.39):

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

Podjetje Y:

$\bar{y} = 8$

$\sigma = 2$

$KV\% = \frac{2}{8} \cdot 100 = 25\ \%$

Odg.: Delež standardnega odklona v aritmetični sredini znaša 25 %.

Podjetje X:

$\bar{y} = 17,9$

Izračunamo varianco iz nerazvrščenih vrednosti po enačbi (v obrazcih št. 1.34):

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

$$\sigma^2 = \frac{1}{20} \cdot [(4 - 17,9)^2 + (5 - 17,9)^2 + (7 - 17,9)^2 + \ldots + (34 - 17,9)^2] = 83,99 \text{ pritožb}^2$$

Izračunamo standardni odklon po enačbi (v obrazcih št. 1.38):

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

$$\sigma = \sqrt{83,99} = 9,16 \text{ pritožb}$$

$$KV\% = \frac{9,16}{17,9} \cdot 100 = 51,17\ \%$$

Odg.: Delež standardnega odklona v aritmetični sredini znaša 51,17 %. V podjetju X se poslovalnice glede na število pritožb kupcev med seboj bolj razlikujejo.

Naloga 14 a)

Podatke uredimo v ranžirno vrsto:

R_i	1	2	3	4	5	6	7	8	9	10	11	12
y_i	22	26	29	30	33	34	38	39	40	41	43	45

Uporabimo enačbo za kvartilni razmik (v obrazcih št. 1.32): $Q = Q_3 - Q_1$ (kjer je $Q_3 = 75$ % in $Q_1 = 25$).

Relativni rang P_i je znan: $P_i = 0,25$ (kjer je $Q_1 = 25$ %), zato pri izračunu uporabimo enačbe (*kvantili iz nerazvrščenih vrednosti – v obrazcih št. 1.18, 1.19 in 1.20*):

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

$R_0 \leq R_i < R_1$

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

Izračunamo rang R_i:
$R_i = 12 \cdot 0{,}25 + 0{,}5 = 3{,}5$

Nato določimo vrednosti:
$R_0 \leq R_i < R_1$
$R_0 = 3 \leq R_i = 3{,}5 < R_1 = 4$
$y_0 \leq y_i < y_1$
$y_0 = 29 \leq yi < y_1 = 30$

Nato izračunamo vrednost y_i:
$y_i = 29 + \frac{3{,}5-3}{4-3} \cdot (30 - 29) = 29{,}5$ minut

Odg.: 25 % zaposlenih je porabilo za izdelavo enega izdelka 29,5 minut ali manj, 75 % zaposlenih pa več kot 29,5 minut.

Pi = 0,75 (kjer je Q_3 = 75 %)
$R_i = 12 \cdot 0{,}75 + 0{,}5 = 9{,}5$
$R_0 \leq R_i < R_1$
$R_0 = 9 \leq R_i = 9{,}5 < R_1 = 10$
$y_0 \leq y_i < y_1$
$y_0 = 40 \leq yi < y_1 = 41$

Nato izračunamo vrednost y_i:
$y_i = 40 + \frac{9{,}5-9}{10-9} \cdot (41 - 40) = 40{,}5$ minut

Odg. 75 % zaposlenih je porabilo za izdelavo enega izdelka 40,5 minut ali manj, 25 % zaposlenih pa več kot 40,5 minut.

Kvartilni razmik:
$Q = Q_3 - Q_1 = 40{,}5 - 29{,}5 = 11$ minut

Odg.: 50 % zaposlenih, ki glede na porabljen čas za izdelavo enega izdelka ležijo na sredini ranžirne vrste, se razlikuje za največ 11 minut.

Izračun mediane (Me = 50 %):
Pi = 0,5
$R_i = 12 \cdot 0,5 + 0,5 = 6,5$
$R_0 \leq R_i < R_1$
$R_0 = 6 \leq R_i = 6,5 < R_1 = 7$
$y_0 \leq y_i < y_1$
$y_0 = 34 \leq yi < y_1 = 38$

Nato izračunamo vrednost y_i:
$y_i = 34 + \frac{6,5-6}{7-6} \cdot (38 - 34) = 36$ minut

Odg. 50 % zaposlenih je imelo porabo časa za izdelavo enega izdelka 36 minut ali manj, 50 % zaposlenih pa več kot 36 minut.

Naloga 14 b)

Povprečno porabljeni čas za en izdelek v organizaciji A:
Uporabimo enačbo – *aritmetična sredina iz nerazvrščenih vrednosti* (v obrazcih št. 1.24):

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

$\bar{y} = \frac{1}{12} \cdot (26 + 38 + 45 + \ldots + 30) = 35$ minut

Odg.: V povprečju so zaposleni v organizaciji A za izdelavo enega izdelka porabili 35 minut.

Povprečno porabljeni čas za en izdelek v organizaciji B:
Uporabimo enačbo – *aritmetična sredina iz razvrščenih vrednosti* (v obrazcih št. 1.25):

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

$\bar{y} = \frac{1}{730} \cdot 24124 = 33,05$ minut

Odg.: V povprečju so zaposleni v organizaciji B za izdelavo enega izdelka porabili 33,05 minut.

Naloga 14 c)

Uporabimo enačbo za variacijski razmik (v obrazcih št. 1.31): $VR = y_{max} - y_{min}$

VR = 45 – 22 = 23 minut

Odg.: Zaposleni se glede na porabljen čas za izdelavo izdelka med seboj razlikujejo za največ 23 minut.

Naloga 14 d)

Varianca in standardni odklon

Organizacija A: $\sigma^2 = \frac{15266}{9} - 35^2 = 47{,}17 \text{ minut}^2$

$\sigma = 6{,}87 \text{ minut}$

Organizacija B: $\sigma^2 = \frac{817296}{730} - 33{,}05^2 = 27{,}28 \text{ minut}^2$

$\sigma = 5{,}22 \text{ minut}$

Pri organizaciji B smo uporabili enačbo (*varianca iz razvrščenih vrednosti – v obrazcih št. 1.37*):

$$\sigma^2 = \frac{1}{N}\sum_{k=1}^{r} f_k y_k^2 - \bar{y}^2$$

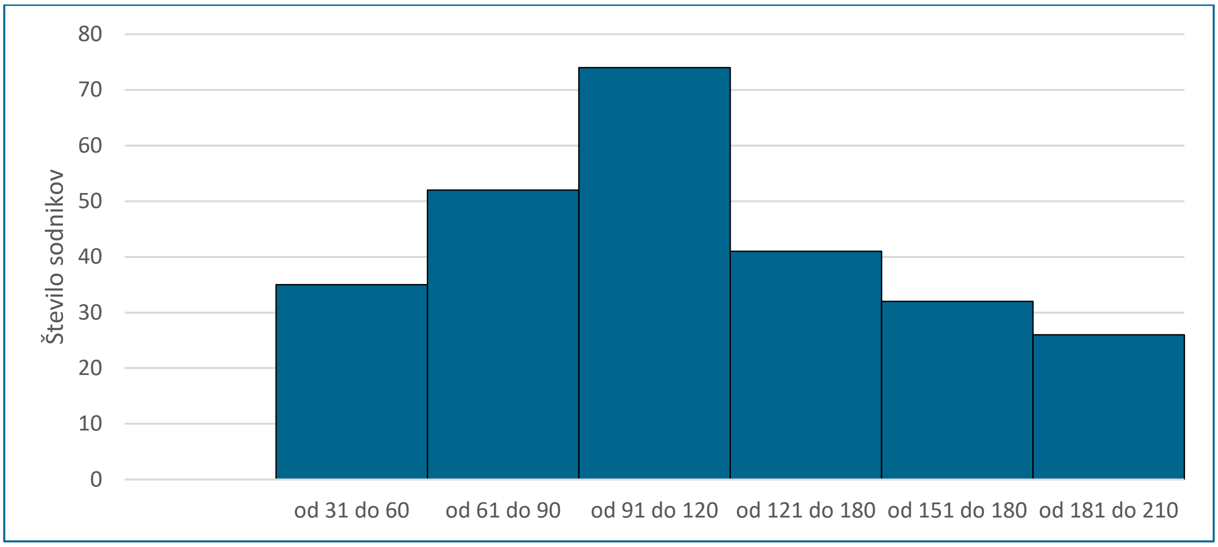
Naloga 14 e)

Organizacija A: KV % $= \frac{6{,}87}{35} \cdot 100 = 19{,}6\,\%$

Organizacija B: KV % $= \frac{5{,}22}{33{,}05} \cdot 100 = 15{,}8\,\%$

Naloga 15 b)

Porazdelitev je asimetrična v desno stran.



Naloga 15 c)

$$\bar{y} = \frac{29260}{260} = 112{,}54 \text{ zadev}$$
$$\sigma^2 = \frac{3815495}{260} - 112{,}54^2 = 2.010{,}08 \text{ zadev}^2$$
$$\sigma = 44{,}83 \text{ zadev}$$
KV % = 39,83 %

Delež standardnega odklona v aritmetični sredini znaša 39,83 %.

Naloga 16 a)

VR= 49 vozil

Naloga 16 c)

$$\bar{y} = 23{,}06 \text{ dni}$$

Naloga 16 d)

$$\sigma = 11{,}57 \text{ vozil}$$
$$\text{KV } \% = 50{,}2 \ \%$$

Naloga 17 b)

Podatke uredimo v ranžirno vrsto.

Q_1 = 50 d.e.

Q_3 = 850 d.e.

Naloga 18 a)

Banka "Y":

KV % = 35,19 %

Banka "X":

Vložena sredstva v 10^2 evrov	Število vlagateljev	Sredina razreda (y_k)
od 1 do manj kot 10	223	5,5
od 10 do manj kot 20	356	15
od 20 do manj kot 50	439	35
od 50 do manj kot 100	245	75
od 100 do manj kot 500	188	300
Skupaj	1451	

Uporabimo enačbo za *aritmetično sredino iz razvrščenih vrednosti* (v obrazcih št. 1.25):

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

$$\bar{y} = \frac{96706{,}5}{1451} = 66{,}648 \cdot 10^2 \text{ eur}$$

Uporabimo enačbo za *varianco iz razvrščenih vrednosti* (v obrazcih št. 1.37):

$$\sigma^2 = \frac{1}{N}\sum_{k=1}^{r} f_k y_k^2 - \bar{y}^2$$

$\sigma^2 = 8599{,}2 \cdot 10^2 \text{ eur}^2$

$\sigma = 92{,}73 \cdot 10^2 \text{ eur}$

KV % = 139,13 %

Odg.: V banki "X" se vlagatelji glede na vložena sredstva med seboj bolj razlikujejo.

Naloga 19 a)

$\bar{y} = \frac{53}{8} = 6{,}6$ nadur

Odg.: Povprečno tedensko število nadur zaposlenih znaša 6,6 nadur.

Naloga 19 b)

$\sigma^2 = 6{,}24 \text{ nadur}^2$

$\sigma = 2{,}5$ nadur

KV % = 37,88 %

Odg.: Delež standardnega odklona v aritmetični sredini znaša 37,88 %.

Naloga 19 c)

Podatke uredimo v ranžirno vrsto:

R_i	1	2	3	4	5	6	7	8
y_i	3	4	5	6	7	8	9	11

Izračun mediane (Me = 50 %):

$P_i = 0{,}5$

$R_i = 8 \cdot 0{,}5 + 0{,}5 = 4{,}5$

$R_0 \leq R_i < R_1$

$R_0 = 4 \leq R_i = 4{,}5 < R_1 = 5$

$y_0 \leq y_i < y_1$

$y_0 = 6 \leq y_i < y_1 = 7$

Nato izračunamo vrednost y_i:

$y_i = 6 + \frac{4{,}5-4}{5-4} \cdot (7-6) = 6{,}5$ nadur

Odg. 50 % zaposlenih ima tedensko število nadur 6,5 nadur ali manj.

Naloga 19 d)

Kvartilni razmik: $Q = Q_3 - Q_1$

$Q = 8{,}5 - 4{,}5 = 4$ nadure

Odg.: 50 % zaposlenih, ki glede na tedensko število nadur ležijo na sredini ranžirne vrste, se razlikuje za največ štiri nadure.

Naloga 20 a)

Modus: 8 in 24 bolniških dni

Naloga 20 b)

Decilni razmik: $D = D_9 - D_1$

$D_9 = 90$ %

$P_i = 0{,}9$

$R_i = 4{,}5$

$y_i = 10$ bolniških dni

$D_1 = 10$ %

$P_i = 0{,}1$

$R_i = 1{,}5$

$y_i = 4$ bolniški dnevi

$D = 10 - 4 = 6$ bolniških dni

Odg.: 80 % zaposlenih, ki glede na število bolniških dni ležijo na sredini ranžirne vrste, se razlikuje za največ šest bolniških dni.

Naloga 20 c)

$D_3 = 30\ \%$
$P_i = 0{,}3$
$R_i = 3{,}5$
$y_i = 8$ bolniških dni

Odg.: Zaposleni, ki spadajo v 30 % zaposlenih z najmanj bolniških dni v podjetju, so imeli osem bolniških dni ali manj.

Naloga 20 d)

$D_7 = 70\ \%$
$P_i = 0{,}7$
$R_i = 7{,}5$
$y_i = 22$ bolniških dni

Odg.: Zaposleni, ki spadajo v 30 % zaposlenih z največ bolniških dni v podjetju, so imeli več kot 22 bolniških dni.

Naloga 21

Ranžirna vrsta:

R_i	1	2	3	4	**5**	6	7	8	9
y_i	75	98	160	180	**195**	220	246	260	290

Enačba za koeficient asimetrije na podlagi mediane (v obrazcih št. 1.42):

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

Mediana:
$P_i = 0{,}5$
$R_i = N \cdot P_i + 0{,}5 = 9 \cdot 0{,}5 + 0{,}5 = 5$
$Me = 195$ d.e.

50 % študentov je porabilo za nakup določenega študijskega gradiva 195 d.e. ali manj, 50 % študentov pa več kot 195 d.e.

$\bar{y} = \frac{1}{9} \cdot (75 + 98 + \ldots + 290) = 191{,}56$ d.e.

$VAR = \sigma^2 = \frac{1}{N}\sum_{i=1}^{N} y_i^2 - \bar{y}^2$, (v obrazcih št. 1.35)

$\sigma^2 = \frac{1}{9} \cdot [(75^2 + 98^2 + \ldots + 290^2)] - 191{,}56^2 = \frac{1}{9} \cdot 371.870 - 191{,}56^2 = 4.623{,}66$ (d.e.)^2

$\sigma = \sqrt{4.623{,}66} = 67{,}997$ d.e.

Izračun koeficienta asimetrije na podlagi mediane (v obrazcih št. 1.42):

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

$$KA_{Me} = \frac{3(191{,}56 - 195)}{67{,}997} = -0{,}152$$

Porazdelitev je šibka in asimetrična v desno.

Ponovitev:

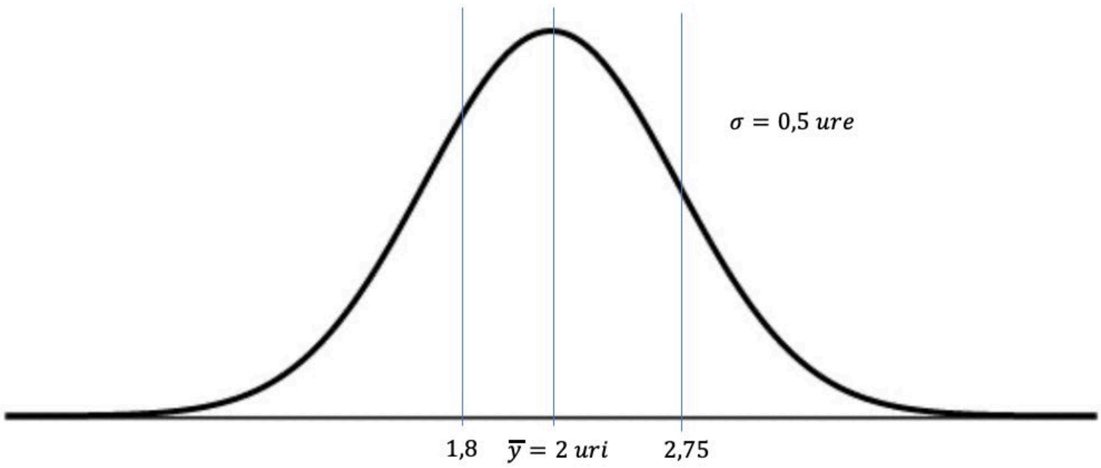
$\bar{y} > Me$ pomeni, da je porazdelitev asimetrična v desno (več kot polovica vrednosti je manjših od aritmetične sredine).

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

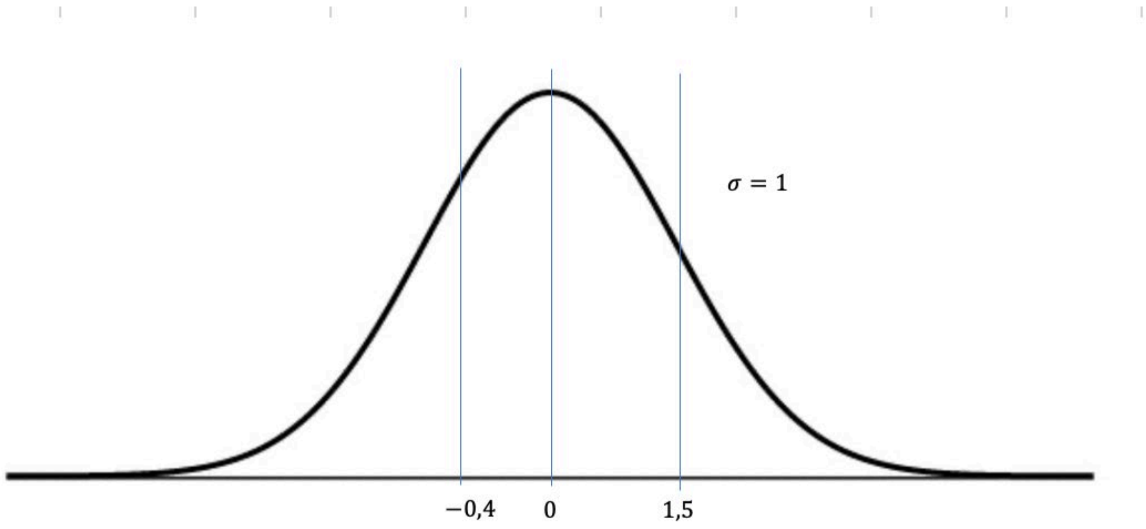
$\bar{y} = Me$ pomeni, da je porazdelitev simetrična.

7.3 Teoretične porazdelitve

Naloga 22 a)



Standardizirana normalna porazdelitev



$z_1 = \frac{y-\bar{y}}{\sigma} = \frac{1{,}8-2}{0{,}5} = -0{,}4$

$z_2 = \frac{2{,}75-2}{0{,}5} = 1{,}5$

P (1,8 ure < y < 2,75 ure) = H (–0,4) + H (1,5) = 0,1554 + 0,4332 = 0,5886 = 58,86 %

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

Verjetnost, da je v naključno izbranem gospodinjstvu družinski računalnik v uporabi za igranje igric med 1,8 in 2,75 urami na dan, je enaka 58,86 %.

Naloga 22 b)

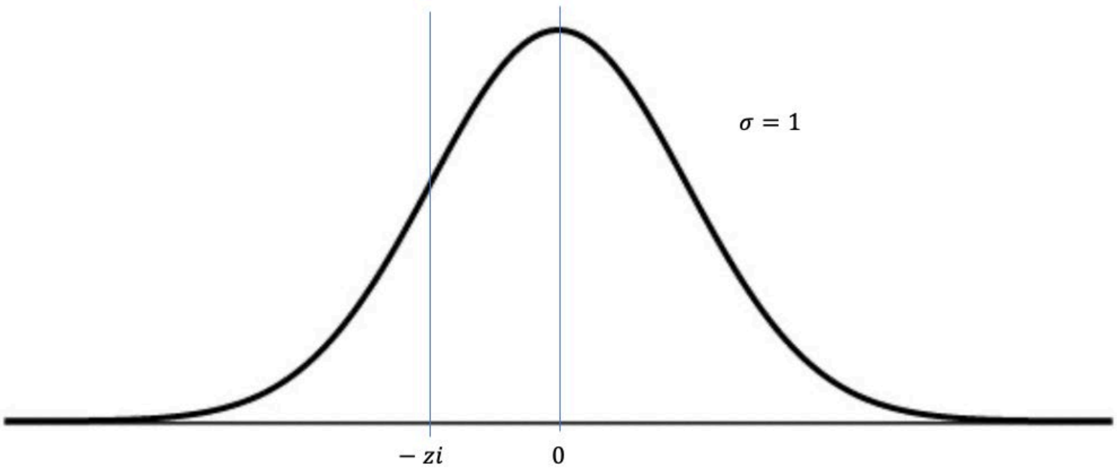
z_i = 6; H(6) = 0,5

P(y > 5) = 0,5 – H(6) = 0,5 – 0,5 = 0

Naloga 22 c)

V tem primeru je podana verjetnost:

P(y < z_i) = 0,25



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

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

$H(–0{,}67) = 0{,}2486 = 0{,}25$

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

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

V 25 % gospodinjstev, kjer je čas uporabe družinskega računalnika za igranje igric najkrajši, znaša ta čas do 1,7 ure dnevno.

Naloga 23

$z_i = \frac{y-\bar{y}}{\sigma} = \frac{210-240}{80} = –0{,}38$

$P(y > 210 \text{ minut}) = 0{,}5 + H(–0{,}38) = 0{,}5 + 0{,}1480 = 0{,}648 = 64{,}8\ \%$

Naloga 24

$z_i = \frac{y-\bar{y}}{\sigma} = \frac{14-17}{5} = -0{,}6$

$P(y > 14 \text{ enot}) = 0{,}5 + H\ (–0{,}6)\ = 0{,}5 + 0{,}2257 = 0{,}7257 = 72{,}57\ \%$

Naloga 25

1. Alternativna investicijska možnost:

$z_i = \frac{y-\bar{y}}{\sigma} = \frac{1500-1780}{115} = –2{,}43$

$P(y > 1500\ €) = H\ (–2{,}43) + 0{,}5 = 0{,}4925 + 0{,}5 = 0{,}9925 = 99{,}25\ \%$

2. Alternativna investicijska možnost:

$z_i = \frac{y-\bar{y}}{\sigma} = \frac{1500-2168}{425} = -1{,}57$

P(y > 1500 €) = H (–1,57) + 0,5 = 0,4418 + 0,5 = 0,9418 = 94,18 %

3. Alternativna investicijska možnost:

$z_i = \frac{y-\bar{y}}{\sigma} = \frac{1500-3000}{515} = -2{,}91$

P(y > 1500 €) = H (–2,91) + 0,5 = 0,4982 + 0,5 = 0,9982 = 99,82 %

Odgovor: Pri drugi alternativni investiciji je verjetnost manjša, da bo donos večji od 1500 €.

Naloga 26 a)

Dosežena norma <u>največ</u> 98 %:

$z_i = \frac{y-\bar{y}}{\sigma} = \frac{98-102}{4} = -1$

P(y ≤ 98 %) = 0,5 – H (–1) = 0,5 – 0,3413 = 0,1587 = 15,87 %

Odgovor: Verjetnost, da bo naključno izbran zaposleni dosegel normo največ 98 %, je enaka 0,1587 oziroma 15,87 %.

Naloga 26 b)

$z_1 = \frac{y-\bar{y}}{\sigma} = \frac{97-102}{4} = -1{,}25$

$z_2 = \frac{y-\bar{y}}{\sigma} = \frac{105-102}{4} = 0{,}75$

P(97 % ≤ y ≤ 105 %) = H (–1,25) + H (0,75) = 0,3944 + 0,2734 = 0,6678 = 66,78 %

Odgovor: Verjetnost, da bo naključno izbran zaposleni dosegel normo med 97 in 105 %, je 0,6678 oziroma 66,78 %.

Naloga 26 c)

$P(y_1 \leq y \leq y_2) = 80\ \%$

$H\ (z_i) = \frac{0{,}80}{2} = 0{,}4$

Vrednost H (z_i) odčitamo iz tabele *ploščine H(z) za standardizirano normalno porazdelitev*, kjer poiščemo najbližje število 0,4 (tj. 0,3997), kar pomeni, da je $z_i = \pm 1{,}28$.

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

$y_1 = 102 - 1{,}28 \cdot 4 = 96{,}88$

$y_2 = 102 + 1{,}28 \cdot 4 = 107{,}12$

Naloga 27

$z_1 = \frac{y-\bar{y}}{\sigma} = \frac{600-1500}{400} = -2{,}25$

$z_2 = \frac{y-\bar{y}}{\sigma} = \frac{1200-1500}{400} = -0{,}75$

$P(600€ \leq y \leq 1200€) = H(-2{,}25) - H(-0{,}75) = 0{,}4878 - 0{,}2734 = 0{,}2144 = 21{,}44\ \%$

Naloga 28 a)

$P(y > y_i) = 40\ \%$

$H(z_i) = 0{,}1$

Vrednost H (z_i) odčitamo iz tabele *ploščine H(z) za standardizirano normalno porazdelitev*, kjer poiščemo najbližje število 0,1 (tj. 0,0987), kar pomeni, da je $z_i = 0{,}25$.

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

$y_i = 90+0{,}25 \cdot 35=98{,}75$ minut

Naloga 28 b)

$P(y < y_i) = 30\ \%$

$H(z_i) = 0{,}2$

Vrednost H (z_i) odčitamo iz tabele *ploščine H(z) za standardizirano normalno porazdelitev* kjer poiščemo najbližje število 0,2 (tj. 0,1985), kar pomeni, da je $z_i = -0{,}52$.

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

$y_i = 90-0{,}52 \cdot 35=71{,}8$ minut

Naloga 28 c)

$z_1 = \frac{y-\bar{y}}{\sigma} = \frac{50-90}{35} = -1{,}14$

$z_2 = \frac{105-90}{35} = 0{,}43$

P (50 minut < y < 105 minut) = H (–1,14) + H (0,43) = 0,3729 + 0,1664 = 0,5393 = 53,93 %

Naloga 28 d)

$z_1 = \frac{y-\bar{y}}{\sigma} = \frac{100-90}{35} = 0{,}29$

$z_2 = \frac{130-90}{35} = 1{,}14$

P (100 minut < y < 130 minut) = H (z_2) – H (z_1) = H (1,14) – H (0,29) = 0,3729 – 0,1141 = 0,2588 = 25,88 %

Naloga 29

$P(D_3 < y < Q_3) = \gamma$.

$\bar{y} = 5{,}5$

$\sigma^2 = 2$

$\sigma = 1{,}41$

D_3 = 30 %, P_i = 0,30 – 0,5 = –0,2 (poiščemo najbližje število 0,2; tj. 0,1985)

$z_1 = -0{,}52$

$y_i = \bar{y} + z_i \cdot \sigma = 5{,}5 - 0{,}52 \cdot 1{,}41 = 4{,}77$

Q_3 = 75 %, P_i = 0,75 – 0,5 = 0,25 (poiščemo najbližje število 0,25; tj. 0,2486)

$z_2 = 0{,}67$

$y_i = \bar{y} + z_i \cdot \sigma = 5{,}5 + 0{,}67 \cdot 1{,}41 = 6{,}44$

$P(D_3 < y < Q_3) = 0{,}2 + 0{,}25 = 0{,}45 = 45\ \%$

$P(4{,}77 < y < 6{,}44) = 45\ \%$

7.4 Enostavna regresijska analiza

Naloga 30 a)

Oblika: Linearna oblika

Smer med spremenljivkama: Pozitivna smer, kar pomeni, da z naraščanjem investicij v tehnologijo *(x)* v povprečju narašča ustvarjen dobiček *(y)*.

Jakost povezanosti med spremenljivkama: Obstaja močna povezanost med odvisno (ustvarjen dobiček) in neodvisno spremenljivko (investicije v tehnologijo).

Naloga 30 b)

Izračun parametrov regresijske premice:

Preden se lotimo izračuna regresijskih koeficientov a in b, izračunamo:

$\Sigma x_i = 115 + 130 + 140 + 149 + 160 + 171 = 865$

$\Sigma x_i^2 = 115^2 + 130^2 + 140^2 + 149^2 + 160^2 + 171^2 = 126.767$

$\Sigma y_i = 328 + 330 + 390 + 361 + 421 + 400 = 2.230$

$\Sigma y_i^2 = 328^2 + 330^2 + 390^2 + 361^2 + 421^2 + 400^2 = 836.146$

$\Sigma x_i \cdot y_i = (115 \cdot 328) + (130 \cdot 330) + (140 \cdot 390) + \ldots + (171 \cdot 400) = 324.769$

$\bar{x} = \frac{865}{6} = 144{,}1667$

$\bar{y} = \frac{2230}{6} = 371{,}6667$

Oba regresijska koeficienta izračunamo po enačbi (v obrazcih št. 2.3 in 2.4):

$$b = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} = \frac{\frac{1}{N}\sum_{i=1}^{N}x_i y_i - \bar{x}\cdot\bar{y}}{\frac{1}{N}\sum_{i=1}^{N}x_i^2 - \bar{x}^2}$$

$$b = \frac{\frac{1}{6}\cdot(324769) - 144{,}1667 \bullet 371{,}6667}{\frac{1}{6}\cdot(126767) - 144{,}1667^2} = 1{,}5887$$

$a = \bar{y} - b\bar{x}$

$a = 371{,}6667 - 1{,}5887 \cdot 144{,}1667 = 142{,}6291$

Izračunana regresijska koeficienta vstavimo v enačbo regresijske premice:

$\hat{y} = a + b \cdot x_i$

$\hat{y} = 142{,}6291 + 1{,}5887 \cdot x_i$

Pomen regresijskega koeficienta a: Pri investicijah v tehnologijo x = 0 lahko v povprečju pričakujemo, da bo ustvarjen dobiček podjetja 142,6291 (v 10^3 EUR).

Pomen regresijskega koeficienta b: Če se investicije v tehnologijo *(x)* povečajo za eno enoto (v 10^6 EUR), se ustvarjen dobiček *(y)* v povprečju poveča za 1,5887 (v 10^3 EUR).

Naloga 30 c)

$x = 180$ (v 10^6 EUR)
$\hat{y} = 142{,}6291 + 1{,}5887 \cdot x_i$
$\hat{y}_{x=180} = 142{,}6291 + 1{,}5887 \cdot 180$
$\hat{y}_{x=180} = 428{,}649$ (v 10^3 EUR)

Višina dobička za podjetje, ki bi investiralo v tehnologijo $x = 180$ (v 10^6 evrov), znaša 428,649 (v 10^3 EUR).

Naloga 30 d)

Parameter, na osnovi katerega določimo smer in jakost linearne korelacijske odvisnosti, je korelacijski koeficient. Izračunamo ga po enačbi (v obrazcih št. 2.9):

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y}$$

Za izračun σ_x in σ_y uporabimo enačbo za skupno varianco (v obrazcih št. 2.7):

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

$\sigma_x^2 = \frac{1}{6} \cdot [(115 - 144{,}1667)^2 + (130 - 144{,}1667)^2 + \ldots + (171 - 144{,}1667)^2] = 343{,}806$
$\sigma_x = \sqrt{343{,}806} = 18{,}543$
$\sigma_y^2 = \frac{1}{6} \cdot [(328 - 371{,}6667)^2 + (330 - 371{,}6667)^2 + \ldots + (400 - 371{,}6667)^2] = 1221{,}556$
$\sigma_y = \sqrt{1221{,}556} = 34{,}950$
$r_{xy} = 1{,}5887 \cdot \frac{18{,}542}{34{,}950} = 0{,}843$

Na osnovi rezultata ($r_{xy} = 0{,}8428$) vidimo, da obstaja močna povezanost med odvisno (ustvarjen dobiček) in neodvisno spremenljivko (investicije v tehnologijo). Smer povezanosti je pozitivna.

Naloga 30 e)

Izračunati moramo delež pojasnjene variance v skupni varianci za odvisno spremenljivko – determinacijski koeficient:
$r_{xy}^2 = 0{,}8428^2 = 0{,}7103$ oz. 71,03 %

Delež pojasnjene variance v skupni varianci za odvisno spremenljivko znaša 71,03 %.

Naloga 30 f)

Standardno napako ocene odvisne spremenljivke izračunamo po enačbi (v obrazcih št. 2.10):

$$\sigma_{ey} = \sigma_y \sqrt{1 - r_{xy}^2}$$

$\sigma_{ey} = 34{,}950 \cdot \sqrt{1 - 0{,}7103} = 18{,}811$

Standardna napaka ocene odvisne spremenljivke je različna od 0, kar pomeni, da na ustvarjen dobiček (odvisna spremenljivka) poleg investicij v tehnologijo (neodvisna spremenljivka) vplivajo še druge spremenljivke in slučajni vplivi.

Naloga 30 g)

Pri domnevi o normalni porazdelitvi ocen na primer velja, da se v razmiku $\hat{y}_x \pm 2 \cdot \sigma_{ey}$ nahaja 95,4 % vseh vrednosti.

$P(\hat{y}_x - 2 \cdot \sigma_{ey} < y_x < \hat{y}_x + 2 \cdot \sigma_{ey}) = 0{,}954 = 95{,}4\ \%$

$P(428{,}649 - 2 \cdot 18{,}811 < y_{x=180} < 428{,}649 + 2 \cdot 18{,}811) = 95{,}4\ \%$

$P(391{,}027 < y_{x=180} < 466{,}271) = 95\ \%$

Pri investicijah v tehnologijo x = 180 (v 10^6 EUR) bo ustvarjen dobiček podjetja med 391,027 EUR in 466,271 (v 10^3 EUR), kar trdimo s 95,4-% verjetnostjo.

Naloga 31

$\Sigma x_i = 27{,}6$

$\Sigma x_i^2 = 165{,}52$

$\Sigma y_i = 977$

$\Sigma y_i^2 = 202.483$

$\Sigma x_i \cdot y_i = 5.776$

$\bar{x} = \frac{27{,}6}{5} = 5{,}52$

$\bar{y} = \frac{977}{5} = 195{,}4$

Izračun regresijskih koeficientov (v obrazcih št. 2.3 in 2.4):

$$b = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2} = \frac{\frac{1}{N}\sum_{i=1}^{N}x_i y_i - \overline{x}\cdot\overline{y}}{\frac{1}{N}\sum_{i=1}^{N}x_i^2 - \overline{x}^2}$$

$$\mathrm{b} = \frac{\frac{1}{5}\cdot(5776) - 5{,}52\cdot 195{,}4}{\frac{1}{5}\cdot(165{,}52) - 5{,}52^2} = 29{,}083$$

$$a = \overline{y} - b\overline{x}$$

a = 195,4 – 29,083 · 5,52 = 34,862

Pomen regresijskega koeficienta a: Pri obrestni meri x = 0 lahko v povprečju pričakujemo, da bo znesek kratkoročnih oblik varčevanja znašal 34,862 d.e.

Pomen regresijskega koeficienta b: Če se obrestna mera *(x)* poveča za eno enoto, se znesek kratkoročnih oblik varčevanja *(y)* v povprečju poveča za 29,083 d.e.

Izračunana regresijska koeficienta vstavimo v enačbo regresijske premice:

$\hat{y} = a + b \cdot x_i$

$\hat{y} = 34{,}862 + 29{,}083 \cdot 7$

$\hat{y} = 238{,}443$

Izračunamo korelacijski koeficient po enačbi (v obrazcih št. 2.9):

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y}$$

Za izračun σ_x in σ_y uporabimo enačbo za skupno varianco (v obrazcih št. 2.7):

$$\sigma_y^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y})^2$$

$\sigma_x^2 = 2{,}6336$

$\sigma_x = 1{,}623$

$\sigma_y^2 = 2315{,}44$

$\sigma_y = 48{,}119$

$r_{xy} = 29{,}083 \cdot \frac{1{,}623}{48{,}119} = 0{,}9809$

Na osnovi rezultata ($r_{xy} = 0{,}9809$) vidimo, da obstaja močna povezanost med odvisno (varčevanje) in neodvisno spremenljivko (obrestna mera). Smer povezanosti je pozitivna.

Izračunati moramo še determinacijski koeficient:

$r_{xy}^2 = 0{,}9809^2 = 0{,}9622$ oz. 96,22 %

Standardno napako ocene odvisne spremenljivke izračunamo po enačbi (v obrazcih št. 2.10):

$\sigma_{ey} = \sigma_y \sqrt{1 - r_{xy}^2}$

$\sigma_{ey} = 48{,}119 \cdot \sqrt{1 - 0{,}9622} = 9{,}355$

Standardna napaka ocene odvisne spremenljivke je različna od 0, kar pomeni, da na znesek kratkoročnih oblik varčevanja (odvisna spremenljivka) poleg obrestne mere (neodvisna spremenljivka) vplivajo še druge spremenljivke in slučajni vplivi.

$P(238{,}443 - 2 \cdot 9{,}355 < y_{x=7} < 238{,}443 + 2 \cdot 9{,}355) = 95{,}4\ \%$

$P(219{,}733 < y_{x=7} < 257{,}153) = 95{,}4\ \%$

Pri obrestni meri x = 7 (v %) bo znesek kratkoročnih oblik varčevanja med 219,733 d.e. in 257,153 d.e., kar trdimo s 95,4-% verjetnostjo.

Naloga 32

$P(\hat{y}_x - 2 \cdot \sigma_{ey} < y_x < \hat{y}_x + 2 \cdot \sigma_{ey}) = 0{,}954 = 95{,}4\ \%$

Izračunamo: $\hat{y} = a + b \cdot x_i$

Za izračun regresijskega koeficienta b uporabimo enačbo (v obrazcih št. 2.2):

$b = \frac{c_{xy}}{\sigma_x^2}$

$b = \frac{270}{29{,}6} = 10{,}037$

Regresijski koeficient a izračunamo po enačbi (v obrazcih št. 2.4): $a = \bar{y} - b \cdot \bar{x}$

$\bar{y} = \frac{1680}{8} = 210$

$\bar{x} = \frac{130}{8} = 16{,}25$

$a = 210 - 10{,}037 \cdot 16{,}25 = 46{,}899$

$\hat{y} = a + b \cdot x_i \quad (x_i = 20)$

$\hat{y} = 46{,}899 + 10{,}037 \cdot 20$

$\hat{y} = 247{,}639$

Standardno napako ocene odvisne spremenljivke izračunamo po enačbi (v obrazcih št. 2.10):

$\sigma_{ey} = \sigma_y \sqrt{1 - r_{xy}^2}$

$r_{xy} = \dfrac{Cxy}{\sigma x \cdot \sigma y}$

$r_{xy}^2 = (\dfrac{Cxy}{\sigma x \cdot \sigma y})^2 = (\dfrac{270}{5{,}187 \cdot 55{,}351})^2 = 0{,}8844$ oz. 88,44 %

$\sigma_{ey} = 55{,}351 \cdot \sqrt{1 - 0{,}8844} = 18{,}819$

$P(247{,}639 - 2 \cdot 18{,}819 < y_{x=20} < 247{,}639 + 2 \cdot 18{,}819) = 95{,}4\ \%$

$P(210{,}001 < y_{x=20} < 285{,}277) = 95{,}4\ \%$

Pri delovni dobi x = 20 let znaša osebni dohodek sodnika med 210,001 d.e. in 285,277 d.e., kar trdimo s 95,4-% verjetnostjo.

Naloga 34 b)

$\Sigma\, x_i = 319$

$\Sigma\, x_i^2 = 26791$

$\Sigma\, y_i = 122$

$\Sigma\, y_i^2 = 3264$

$\Sigma\, x_i \cdot y_i = 9012$

$\bar{x} = 63{,}8$

$\bar{y} = 24{,}4$

Izračun regresijskih koeficientov:

b = 0,193

a = 12,087

Izračunana regresijska koeficienta vstavimo v enačbo regresijske premice:

$\hat{y} = a + b \cdot x_i$

$\hat{y} = 12{,}087 + 0{,}193 \cdot x_i$

Korelacijski koeficient izračunamo po enačbi (v obrazcih št. 2.9):

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y}$$

Za izračun σ_x in σ_y uporabimo enačbo za skupno varianco (v obrazcih št. 2.7):

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

$\sigma_x^2 = 1287{,}78$

$\sigma_x = 35{,}885$

$\sigma_y^2 = 57{,}44$
$\sigma_y = 7{,}579$
$r_{xy} = 0{,}193 \cdot \frac{35{,}885}{7{,}579} = 0{,}914$

Determinacijski koeficient:
$r_{xy}^2 = 0{,}914^2 = 0{,}8354$ oz. 83,54 %

Standardno napako ocene odvisne spremenljivke izračunamo po enačbi (v obrazcih št. 2.10):

$\sigma_{ey} = \sigma_y \sqrt{1 - r_{xy}^2}$

$\sigma_{ey} = 7{,}579 \cdot \sqrt{1 - 0{,}8354} = 3{,}075$

Naloga 34 c)

$x = 180$
$\hat{y} = 12{,}087 + 0{,}193 \cdot x_i$
$\hat{y}_{x=180} = 12{,}087 + 0{,}193 \cdot 180$
$\hat{y}_{x=180} = 46{,}827$
$P(46{,}827 - 2 \cdot 3{,}075 < y_{x=180} < 46{,}827 + 2 \cdot 3{,}075) = 95\ \%$
$P(40{,}677 < y_{x=180} < 52{,}977) = 95\ \%$

Pri številu zaposlenih x = 180 zaposlenih znaša znesek stroškov za okoljevarstvene dejavnosti podjetja med 40,677 d.e. in 52,977 d.e., kar trdimo s 95-% verjetnostjo.

Naloga 35

Povprečni odstotek zasedenosti je neodvisna spremenljivka x.

Cena v dolarjih je odvisna spremenljivka y.

$\Sigma\, x_i = 501{,}2$
$\Sigma\, x_i^2 = 36.105{,}98$
$\Sigma\, y_i = 689{,}91$
$\Sigma\, y_i^2 = 74.444{,}984$
$\Sigma\, x_i \cdot y_i = 50.210{,}032$
$\overline{x} = 71{,}6$
$\overline{y} = 98{,}559$

Izračun regresijskih koeficientov:
$b = 3{,}691$
$a = -165{,}717$

Izračunana regresijska koeficienta vstavimo v enačbo regresijske premice:
$\hat{y} = -165{,}717 + 3{,}691 \cdot x_i$
$\hat{y}_{x=95} = -165{,}717 + 3{,}691 \cdot 95$
$\hat{y}_{x=95} = 516{,}362$

Izračunamo korelacijski koeficient:
$\sigma_x^2 = 31{,}437$
$\sigma_x = 5{,}607$
$\sigma_y^2 = 921{,}206$
$\sigma_y = 30{,}351$
$r_{xy} = 3{,}691 \cdot \frac{5{,}607}{30{,}351} = 0{,}682$

Na osnovi rezultata ($r_{xy} = 0{,}682$) vidimo, da obstaja srednje močna povezanost med odvisno (cena v dolarjih) in neodvisno spremenljivko (povprečni % zasedenosti). Smer povezanosti je pozitivna.

Izračunamo determinacijski koeficient:
$r_{xy}^2 = 0{,}682^2 = 0{,}4651$ oz. 46,51 %

Delež pojasnjene variance v skupni varianci za odvisno spremenljivko znaša 46,51 %.
Izračunamo standardno napako ocene odvisne spremenljivke (v obrazcih št. 2.10):
$\sigma_{ey} = 30{,}351 \cdot \sqrt{1 - 0{,}4651} = 22{,}198$

Izračun intervalne ocene s 95,4-% verjetnostjo:

$P(516{,}362 - 2 \cdot 22{,}198 < y_{x=95} < 516{,}362 + 2 \cdot 22{,}198) = 95{,}4\ \%$
$P(471{,}966 < y_{x=95} < 560{,}758) = 95{,}4\ \%$

Pri povprečni odstotni zasedenosti x= 95 (v %) bo cena za sobo med 471,966 USD in 560,758 USD, kar trdimo s 95,4-% verjetnostjo.

Naloga 36 a)

$\Sigma x_i = 722$
$\Sigma x_i^2 = 87244$
$\Sigma y_i = 56$
$\Sigma y_i^2 = 420$
$\Sigma x_i \cdot y_i = 5700$
$\overline{x} = 90{,}25$
$\overline{y} = 7$

Izračun regresijskih koeficientov:

$$b = \frac{\frac{1}{8} \cdot (5700) - 90{,}25 \cdot 7}{\frac{1}{8} \cdot (87244) - 90{,}25^2} = 0{,}029$$

$$a = 1 - 0{,}029 \cdot 90{,}25 = 4{,}383$$

Izračun korelacijskega koeficienta (v obrazcih št. 2.9):

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y}$$

Za izračun σ_x in σ_y uporabimo enačbo za skupno varianco (v obrazcih št. 2.7):

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

$\sigma_x^2 = 2760{,}438$
$\sigma_x = 52{,}54$
$\sigma_y^2 = 3{,}5$
$\sigma_y = 1{,}87$

$$r_{xy} = 0{,}029 \cdot \frac{52{,}54}{1{,}87} = 0{,}815$$

Izračun determinacijskega koeficienta:

$r_{xy}^2 = 0{,}815^2 = 0{,}6642$ oz. 66,42 %

Standardno napako ocene odvisne spremenljivke izračunamo po enačbi (v obrazcih št. 2.10):

$$\sigma_{ey} = \sigma_y \sqrt{1 - r_{xy}^2}$$

$$\sigma_{ey} = 1{,}87 \cdot \sqrt{1 - 0{,}6642} = 1{,}084$$

Naloga 36 b)

Enačba regresijske premice:

$\hat{y} = a + b \cdot x_i$

$\hat{y} = 4{,}383 + 0{,}029 \cdot x_i$

Pomen regresijskega koeficienta a: Pri porabi časa, ki ga študenti namenijo za učenje zadnji dan pred izpitom Poslovna statistika x = 0, lahko v povprečju pričakujemo, da bo ocena izpita znašala 4,383 enot.

Pomen regresijskega koeficienta b: Če se poraba časa, ki ga študenti namenijo za učenje zadnji dan pred izpitom Poslovna statistika *(x)*, poveča za eno enoto, se ocena izpita *(y)* v povprečju poveča za 0,029 enot.

Naloga 37 b)

$n = 7$

$\Sigma\, x_i = 1.446$

$\Sigma\, x_i^2 = 306.990$

$\Sigma\, y_i = 1.598$

$\Sigma\, y_i^2 = 383.188$

$\Sigma\, x_i \cdot y_i = 338.827$

$\bar{x} = 206{,}571$

$\bar{y} = 228{,}286$

$b = 1{,}053$

Naloga 37 c)

$\sigma_x^2 = 40.681{,}55$

$\sigma_x = 201{,}697$

$\sigma_y^2 = 2.626{,}78$

$\sigma_y = 51{,}252$

$r_{xy} = 4{,}144$

Naloga 38 b)

y = Motiviranost zaposlenih

x = Izplačilo denarne nagrade

$\Sigma\, x_i = 150 + 180 + 225 + 260 + 285 + 300 = 1400$

$\Sigma\, x_i^2 = 150^2 + 180^2 + 225^2 + 260^2 + 285^2 + 300^2 = 344.350$

$\Sigma\, y_i = 26 + 35 + 63 + 76 + 88 + 95 = 383$

$\Sigma\, y_i^2 = 26^2 + 35^2 + 63^2 + 76^2 + 88^2 + 95^2 = 28.415$

$\Sigma\, x_i \cdot y_i = (150 \cdot 26) + (180 \cdot 35) + (225 \cdot 63) + \ldots + (300 \cdot 95) = 97.715$

$\bar{x} = \frac{1400}{6} = 233{,}333$

$\bar{y} = \frac{383}{6} = 63{,}833$

Oba regresijska koeficienta izračunamo po enačbi (v obrazcih št. 2.3 in 2.4):

$$b = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} = \frac{\frac{1}{N}\sum_{i=1}^{N}x_i y_i - \bar{x}\cdot\bar{y}}{\frac{1}{N}\sum_{i=1}^{N}x_i^2 - \bar{x}^2}$$

$$a = \bar{y} - b\bar{x}$$

$$\text{b} = \frac{\frac{1}{6}\cdot(97715) - 233{,}333 \cdot 63{,}833}{\frac{1}{6}\cdot(344350) - 233{,}333^2} = 0{,}472$$

$a = 63{,}833 - 0{,}472 \cdot 233{,}333 = -46{,}300$

Izračunana regresijska koeficienta vstavimo v enačbo regresijske premice:

$\hat{y} = a + b \cdot x_i$

$\hat{y} = -46{,}300 + 0{,}472 \cdot x_i$

Pomen regresijskega koeficienta a: Pri izplačilu denarne nagrade (v €) x = 0 lahko v povprečju pričakujemo, da bo motivacija zaposlenih –46,300 enot.

Pomen regresijskega koeficienta b: Če se izplačilo denarne nagrade *(x)* poveča za eno enoto (v €), se motivacija zaposlenih *(y)* v povprečju poveča za 0,472 enot.

Naloga 38 c)

x = 310 €

$\hat{y} = -46{,}300 + 0{,}472 \cdot x_i$

$\hat{y}_{x=310} = -46{,}300 + 0{,}472 \cdot 310$

$\hat{y}_{x=310}$ = 100 enot (motivacija zaposlenih na lestvici od 1 do 100)

Naloga 38 d)

Parameter, na osnovi katerega določimo smer in jakost linearne korelacijske odvisnosti, je korelacijski koeficient.

Izračunamo ga po enačbi (v obrazcih št. 2.9):

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y}$$

Za izračun σ_x in σ_y uporabimo enačbo za skupno varianco (v obrazcih št. 2.7):

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

$\sigma_y^2 = \frac{1}{6} \cdot [(26 - 63{,}833)^2 + (35 - 63{,}833)^2 + \ldots + (95 - 63{,}833)^2] = 661{,}139$

$\sigma_y = \sqrt{661{,}139} = 25{,}713$

$\sigma_x^2 = \frac{1}{6} \cdot [(150 - 233{,}333)^2 + (180 - 233{,}333)^2 + \ldots + (300 - 233{,}333)^2] = 2947{,}222$

$\sigma_x = \sqrt{2947{,}222} = 54{,}288$

$r_{xy} = 0{,}472 \cdot \frac{54{,}288}{25{,}713} = 0{,}997$

Naloga 38 e)

Izračunati moramo delež pojasnjene variance v skupni varianci za odvisno spremenljivko (determinacijski koeficient):

$r_{xy}^2 = 0{,}997^2 = 0{,}994$ oz. 99,4 %

Delež pojasnjene variance v skupni varianci za odvisno spremenljivko znaša 99,4 %.

Naloga 38 f)

Standardno napako ocene odvisne spremenljivke izračunamo po enačbi (v obrazcih št. 2.10):

$$\sigma_{ey} = \sigma_y\sqrt{1 - r_{xy}^2}$$

$\sigma_{ey} = 25{,}713 \cdot \sqrt{1 - 0{,}994} = 1{,}992$

Naloga 38 g)

x = 295 €

$\hat{y} = -46{,}300 + 0{,}472 \cdot x_i$

$\hat{y}_{x=295} = -46{,}300 + 0{,}472 \cdot 295$

$\hat{y}_{x=295} = 92{,}94$ enot
$P(92{,}94 - 2 \cdot 1{,}992 < y_{x=295} < 92{,}94 + 2 \cdot 1{,}992) = 95{,}4\ \%$
$P(88{,}956 < y_{x=295} < 96{,}924) = 95\ \%$

7.5 Osnove vzorčenja

Naloga 39

Dvostranski interval za povprečno porabo izdelka v osnovni statistični množici s 95-% verjetnostjo:

Imamo velik vzorec, ker je n = 200, ter dvostransko intervalno ocenjevanje aritmetične sredine:
$\gamma = 95\ \%,\ \alpha = 5\ \%,\ z = \pm 1{,}96$

Izračunamo standardno napako ocene aritmetične sredine (v obrazcih št. 4.4):
$$SE_{\bar{y}} = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$$
$$se_{\bar{y}} = \frac{6{,}2}{\sqrt{200}} = 0{,}438$$

Uporabimo enačbo za dvostransko ocenjevanje aritmetične sredine za velik vzorec (v obrazcih št. 4.1):
$P\ (\bar{Y} - z \cdot se_{\bar{y}} < \bar{y} < \bar{Y} + z \cdot se_{\bar{y}}) = \gamma$
$P(20{,}8 - 1{,}96 \cdot 0{,}438 < \bar{y} < 20{,}8 - 1{,}96 \cdot 0{,}438) = 95\ \%$
$P(19{,}941 < \bar{y} < 21{,}658) = 95\ \%$

S 95-% verjetnostjo ocenjujemo, da je povprečna poraba izdelka v osnovni statistični množici med 19,941 in 21,658 kosov.

Naloga 40

Imamo mali vzorec, ker je n = 6, ter dvostransko intervalno ocenjevanje aritmetične sredine (v obrazcih št. 4.21):
$\alpha = 5\ \%$
$\bar{Y} = 32{,}83$
$s^2 = 537{,}77$
$s = 23{,}19$

$se_{\bar{y}} = 9{,}467$

$t_{n-1;\,\alpha/2} = t_{5;\,0{,}025} = 2{,}571$ (gledamo tabelo *kritična vrednost za t porazdelitev*)
$\alpha/2 = 0{,}05/2 = 0{,}025$

$P(32{,}83 - 2{,}571 \cdot 9{,}467 < \bar{y} < 32{,}83 + 2{,}571 \cdot 9{,}467) = 95\ \%$
$P(8{,}490 < \bar{y} < 57{,}169) = 95\ \%$

S 95-% verjetnostjo ocenjujemo, da povprečen odstotek žensk med zaposlenimi v vseh podjetjih znaša med 8,490 % in 57,169 %.

Naloga 41

$n = 625$ (velik vzorec)
$n_a = 125$
$\alpha = 20\ \% \rightarrow z = \pm 1{,}28$

Izračun strukturnega odstotka iz vzorca (v obrazcih št. 4.13):

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

$$p = \frac{125}{625} \cdot 100 = 20\ \%$$

Izračun standardne napake ocene strukturnega odstotka (v obrazcih št. 4.14):

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

$$se_{\pi} = \sqrt{\frac{20(100-20)}{625}} = 1{,}6$$

Dvostransko ocenjevanje strukturnega odstotka (v obrazcih št. 4.12):
$P(p - z \cdot se_{\pi} < \pi < p + z \cdot se_{\pi}) = \gamma$
$P(20 - 1{,}28 \cdot 1{,}6 < \pi < 20 - 1{,}28 \cdot 1{,}6) = 80\ \%$
$P(17{,}95 < \pi < 22{,}05) = 80\ \%$

Z 80-% verjetnostjo ocenjujemo, da je odstotek (%) nezaposlenih, ki prejemajo plačila na osnovi »dela na črno«, med 17,95 % in 22,05 %.

Naloga 42

H0: $\bar{y}_D$ = 5.000 EUR

H1: $\bar{y}_D \neq$ 5.000 EUR

$\bar{y}$ = 5.200

s^2 = 800.000

s = 894,43

$se_{\bar{y}}$ = 149,07

α = 10% (oziroma 0,10), kar pomeni, da je kritična vrednost spremenljivke z = ± 1,645

Izračunamo *z* po enačbi (izračun testne vrednosti pri preizkušanju domneve o aritmetični sredini – z-test, v obrazcih št. 4.22):

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

$z = \frac{5200-5000}{149{,}07}$ = 1,34 (vidimo, da dobljeno število pade v interval ± 1,645)

Sprejmemo domnevo H_0.

Naloga 43

α = 1 % ⟶ $t_{n-1;\,\alpha/2} = t_{4;0{,}005}$ = 4,604 (gledamo tabelo kritična vrednost za t porazdelitev)

Zastavimo ničelno in raziskovalno domnevo:

H0: $\bar{y}_D$ = 20 cm

H1: $\bar{y}_D \neq$ 20 cm

$SE_{\bar{y}} = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$ (v obrazcih št. 4.23)

$se_{\bar{y}} = \frac{0{,}2}{\sqrt{5}}$ = 0,089

Izračunamo *t* po enačbi (izračun testne vrednosti pri preizkušanju domneve o aritmetični sredini – t-test, v obrazcih št. 4.24):

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

$t = \frac{20{,}3 - 20}{0{,}089}$ = 3,37 (vidimo, da dobljeno število pade v interval ± 4,604)

Sprejmemo domnevo H0.

Naloga 44 a)

Velik vzorec; dvostransko intervalno ocenjevanje aritmetične sredine

$\overline{Y} = 60{,}515$

Za izračun vzorčne variance iz frekvenčne porazdelitve uporabimo enačbo (v obrazcih št. 4.9):

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

$s^2 = 3446{,}121$

$s = 58{,}704$

$se_{\bar{y}} = 3{,}184$

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

$P(60{,}515 - 1{,}96 \cdot 3{,}184 < \bar{y} < 60{,}515 + 1{,}96 \cdot 3{,}184) = 95\ \%$

$P(54{,}274 < \bar{y} < 66{,}756) = 95\ \%$

S 95-% verjetnostjo ocenjujemo, da se povprečni znesek, ki so ga v vrednostne papirje vložili imetniki vrednostnih papirjev pri banki X, giblje med 54,274 € in 66,756 €.

Naloga 44 b)

$N = 3.648$

$\overline{Y} = 60{,}515$

$SE_{\bar{y}} = 3{,}184$

Ocena totala iz vzorca (v obrazcih št. 4.10 in 4.11): $\hat{Y} = N \cdot \overline{Y} = 3.648 \cdot 60{,}515 = 220.758{,}72$

Standardna napaka ocene totala (v obrazcih št. 4.10 in 4.11): $SE_y = N \cdot SE_{\bar{y}} = 3.648 \cdot 3{,}184 = 11.615{,}232$

Kritična vrednost spremenljivke $z = \pm 1{,}645$

$P(220.758{,}72 - 1{,}645 \cdot 11.615{,}232 < Y < 220.758{,}72 + 1{,}645 \cdot 11.615{,}232) = 90\ \%$

$P(201.651{,}663 < Y < 239.865{,}777) = 90\ \%$

Naloga 44 c)

Velik vzorec; dvostransko preizkušanje domneve o aritmetični sredini

H0: $\bar{y}_D = 75$ €

H1: $\bar{y}_D \neq 75$ €

$\overline{Y} = 60{,}515$

$SE_{\bar{y}} = 3{,}184$

Kritična vrednost spremenljivke $z = \pm\ 1{,}645$ ($\alpha = 0{,}10$)

$z = -4,\ 549$

Sprejmemo domnevo H_1.

Naloga 44 d)

Velik vzorec; dvostransko preizkušanje domneve o strukturnem deležu.

Zastavimo ničelno in raziskovalno domnevo:

H0: $\pi_D\ = 35\ \%$

H1: $\pi_D\ \neq 35\ \%$

Kritična vrednost spremenljivke $z = \pm\ 2{,}58$ ($\alpha = 1\%$)

$n = 340$

$n_a = 92 + 41 = 133$

Za izračun strukturnega odstotka iz vzorca uporabimo enačbo (v obrazcih št. 4.13):

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

$p = 39{,}12$

Za izračun standardne napake ocene strukturnega odstotka uporabimo enačbo (v obrazcih št. 4.14):

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

$SE\pi = 2{,}65$

$$z = \frac{39{,}12 - 35}{2{,}65} = 1{,}55$$

Sprejmemo domnevo H_0.

Izračun za največ 35 % (enostransko preizkušanje domneve):

H0: $\pi_D\ \leq 35\ \%$

H1: $\pi_D > 35\ \%$

$n = 340$

$n_a = 92 + 41 = 133$

$p = 39{,}12$

$SE\pi = 2{,}65$

Kritična vrednost spremenljivke $z = +\ 2{,}33$

$z = 1{,}55$

Sprejmemo domnevo H_0.

Naloga 44 e)

Velik vzorec; dvostransko preizkušanje domneve o totalu

H0: $\bar{y}_D = 20.000.000$ EUR

H1: $\bar{y}_D \neq 20.000.000$ EUR

$N = 3648$

$\overline{Y} = 60{,}515$

$s = 58{,}704$

$SE_{\bar{y}} = 3{,}184$

Ocena totala iz vzorca: $\hat{Y} = 220.758{,}72$

Standardna napaka ocene totala: $SE_Y = 11.615{,}232$

Kritična vrednost spremenljivke z: $z = 1{,}28$ ($\gamma = 80\% \rightarrow \alpha = 20\ \%$)

$z = 1{,}787$

Sprejmemo domnevo H_1.

Naloga 45 a)

Mali vzorec; dvostransko ocenjevanje aritmetične sredine.

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

$\bar{y} = 700$

$s = 259{,}808$

$se_{\bar{y}} = 86{,}60$

Kritična vrednost spremenljivke $t_{8;0,10} = 1{,}379$

$P(700 - 1{,}397 \cdot 86{,}60 < \bar{y} < 700 + 1{,}397 \cdot 86{,}60) = 80\ \%$

$P(579{,}02 < \bar{y} < 820{,}98) = 80\ \%$

Naloga 45 b)

$\bar{Y} = 700$

$s = 259{,}808$

$se_{\bar{y}} = 86{,}60$

$N = 2.860$

Ocena totala iz vzorca: $\hat{Y} = 2.860 \cdot 700 = 2.002.000$

$SE_Y = 2.860 \cdot 86{,}60 = 247.676$

Mali vzorec; dvostransko preizkušanje domneve o totalu

$H0: \bar{y}_D = 1.800.000$ EUR

$H1: \bar{y}_D \neq 1.800.000$ EUR

Kritična vrednost spremenljivke t: $t_{8;0,05} = \pm\ 1{,}86$ ($\alpha = 0{,}10$)

$t = 0{,}816$

Sprejmemo domnevo H_0.

Naloga 46

$n = 50$

$\bar{Y} = 102$

$s^2 = 6{,}25 \ \rightarrow s = 2{,}5$

$se_{\bar{y}} = \frac{2{,}5}{\sqrt{50}} = 0{,}35$

$\alpha = 5\% \ \rightarrow z = \pm\ 1{,}96$

Zastavimo ničelno in raziskovalno domnevo:

$H0: \bar{y}_D = 100$ EUR

$H1: \bar{y}_D \neq 100$ EUR

Izračunamo *z* po enačbi (izračun testne vrednosti pri preizkušanju domneve o aritmetični sredini – z-test, v obrazcih št. 4.22):

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

$z = \frac{102 - 100}{0{,}35} = 5{,}71$ (vidimo, da dobljeno število ne pade v interval ± 1,96)

Na osnovi rezultata vidimo, da delavci ne izdelajo povprečno 100 izdelkov v določeni časovni enoti, tako da zavrnemo domnevo H0 in sprejmemo raziskovalno domnevo H1.

Naloga 47

Število delavcev	Število podjetij (f_k)	Sredina razreda (y_k)	$(y_k - \bar{Y})^2$
Od 1 do 5	3	3	272,25
Od 6 do 10	4	8	132,25
Od 11 do 15	2	13	42,25
Od 16 do 20	1	18	2,25
Od 21 do 25	1	23	12,25
Od 26 do 30	5	28	72,25
Od 31 do 35	2	33	182,25
Od 36 do 40	2	38	342,25
Skupaj	20		

Uporabimo enačbo za aritmetično sredino iz frekvenčne porazdelitve (v obrazcih št. 1.25):

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

$\bar{Y} = \frac{1}{20} \cdot 390 = 19{,}5$ EUR

Uporabimo enačbo za vzorčno varianco iz frekvenčne porazdelitve (v obrazcih št. 4.9):

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

$s^2 = \frac{1}{19} \cdot 2855 = 150{,}26$ EUR^2

Standardni odklon:

$s = \sqrt{150{,}26} = 12{,}26$ EUR

Nato izračunamo standardno napako ocene aritmetične sredine (v obrazcih št. 4.4):

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

$se_{\bar{y}} = \frac{12{,}26}{\sqrt{20}} = 2{,}74$

Imamo mali vzorec, ker je n = 20, ter dvostransko intervalno ocenjevanje aritmetične sredine (v obrazcih št. 4.21):

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

$P(19{,}5 - 2{,}861 \cdot 2{,}74 < \overline{y} < 19{,}5 + 2{,}861 \cdot 2{,}74) = 99\ \%$

$P(11{,}661 < \overline{y} < 27{,}339) = 99\ \%$

Z 99-% verjetnostjo ocenjujemo, da je povprečno število zaposlenih delavcev v vseh podjetjih te gospodarske panoge med 11,661 in 27,339 zaposlenih delavcev.

$t_{n-1;\,\alpha/2} = t_{19;\,0{,}005} = 2{,}861$ (gledamo *tabelo kritična vrednost za t porazdelitev*) $\alpha/2 = 0{,}01/2 = 0{,}005$ $(\gamma = 99\ \% \rightarrow \alpha = 1\ \%)$

s = 5 minut

$se_{\bar{y}} = 0{,}408$

$\gamma = 85\ \% \rightarrow \alpha = 15\ \% \rightarrow z = 1{,}44$

$P(12{,}5 - 1{,}44 \cdot 0{,}408 < \overline{y} < 12{,}5 + 1{,}44 \cdot 0{,}408) = 85\ \%$

$P(11{,}91 < \overline{y} < 13{,}09) = 85\ \%$

Naloga 48 b)

$\overline{Y} = 12{,}5$ minut

$se_{\bar{y}} = 0{,}408$

H0: $\bar{y}_D = 10$ minut

H1: $\bar{y}_D \neq 10$ minut

Kritična vrednost spremenljivke $z = \pm\ 2{,}58$ ($\alpha = 1\ \%$)

$z = 6{,}13$

Sprejmemo domnevo H_1.

Naloga 49

$\alpha = 1\ \% \rightarrow t_{n-1;\,\alpha/2} = t_{9;0{,}005} = 3{,}250$

$se_{\bar{y}} = 7{,}906$

$t = -0{,}632$

Sprejmemo domnevo H0.

Naloga 50

$\bar{Y} = 68.750$ d.e.

$s = 460$ d.e.

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

$se_{\bar{y}} = 72{,}73$ d.e.

$P(68.750 - 1{,}96 \cdot 72{,}73 < \bar{y} < 68.750 + 1{,}96 \cdot 72{,}73) = 95\ \%$

$P(68.607{,}75 < \bar{y} < 68.892{,}55) = 95\ \%$

Naloga 51 a)

$n = 7$ (mali vzorec)

$\gamma = 80\ \% \rightarrow \alpha = 20\ \% \rightarrow t_{6;0,10} = 1{,}440$

$\bar{Y} = 27{,}286$

$s = 37{,}622$

$se_{\bar{y}} = 14{,}219$

$P(27{,}286 - 1{,}440 \cdot 14{,}219 < \bar{y} < 27{,}286 - 1{,}440 \cdot 14{,}219) = 80\ \%$

$P(6{,}811 < \bar{y} < 47{,}761) = 80\ \%$

Naloga 51 b)

Enostransko preizkušanje domneve:

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

H_0: $\bar{y}_D \leq 30$

H_1: $\bar{y}_D > 30$

$se_{\bar{y}} = 14{,}219$

$t = -0{,}191$

Sprejmemo domnevo H_0.

Naloga 52

$\bar{Y} = 20{,}5$

$s = 1{,}9$

$se_{\bar{y}} = 0{,}95$

H_0: $\bar{y}_D$ = 22 miligramov

H_1: $\bar{y}_D \neq$ 22 miligramov

Kritična vrednost spremenljivke t: $t_{3;0,025} = \pm 3,182$ ($\alpha = 0,05$)

$t = -1,58$

Sprejmemo domnevo H_0.

Naloga 53 a)

$\bar{Y} = 223,2$

$s = 28,665$

$se_{\bar{y}} = 12,819$

$z = \pm 1,96$

$P(223,2 - 1,96 \cdot 12,819 < \bar{y} < 223,2 + 1,96 \cdot 12,819) = 95\ \%$

$P(198,075 < \bar{y} < 248,325) = 95\ \%$

Naloga 53 b)

H_0: $\bar{y}_D \geq 220$

H_1: $\bar{y}_D < 220$

Kritična vrednost spremenljivke $z = -1,28$ ($\alpha = 10\ \%$)

$z = 0,249$

Sprejmemo domnevo H_0.

7.6 Časovne vrste

Naloga 54 a)

Leto	Živorojeni	$I_{t/3}$ v %	V_t
1	1,94	103,7	/
2	1,89	101,1	97,4
3	1,87	100	98,9
4	1,82	97,3	97,3
5	1,78	95,2	97,8
6	1,75	93,6	98,3
7	1,81	96,8	103,4

Naloga 54 b)

Leto	Srednje štev. preb. (30.6.)	Živorojeni	Koeficient natalitete	Umrli	Koeficient mortalitete
1	198	1,94	9,8	1,93	9,7
2	198	1,89	9,5	1,89	9,5
3	199	1,87	9,4	1,86	9,4
4	198	1,82	9,2	1,89	9,5
5	198	1,78	9,0	1,90	9,6
6	198	1,75	8,8	1,88	8,8
7	199	1,81	9,1	1,85	9,3

Koeficient natalitete meri število živorojenih na 1000 prebivalcev, koeficient mortalitete pa število mrtvih na 1000 prebivalcev.

Naloga 55 a)

Leto	$I_{t/2017}$	Koeficient dinamike K_t	Y_t
2017	100	/	130
2018	110	1,1	143
2019	125	1,136	162,448
2020	117	0,936	152,051
2021	120	1,026	156,004
2022	118	0,983	153,352

Za koeficient dinamike uporabimo enačbo (v obrazcih št. 1.10):

$K_t = \frac{Yt}{Yt-1}$ $K_1 = /$ *za T= 2,…T*

Na primer:

$K_{t/2018} = \frac{110}{100} = 1,1$

$K_{t/2019} = \frac{125}{110} = 1,136$

Naloga 55 b)

Uporabimo enačbo za povprečni koeficient dinamike (v obrazcih št. 1.26):

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

$$K = \sqrt[6-1]{\frac{118}{100}} = \sqrt[5]{1,18} = 1,034$$

Uporabimo enačbo za povprečno stopnjo rasti (v obrazcih št. 1.30):

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

$S = (1{,}034 - 1) \cdot 100 = 3{,}4$ %

Odgovor: Število študentov, ki so opravljali prakso v podjetju X, se je v letih od 2017 do 2022 povečalo za 3,4 % na leto.

Naloga 56 b), c)

Leto	1	2	3	4	5	6	7	8	9
S_t%	/	+18,0	–2,1	+2,7	+2,5	+2,5	–5,0	+7,0	–5,8
V_t	/	118	97,9	102,7	102,5	102,5	95	107	94,2
K_t	/	1,18	0,979	1,027	1,025	1,025	0,95	1,07	0,942
$I_{t/4}$	84,3	99,5	97,4	100	102,5	105,1	99,8	106,8	100,6
Y_t (v kg)	71,2	84	82,3	84,5	86,6	88,8	84,3	90,2	85,0

Naloga 57

Uporabimo enačbo za statistični koeficient (v obrazcih št. 1.15):

$$K = \frac{x \cdot E}{\bar{y} \cdot i}$$

x (seštevek intervalnega podatka) = 652 + 730 + 840 + 752 = 2974

E (na koliko enot izražamo kazalec: na 10 zaposlenih) = 10

i (podana je vrednost proizvodnje za 4 mesece) = 4

Trenutni podatek se nanaša na začetek in konec časovnih razmikov, zato uporabimo enačbo (v obrazcih št. 1.17) : $\bar{y} = \frac{1}{T}\left(\frac{1}{2} \cdot y_0 + y_1 + y_2 + \ldots + \frac{1}{2} \cdot y_T\right)$

T (število časovnih razmikov) = 4

$$\bar{y} = \frac{1}{4}\left(\frac{1}{2} \cdot 214 + 240 + 226 + 208 + \frac{1}{2} \cdot 200\right) = 220{,}25$$

Povprečni statistični koeficient – povprečna vrednost proizvodnje na 10 zaposlenih – za opazovano obdobje je enak (v 10^6 EUR):

$$K = \frac{2974 \cdot 10}{220{,}25 \cdot 4} = 33{,}757$$

Naloga 58 a)

Mesec	I.	II.	III.	IV.	V.
Povprečna mesečna plača zaposlenih (v d.e.)	2600	2800	3100	2900	3200
Povprečna mesečna plača zaposlenih – $I_{t/V}$ (v %)	81,25	87,5	96,9	90,6	100

Naloga 58 b)

Mesec	I.	II.	III.	IV.	V.
Mesečna stopnja rasti vrednosti prodaje S_t %	–	+15	–10	+8	–12
Mesečna stopnja rasti vrednosti prodaje K_t	–	1,15	0,90	1,08	0,88
Mesečna stopnja rasti vrednosti prodaje – $I_{t/I}$ (v %)	100	115	103,5	111,8	98,4

Naloga 58 c)

Povprečna mesečna plača zaposlenih

$$K = \sqrt[4]{\frac{3200}{2600}} = 1{,}053$$

S = +5,3 %

Vrednost prodaje:

$$K = \sqrt[4]{\frac{98{,}4}{100}} = 0{,}9960$$

S = –0,4 %

V opazovanem obdobju so povprečne plače naraščale povprečno za 5,3 % na mesec, vrednost proizvodnje pa se je v povprečju zmanjševala za 0,4 % mesečno.

Naloga 58 d)

Napoved povprečne mesečne plače za oktober (v d.e.):

$$Y_t = 3.200 \cdot 1{,}053^5 = 4.142{,}8$$

Naloga 59 a)

Verižni indeks za 4. časovno enoto, v % (v obrazcih št. 1.8):

$$V_t = 100 \times \frac{Y_t}{Y_{t-1}} \qquad za\ t = 2,3,\ \cdots,\ T$$

$$V_4 = \frac{400}{118} \cdot 100 = 338{,}98$$

Naloga 59 b)

Izračunamo po enačbi indeksov s stalno osnovo iz časovnih vrst (v obrazcih št. 1.7):

$$I_{t/o} = 100 \times \frac{Y_t}{Y_o} \qquad za\ t = 1,2,\cdots,T$$

Indeks z osnovo 1, za 7. časovno enoto:, v %

$$I_{7/1} = \frac{300}{115} \cdot 100 = 260{,}87$$

Naloga 60 a, b)

Leto	1	2	3	4
S_t (v %)	/	+18,4	+11,7	–9,2
K_t	/	1,184	1,117	0,908
$I_{t/2}$ (v %)	84,46	100	111,7	101,42
Y_t	128,92	152,64	170,5	154,81

Naloga 60 c)

Uporabimo enačbo za povprečni koeficient dinamike (v obrazcih št. 1.26):

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

$$K = \sqrt[3]{\frac{101{,}42}{84{,}46}} = 1{,}063$$

Naloga 60 d)

Napoved za 7. časovno enoto, v d.e.:

$$Y_7 = 154{,}81 \cdot 1{,}063^3 = 185{,}90$$

Naloga 61 b)

Uporabimo enačbo za povprečni koeficient dinamike (v obrazcih št. 1.26):

$$K = \sqrt[T-1]{\frac{Y_T}{Y_1}}$$

$$K = \sqrt[5]{\frac{325}{215}} = \sqrt[5]{1{,}512} = 1{,}086$$

$$S = (K-1) \cdot 100 = (1{,}086-1) \cdot 100 = 8{,}6\ \%$$

Število študentov se je v šestih letih povečevalo povprečno za 8,6 % na leto.

$Y_9 = 325 \cdot K^3 = 325 \cdot 1{,}086^3 = 416{,}27$

Naloga 62 c)

Leto	V_t	K_t	Y_t
2018	/	/	592,86
2019	70	0,7	415
2020	100	1	415
2021	85	0,85	352,75
2022	95	0,95	335,11
2023	110	1,1	368,62

Naloga 63 b)

Leto	t	yi	t·yi	t^2
1	1	100	100	1
2	2	124	248	4
3	3	245	735	9
4	4	300	1200	16
5	5	320	1600	25
Skupaj	15	1089	3883	55

$5a + 15b = 1089 \quad / \cdot (-3)$

$15a + 55b = 3883$

$0 + 10b = 616$

$b = \frac{616}{10} = 61{,}6$

$5a + 15 \cdot 61{,}6 = 1089$

$5a + 924 = 1089$

$5a = 165$

$a = \frac{165}{5} = 33$

$\hat{y} = 33 + 61{,}6 \cdot t$

$\hat{y} = 33 + 61{,}6 \cdot 6$

$\hat{y} = 402{,}6$

Naloga 64 a)

Sistem normalnih enačb za izračun funkcije trenda (v obrazcih št. 3.1):

$5a + 15b = 24.700$

$15a + 55b = 75.200$

Funkcija trenda: $\hat{Y} = 4610 + 110t$

Napoved za število nočitev za 6. leto: $\hat{Y}_t = 4610 + 110 \cdot 6 = 5270$

Naloga 64 b)

Leto	I–IV	V–VIII	IX–XII	SKUPAJ
1	1.200	2.500	1.000	4.700
2	1.100	2.600	1.200	4.900
3	1.000	2.800	1.100	4.900
4	1.300	2.500	1.200	5.000
5	1.200	2.700	1.300	5.200
SKUPAJ	5.800	1.3100	5.800	24.700

Povprečno štirimesečno število nočitev v obdobju petih let: 24.700/3 = 8.233,3 nočitev.

$$\mathrm{SI1} = \frac{5800}{8233{,}3} \cdot 100 = 70{,}4\ \%$$
$$SI2 = \frac{13100}{8233{,}3} \cdot 100 = 159{,}2\ \%$$
$$\mathrm{SI3} = \frac{5800}{8233{,}3} \cdot 100 = 70{,}4\ \%$$

Povprečno štirimesečno število nočitev v 6. letu: 5270/3 = 1.756,6.

Napoved po štirimesečjih za 6. leto:

$$\hat{Y}_{I-IV;6.leto} = 0{,}704 \cdot 1.756{,}6 = 1.236{,}7$$
$$\hat{Y}_{V-VIII;6.leto} = 1{,}592 \cdot 1.756{,}6 = 2.796{,}6$$
$$\hat{Y}_{IX-XII;6.leto} = 0{,}704 \cdot 1.756{,}6 = 1.236{,}7$$

Naloga 65 a)

Ocena stroškov za prihodnje leto, v d.e.

$$\hat{Y}_5 = 58{,}5 + 2{,}1 \cdot 5 = 69$$

Naloga 65 c)

Povprečni četrtletni stroški za 5. leto: 69/4 = 17,25 d.e.

Ocena stroškov za 4. četrtletje, 5. leto, v d.e.:

$\hat{Y}_{4,\text{četrtletje},5.\text{leto}} = 2{,}255 \cdot 17{,}25 = 38{,}90$

Naloga 66 a)

Uporabimo enačbo (v obrazcih št. 1.7):

$$I_{i/o} = 100 \times \frac{Y_i}{Y_o} \qquad za\ i = 1,2,\cdots, N$$

$$\text{I}_{\ 2016/2022} = \frac{Y\ 2016}{Y\ 2022} \cdot 100 = \frac{340}{390} \cdot 100 = 87{,}2\ \%$$

Naloga 66 b)

Uporabimo enačbe:

Verižni indeksi (v obrazcih št. 1.8)

$$V_1 = - \ in\ V_t = 100 \times \frac{Y_t}{Y_{t-1}} \qquad za\ t = 2,3,\ \cdots,\ T$$

Koeficienti dinamike (v obrazcih št. 1.10)

$$K_1 = \text{--}\ in\ K_t = \frac{Y_t}{Y_{t-1}} \qquad za\ t = 2,3,\ \cdots,\ T$$

Stopnje rasti (v obrazcih št. 1.11)

$$S_1 = - S_t = V_t - 100 \qquad za\ t = 2,3,\ \cdots,\ T$$

Leto	Y_t	K_t	V_t	S_t
2016	340	/	/	/
2017	310	0,912	91,2	-8,8
2018	290	0,935	93,5	-6,5
2019	270	0,931	93,1	-6,9
2020	350	1,296	129,6	+29,6
2021	370	1,057	105,7	+5,7
2022	390	1,054	105,4	+5,4

Naloga 66 c)

Funkcija trenda: $\hat{Y} = 284{,}2 + 11{,}8t$

Napoved za vpis študentov za leto 2023:

$\hat{Y}_t = 284{,}2 + 11{,}8 \cdot 8 = 378{,}6$

Naloga 66 d)

Povprečen vpis študentov v obdobju sedmih let = $\frac{2320}{7}$ = 331,4

SI_3: $\frac{290}{331{,}4} \cdot 100 = 87{,}51$ %

SI_7: $\frac{390}{331{,}4} \cdot 100 = 117{,}68$ %

Ploščine H(z) za standardizirano normalno porazdelitev

$z = \frac{y - \bar{y}}{\sigma}$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0.0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4014
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4983	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987									
3,5	0,4997									
4,0	0,4999									

Primer: Za $z = 1{,}96$ iz preglednice odčitamo površino 0,4750.
(Vir: Artenjak, 2000)

Kritične vrednosti za t porazdelitev

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

prostostne stopinje	0,10	0,05	0,025	0,01	0,005
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	2,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
60	1,296	1,671	2,000	2,390	2,660
120	1,289	1,658	1,980	2,358	2,617
	1,282	1,645	1,960	2,326	2,576

(Vir: Artenjak, 2000)

Univerzitetna založba
Univerze v Mariboru

OBRAZCI[1]

1. del: Urejanje, prikazovanje in analiza podatkov

N	število enot v osnovni statistični množici
y	spremenljivka
k	razred
r	število razredov
f_k	število enot v k-tem razredu
y_{min}	najmanjša vrednost spremenljivke
y_{max}	največja vrednost spremenljivke
$y_{k,\min}$	spodnja meja razreda k
$y_{k,\max}$	zgornja meja razreda k
$i_k = y_{k,max} - y_{k,min}$	širina razreda k
$y_k = \frac{y_{k,min} + y_{k,max}}{2}$	sredina razreda k

Gostota frekvence

$$g_k = \frac{f_k}{i_k} \quad za\ k = 1,2, ..., r \tag{1.1}$$

Kumulativni členi frekvenčne porazdelitve

$$F_1 = f_1,\ F_k = F_{k-1} + f_k \qquad za\ k = 2,3,\cdots, r \tag{1.2}$$

[1] Obrazci so povzeti po priloga obrazci iz učbenika Artenjak (2003).

Strukturni odstotek:

$$f\%(y_k) = 100\frac{f(y_k)}{N} = 100 \cdot f^o(y_k) \qquad za\ k = 1,2,\ \cdots,\ r \qquad (1.3)$$

Ločne stopinje:

$$krog\ f^t(y_k) = 3{,}6 \cdot f\%(y_k)\ \ in\ \ f^t(y_k, x_j) = 3{,}6 \cdot f\%(y_k, x_j) \qquad (1.4)$$

$$polkrog\ f^t(y_k) = 1{,}8 \cdot f\%(y_k)\ \ in\ \ f^t(y_k, x_j) = 1{,}8 \cdot f\%(y_k, x_j) \qquad (1.4A)$$

Indeksi s stalno osnovo

$$I_{i/o} = 100 \times \frac{Y_i}{Y_o} \qquad za\ i = 1,2,\ \cdots,\ N \qquad (1.5)$$

Preračunavanje indeksov

$$I_{i/j} = 100 \times \frac{I_{i/o}}{I_{j/o}} \qquad (1.6)$$

Indeksi s stalno osnovo iz časovnih vrst

$$I_{t/o} = 100 \times \frac{Y_t}{Y_o} \qquad za\ t = 1,2,\ \cdots,\ T \qquad (1.7)$$

Verižni indeksi

$$V_1 = -\ in\ V_t = 100 \times \frac{Y_t}{Y_{t-1}} \qquad za\ t = 2,3,\ \cdots,\ T \qquad (1.8)$$

Odnosi med členi v časovni vrsti

$$\frac{Y_t}{Y_{t-1}} = \frac{I_{t/o}}{I_{t-1/o}} = \frac{V_t}{100} \qquad (1.9)$$

Koeficienti dinamike

$$K_1 = -\ in\ K_t = \frac{Y_t}{Y_{t-1}} \qquad za\ \ t = 2,3,\ \cdots,\ T, \qquad (1.10)$$

Stopnje rasti

$$S_1 = -S_t = V_t - 100 \qquad za\ \ t = 2,3,\ \cdots,\ T \qquad (1.11)$$

$$S_1 = -S_t = 100(K_t - 1) \qquad za\ \ t = 2,3,\ \cdots,\ T \qquad (1.12)$$

Povezava med koeficientom dinamike in verižnim indeksom

$$K_t = (V_t / 100);\ V_t = 100K_t \qquad (1.13)$$

Statistični koeficient

$$K = \frac{X}{Y} \times E \tag{1.14}$$

Koeficient obračanja zalog

$$K = \frac{X \cdot E}{\overline{Y} \cdot i}, \tag{1.15}$$

Povprečna vrednost zalog

$$\overline{Y} = \frac{1}{T}\left(Y_1 + Y_2 + \cdots + Y_T\right), \tag{1.16}$$

$$\overline{Y} = \frac{1}{T}\left(\tfrac{1}{2}Y_o + Y_1 + \cdots + Y_{T-1} + \tfrac{1}{2}Y_T\right), \tag{1.17}$$

Kvantili iz nerazvrščenih vrednosti

$$R_i = N \times P_i + 0{,}5 \tag{1.18}$$

$$R_0 \leq R_i < R_1 \tag{1.19}$$

$$y_i = y_0 + \frac{R_i - R_0}{R_1 - R_0} \times (y_1 - y_0) \tag{1.20}$$

Kvantilni rangi iz nerazvrščenih vrednosti

$$y_0 \leq y_i < y_1 \tag{1.21}$$

$$R_i = R_0 + \frac{y_i - y_0}{y_1 - y_0} \times (R_1 - R_0) \tag{1.22}$$

$$P_i = \frac{R_i - 0{,}5}{N}, \tag{1.23}$$

Aritmetična sredina iz nerazvrščenih vrednosti

$$\bar{y} = \frac{1}{N}(y_1 + y_2 + \cdots + y_N) = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{1.24}$$

Aritmetična sredina iz razvrščenih vrednosti

$$\bar{y} = \frac{1}{N}\sum_{k=1}^{r} f_k y_k \ , \tag{1.25}$$

Geometrijska sredina

Povprečni koeficient dinamike

$$K = \sqrt[T-1]{\frac{Y_T}{Y_1}}, \tag{1.26}$$

$$K = \frac{V}{100} = \frac{\sqrt[T-1]{V_2 \times V_3 \times \cdots \times V_T}}{100} \tag{1.27}$$

$$K = \sqrt[T-1]{\frac{I_{T/o}}{I_{1/o}}} \tag{1.28}$$

$$K = \sqrt[T-1]{K_2 \times K_3 \times \cdots \times K_T} \tag{1.29}$$

Povprečna stopnja rasti

$$S = (K-1)100 = 100K - 100 = V - 100 \tag{1.30}$$

Variacijski razmik

$$VR = y_{max} - y_{min}, \tag{1.31}$$

Kvartilni razmik

$$Q = Q_3 - Q_1 \tag{1.32}$$

Decilni razmik

$$D = D_9 - D_1 \tag{1.33}$$

Varianca iz nerazvrščenih vrednosti

$$VAR = \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2 \tag{1.34}$$

ali

$$VAR = \sigma^2 = \frac{1}{N}\sum_{i=1}^{N} y_i^2 - \bar{y}^2, \tag{1.35}$$

Varianca iz razvrščenih vrednosti

$$\sigma^2 = \frac{1}{N}\sum_{k=1}^{r} f_k (y_k - \bar{y})^2 \tag{1.36}$$

$$\sigma^2 = \frac{1}{N}\sum_{k=1}^{r} f_k y_k^2 - \bar{y}^2 \tag{1.37}$$

Standardni odklon

$$SD = \sigma = \sqrt{VAR} = \sqrt{\sigma^2}, \tag{1.38}$$

Koeficient variabilnosti v odstotku

$$KV\% = \frac{\sigma}{\bar{y}} \times 100 \qquad (1.39)$$

Standardizirana spremenljivka

$$z_i = \frac{y_i - \bar{y}}{\sigma} \qquad za\ i = 1,2,\cdots,N \qquad (1.40)$$

Koeficient asimetrije na podlagi modusa

$$KA_{Mo} = \frac{\bar{y} - Mo}{\sigma} \qquad (1.41)$$

Koeficient asimetrije na podlagi mediane

$$KA_{Me} = \frac{3(\bar{y} - Me)}{\sigma} \qquad (1.42)$$

2. del: Korelacija in regresija

Sistem normalnih enačb za linearno funkcijo

$$\mathrm{N} \cdot \mathrm{a} + \left(\sum_{i=1}^{N} x_i\right) \cdot \mathrm{b} = \sum_{i=1}^{N} y_i \qquad (2.1)$$

$$\left(\sum_{i=1}^{N} x_i\right) \cdot \mathrm{a} + \left(\sum_{i=1}^{N} x_i^2\right) \cdot \mathrm{b} = \sum_{i=1}^{N} y_i x_i$$

$$b = \frac{c_{xy}}{\sigma_x^2}, \qquad (2.2)$$

$$b = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} = \frac{\frac{1}{N}\sum_{i=1}^{N} x_i y_i - \bar{x} \cdot \bar{y}}{\frac{1}{N}\sum_{i=1}^{N} x_i^2 - \bar{x}^2} \qquad (2.3)$$

$$a = \bar{y} - b\bar{x} \qquad (2.4)$$

Kovarianca

$$c_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$

ali (2.5)

$$c_{xy} = \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \bar{x} \cdot \bar{y},$$

Determinacijski koeficient

$$r_{xy}^2 = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{2.6}$$

Skupna variance

$$\sigma_y^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2 \tag{2.7}$$

Korelacijski koeficient

$$r_{xy} = \frac{c_{xy}}{\sigma_x \cdot \sigma_y} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 \cdot \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2}} = \tag{2.8}$$

$$= \frac{\frac{1}{N}\sum_{i=1}^{N}x_i y_i - \bar{x}\cdot\bar{y}}{\sqrt{\left(\frac{1}{N}\sum_{i=1}^{N}x_i^2 - \bar{x}^2\right)\left(\frac{1}{N}\sum_{i=1}^{N}y_i^2 - \bar{y}^2\right)}}$$

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} \tag{2.9}$$

Standardna napaka

$$\sigma_{ey} = \sigma_y\sqrt{1 - r_{xy}^2} \tag{2.10}$$

3. del: Časovne vrste

Trendne funkcije

Enačba premice	$\hat{Y} = a + bt$
Parabola druge stopnje	$\hat{Y} = a + bt + ct^2$
Parabola tretje stopnje	$\hat{Y} = a + bt + ct^2 + dt^3$
Eksponentna funkcija	$\hat{Y} = ab^t$
Gompertzova krivulja	$\hat{Y} = Ka^{b^t}$
Pearl-Reedova logistična krivulja	$\hat{Y} = \frac{\hat{Y}_\infty}{1 + e^{a+bt}}$

Sistem normalnih enačb za linearni trend

$$\begin{aligned} Ta + \left(\sum_{t=1}^{T} t\right)b &= \sum_{t=1}^{T} Y_t \\ \left(\sum_{t=1}^{T} t\right)a + \left(\sum_{t=1}^{T} t^2\right)b &= \sum_{t=1}^{T} tY_t, \end{aligned} \tag{3.1}$$

$$b = \frac{\frac{1}{T}\sum_{t=1}^{T}(t-\bar{t})(Y_t-\bar{Y})}{\frac{1}{T}\sum_{t=1}^{T}(t-\bar{t})^2} =$$

$$b = \frac{\frac{1}{T}\sum_{t=1}^{T}tY_t-\bar{t}\cdot\bar{Y}}{\frac{1}{T}\sum_{t=1}^{T}t^2-\bar{t}^2} \tag{3.2}$$

$$a = \bar{Y} - b\bar{t} \tag{3.3}$$

Sezonski indeksi

$$SI_p = \frac{100\cdot r}{\sum_{t=1}^{T}\sum_{p=1}^{r}Y_{tp}}\sum_{t=1}^{T}Y_{tp} \qquad p = 1,2,\cdots,r \tag{3.4}$$

4. del: Osnovni pojmi statističnega sklepanja

Dvostransko ocenjevanje aritmetične sredine

$$\bar{Y} - z\cdot SE_{\bar{y}} < \bar{y} < \bar{Y} + z\cdot SE_{\bar{y}}, \tag{4.1}$$

$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n}y_i$ *vzorčna aritmetična sredina iz nerazvrščenih vrednosti* (4.2)

$\bar{Y} = \frac{1}{n}\sum_{k=1}^{r}f_k y_k$ *vzorčna aritmetična sredina iz razvrščenih vrednosti* (4.3)

$SE_{\bar{y}} = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$ *standardna napaka ocene aritmetične sredin* (4.4)

$$z = 1{,}645 \Rightarrow P(\bar{Y} - 1{,}645\cdot SE_{\bar{y}} < \bar{y} < \bar{Y} + 1{,}645\cdot SE_{\bar{y}}) = 0{,}90, \alpha = 0{,}10 \tag{4.5}$$

$$z = 1{,}96 \Rightarrow P(\bar{Y} - 1{,}96\cdot SE_{\bar{y}} < \bar{y} < \bar{Y} + 1{,}96\cdot SE_{\bar{y}}) = 0{,}95,\ \alpha = 0{,}05 \tag{4.6}$$

$$z = 2{,}58 \Rightarrow P(\bar{Y} - 2{,}58\cdot SE_{\bar{y}} < \bar{y} < \bar{Y} + 2{,}58\cdot SE_{\bar{y}}) = 0{,}99,\ \alpha = 0{,}01 \tag{4.7}$$

Ocenjena variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{Y})^2 \tag{4.8}$$

$$s^2 = \frac{1}{n-1}\sum_{k=1}^{r}f_k(y_k - \bar{Y})^2, \tag{4.9}$$

Dvostransko ocenjevanje totala

$$\hat{Y} - z \cdot SE_Y < Y < \hat{Y} + z \cdot SE_Y \qquad (4.10)$$

$$N \cdot \overline{Y} - z \cdot N \cdot SE_{\bar{y}} < Y < N \cdot \overline{Y} + z \cdot N \cdot SE_{\bar{y}} \qquad (4.11)$$

Dvostransko ocenjevanje strukturnega odstotka

$$p - z \cdot SE_\pi < \pi < p + z \cdot SE_\pi , \qquad (4.12)$$

$$p = 100\frac{n_a}{n} \quad \textit{strukturni odstotek iz vzorca} \qquad (4.13)$$

$$SE_\pi = \sqrt{\frac{p(100-p)}{n}} \quad \textit{standardna napaka ocene strukturnega odstotka} \qquad (4.14)$$

Kritične vrednosti za spremenljivko z

Stopnja tveganja α	Enostransko ocenjevanje	Dvostransko ocenjevanje
10 %	1,28 (ali –1,28)	± 1,645
5 %	1,645 (ali –1,645)	± 1,96
1 %	2,33 (ali –2,33)	± 2,58

Enostransko ocenjevanje statističnih parametrov

$$\bar{y} > \overline{Y} - z\frac{\sigma}{\sqrt{n}} \qquad (4.15)$$

$$Y > \hat{Y} - z \cdot N \cdot \frac{\sigma}{\sqrt{n}} \qquad (4.16)$$

$$\pi > p - z \cdot \sqrt{\frac{p(100-p)}{n}} \qquad (4.17)$$

$$\bar{y} < \overline{Y} + z\frac{\sigma}{\sqrt{n}} \qquad (4.18)$$

$$Y < \hat{Y} + z \cdot N \cdot \frac{\sigma}{\sqrt{n}} \qquad (4.19)$$

$$\pi < p + z \cdot \sqrt{\frac{p(100-p)}{n}} \qquad (4.20)$$

Dvostransko ocenjevanje aritmetične sredine iz malih vzorcev

$$\overline{Y} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \bar{y} < \overline{Y} + t_{\alpha/2}\frac{s}{\sqrt{n}}, \qquad (4.21)$$

Preizkušanje domneve o aritmetični sredini

$$z = \frac{\overline{Y} - \bar{y}_D}{SE_{\bar{y}}} \qquad (4.22)$$

$$SE_{\bar{y}} = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}} \qquad (4.23)$$

$$t = \frac{\overline{Y} - \overline{y}_D}{SE_{\overline{y}}} \tag{4.24}$$

Kritične vrednosti standardizirane spremenljivke

$z_s = -1{,}645 < z < z_z = +1{,}645 \quad \Rightarrow \quad \alpha = 0{,}10$

$z_s = -1{,}96 \quad < z < z_z = +1{,}96 \quad \Rightarrow \quad \alpha = 0{,}05$

$z_s = -2{,}58 \quad < z < z_z = +2{,}58 \quad \Rightarrow \quad \alpha = 0{,}01$

Kritične vrednosti spremenljivke

$\overline{y} \; = \overline{y}_D \pm z \cdot SE_{\overline{y}}$

LITERATURA

Aickin, M. (2010). Variance and Covariance, Reliability and Regression: A Brief Companion of Formulas and Methods for Data Analysis. ZDA: CreateSpace Independent Publishing Platform.

Artenjak, J. (2003). Poslovna statistika. Prenovljena in dopolnjena izdaja. Maribor: UM Ekonomsko-poslovna fakulteta.

Balakrishnan, N., Nevzorov V. B. (2003). A Primer on Statistical Distributions. ZDA: Wiley-Interscience

Brockwell, P. J., Davis, R. A. (2016). Introduction to Time Series and Forecasting. Springer Nature.

Evans, M., Hastings, N., Peacock, B., Forbes, C. (2010). Statistical Distributions. ZDA: Wiley

Freedman, D., Pisani, R., Purves, R. (2007). Statistics. New York: W.W. Norton & Company

Ghauri, P., Grønhaug, K., Strange, R. (2020). Research Methods in Business Studies. United Kingdom: University Printing House.

Heumann, C., Schomaker, M., Shalabh, S. (2016). Introduction to Statistics and Data Analysis. Singapore: Springer

Holmes, A., Illowsky, B., Dean, S. (2018). Introductory Business Statistics. Houston: Rice University.

Lind, D. A, Marchal, W. G., Wathen, S. A. (2021). Basic Statistics in Business and Economics. New York: McGraw-Hill Education

Mišić, E. (2019). Indeksna števila. RS: Statistični urad.

Montgomery, D. C., Peck, E. A., Vining, G. G. (2021). Introduction to Linear Regression Analysis. ZDA: Wiley

Moore, D. S., McCabe, G. P., Craig, B. A. (2016). Introduction to the Practice of Statistics. USA: W. H. Freeman

Ralph, J., O'Neill, R., Winton, J. (2015). A Practical Introduction to Index Numbers. ZDA: Wiley

Seber, G. A. F., Lee, A. J. (2003). Linear Regression Analysis. ZDA: Wiley

Selvamuthu, D., Das, D. (2018). Introduction to Statistical Methods, Design of Experiments and Statistical Quality Control. Singapore: Springer

Tabachnick, B. G., Fidell, L. S. (2013). Using multivariate statistics. Boston: Pearson.

Tominc, P. (2016). Statistika (2. del predmeta). Učno gradivo pri predmetu Statistika (2. del predmeta), interno gradivo. Maribor: EPF.

Tominc, P., Kramberger, T. (2007). Statistične metode v logistiki. Celje: UM Fakulteta za logistiko.

POSLOVNA STATISTIKA:

POLONA TOMIC, MAJA ROŽMAN
Univerza v Mariboru, Ekonomsko-poslovna fakulteta, Maribor, Slovenija
polona.tominc@um.si, maja.rozman@um.si

Predmet Poslovna statistika, za katerega je namenjena ta zbirka vaj, je pomemben za to, da študenti razvijejo sposobnost razumevanja informacij v podatkih. Gradivo obravnava naslednje vsebinske sklope: prikazovanje podatkov v tabelah in grafih, relativna števila, mere centralne tendence, mere variabilnosti, asimetrije in sploščenosti, intervalno ocenjevanje vrednosti statističnih parametrov in osnove preizkušanja domnev o statističnih parametrih, osnove enostavne regresije ter osnove analize in napovedovanja vrednosti v časovnih vrstah.

Ključne besede:
urejanje in prikazovanje podatkov, deskriptivna statistika, enostavna regresijska analiza, osnove vzorčenja, časovne vrste

Univerzitetna založba Univerze v Mariboru

DOI https://doi.org/10.18690/um.epf.6.2023

ISBN 978-961-286-741-6

BUSINESS STATISTICS:

POLONA TOMIC, MAJA ROŽMAN
University of Maribor, Faculty of Economics and Business, Maribor, Slovenia
polona.tominc@um.si, maja.rozman@um.si

Keywords:
data visualisation,
descriptive statistics,
simple regression
analysis,
basics of sampling,
time series

The course Business Statistics, for which this collection of exercises/tutorials is intended, is important for students to develop the ability to understand the information in data. These tutorials contain the following chapters: data visualisation using tables and graphs, relative numbers, measures of central tendency, measures of dispersion, skewness and kurtosis, confidence intervals for statistical parameters and the basis of hypothesis testing of statistical parameters, basics of simple regression and basics of time series analysis and forecasting.

https://doi.org/10.18690/um.epf.6.2023 DOI

978-961-286-741-6 ISBN

University of Maribor Press

University of Maribor

Faculty of Economics and Business