Suzi *

Opisna statistika

1

~ Strukturna števila

DEL S CELOTO

- strukturni delež: $p^0 = \frac{Y_j}{Y}$ ~ delni agregat / ~ agregat $\quad p^0 = \frac{N_j}{N}$ ~ št. enot dela pojava / ~ št. vseh enot ~ kot decimalno število
- odstotni delež: $p^{\%} = \frac{Y_j}{Y} \times 100 \quad p^{\%} = \frac{N_j}{N} \times 100$ ~ v %

npr: $p^0_{\frac{ajdovščina}{Slo}} = \frac{Y_{ajdovščina}}{Y_{Slo}}$

- strukturni delež površine občine Ajdovščina v celotni površini Slo.
- odstotni delež št. ... skupine občin v celotnem št. ... št. v Slo.

~ Statistični koeficient

- splošna formula: $K = E \times \frac{Y}{X}$ ~ vrednost primerjanega pojava / ~ vrednost primerjalnega pojava

koeficient, enota, s katero uravnavamo vr. koeficienta

$K = 0{,}000\,000\,012$
$E = 1\,000\,000$
$K = 0{,}012$

- št. živorojenih otrok na 1000 preb: $K_ž = 1000 \times \frac{Y_ž}{X} = \frac{\text{št. živ. otrok}}{\text{št. preb}} \times 1000$ Ob predpostavki da bi bila stopnja rodnosti celo leto 1992 takšna kot je bila januarja tega leta, bi bila 11,41 živorojenih na 1000 preb.
- št. umrlih na 1000 preb: $K_u = 1000 \times \frac{Y_u}{X}$
- naravni prirast: $K_{np} = K_ž - K_u$

~ Krajevni indeksi → primerjave glede na kraj

primerjana vr.

INDEKS S STALNO OSNOVO ~ 1993: podatke za ostala leta delimo s podatkom l. 1993. × 100

- indeksi s stalno osnovo: $I_{j/0} = 100 \times \frac{Y_j}{Y_0}$ ~ mestevalu / ~ začetna vrednost (osnova)

~ Stvarni indeksi → primerjave med stvarnimi enotami

- indeksi s stalno osnovo: $I_{j/0} = 100 \times \frac{Y_j}{Y_0}$

~ Časovni indeksi → primerjave izbranega pojava v času

- indeksi s stalno osnovo: $I_{j/0} = 100 \times \frac{Y_j}{Y_0}$ ~ tekoči mesec / ~ december preteklega leta

Y_i	V_j
625	–
600	96,0

$\frac{600}{625} \times 100$

- verižni indeksi: $V_j = 100 \times \frac{Y_j}{Y_{j-1}}$ ~ tekoči mesec / ~ pretekli mesec ~ podatek iz tekočega leta delimo s podatkom iz preteklega leta × 100. prvo leto = –
- periodični indeksi: $I_{j,p} = 100 \times \frac{Y_{j,p}}{Y_{j,p-1}}$ ~ tekoči mesec / ~ isti mesec preteklega leta
- koeficient dinamike: $K_j = \frac{Y_j}{Y_{j-1}} = \frac{V_j}{100} = 1 + \frac{S_j}{100}$ ($K_{99} = \frac{V_{99}}{100}$) ~ je faktor, s katerim moramo pomnožiti vr. v prejšnjem obdobju, da dobimo vrednost v opazovanem obdobju
- stopnja rasti: $S_j = 100 \times \frac{Y_j - Y_{j-1}}{Y_{j-1}} = V_j - 100 = 100 \times K_j - 100$ → predstavlja relativno spremembo vr. pojava med dvema zaporednima časovnima obdobjema, to je prirast ali upad vrednosti pojava

~ Razlike med relativnimi "števili"

- absolutna razlika: $D_{j/0} = Y_j - Y_0 \quad D = 70\% - 60\% = 10\%$ točk
- relativna razlika: $D_{j/0}\% = 100 \times \frac{Y_j - Y_0}{Y_0} \quad D_{j/0}\% = \frac{60-70}{60} \times 100 = 16{,}7\%$

~ Razvrstitev vrstic in rang

- povprečni rang: $\bar{R} = \frac{R_j + R_{j+1} + \ldots + R_{j+n}}{N}$

~ Kvantilni rang

- kvantilni rang: $P = \frac{R_p - 0{,}5}{N}$ ~ sredina / ~ vsi rangi skupaj
- SPLOŠNI PRIMER
  - ~ poiščemo sosednji vrednosti: $Y_{-1} \leq Y_P \leq Y_0$
  - ~ v ranžirni vrsti preberemo ustrezna ranga R_{-1} in R_0: $R_{-1} \leq R_y \leq R_0$
  - ~ linearna interpolacija: $R_y = R_{-1} + (R_0 - R_{-1}) \frac{(Y_P - Y_{-1})}{(Y_0 - Y_{-1})}$
  - ~ poznamo ...: $P = \frac{R_p - 0{,}5}{N} \rightarrow R_p = P \cdot N + 0{,}5$ ~ 54% občin iz dane statistične množice ima manjše št. prebivalcev kot Bohinj.

npr: $P = \frac{8 - 0{,}5}{14}$

8 = rang Bohinj
14 = vsi rangi skupaj

2

$R=10$, $n=200$ $= \frac{10-0{,}5}{200} = 0{,}05$ ~ pred njo okoli 5% enot

~ Kvantil

- neposreden izračun kvantila: $R_p = P \times N + 0{,}5$ — EXCEL: PERCENTILES (niz, P) (0,54)
- SPLOŠNI PRIMER:
  - ~ izračunamo ustrezni rang: $R_p = P \times N + 0{,}5$ ~ npr. $R_p = 1{,}2$ ↳ $R_{-1} = 1$, $R_0 = 2$
  - ~ poiščemo sosednja ranga R_{-1} in R_0: $R_{-1} \leq R_p \leq R_0$
  - ~ v ranžirni vrsti preberemo ustrezni vr. npr. y_{-1} in y_0: $y_{-1} \leq y_p \leq y_0$
  - ~ linearna interpolacija: $y_p = y_{-1} + (y_0 - y_{-1}) \frac{(R_p - R_{-1})}{(R_0 - R_{-1})}$ ~ med danimi občinami ima 50% občin manj kot ___ preb., 50% pa več. Mediana je ___ preb.

~ Aritmetična sredina

- povprečna vrednost: $Y = Y_1 + Y_2 + \ldots + Y_N = \sum_{i=1}^{N} Y_i$
- aritmetična sredina: $M_Y = \bar{Y} = \frac{Y}{N}$ ~ vsota od 1 do N
- –"– zapisano za posamične podatke: $M_Y = \bar{Y} = \frac{\sum_{i=1}^{N} y_i}{N}$

~ Agregatni indeksi

- indeks s stalno osnovo: $I_{j/0} = 100 \times \frac{Y_j}{Y_0}$
- verižni indeks: $V_j = 100 \times \frac{Y_j}{Y_{j-1}}$
- stopnja rasti: $S_j = 100 \times \frac{Y_j - Y_{j-1}}{Y_{j-1}} = V_j - 100 = 100 \times K_j - 100$
- periodični indeks: $I_{j,p} = 100 \times \frac{Y_{j,p}}{Y_{j,p-1}}$ ~ letošnji / ~ lanski — j = mesec, p = leto

~ Agregatni indeksi cen – primerjava cen v različnih obdobjih pri enakih količinah

- agregatni indeks cen: $I_p = 100 \times \frac{\sum_{i=1}^{N} p_{j,i}\, q_i}{\sum_{i=1}^{N} p_0\, q_i}$ → tekoče obdobje; → iste količine; → agregat količin in cen v danem obdobju

~ Agregatni indeks količin – primerjava količin dveh izračunskih obdobij pri enakih cenah

X $I_q = 100 \times \frac{\sum_{i=1}^{N} q_{ji}\, p_i}{\sum_{i=1}^{N} q_{0i}\, p_i}$

~ Geometrijska sredina

- splošna formula: $G_Y = \sqrt[N]{Y_1 \times Y_2 \times \cdots \times Y_N}$ — EXCEL: GEOMEAN (niz)
- povprečni verižni indeks: $G_V = \bar{V} = \sqrt[N]{V_1 \times V_2 \times \cdots \times V_N}$
- povprečni koeficient dinamike: $G_K = \bar{K} = M_K = \sqrt[N]{K_1 \times K_2 \times \cdots \times K_N}$
- povprečna stopnja rasti: $\bar{S} = \bar{V} - 100 = 100 \times \bar{K} - 100$

~ Harmonična sredina

- harmonična vsota: $Y_H = \frac{1}{Y_1} + \frac{1}{Y_2} + \ldots + \frac{1}{Y_N} = \sum_{i=1}^{N} \frac{1}{Y_i}$ ~ obratne vrednosti spremenljivk
- splošna formula: $H_Y = \frac{N}{Y_H} = \frac{\text{št. vseh enot}}{\text{harmonični total}}$
- zapisano za posamične podatke: $H_Y = \frac{N}{\sum_{i=1}^{N} \frac{1}{Y_i}}$ — EXCEL: HARMEAN (niz)

~ Srednje vrednosti določene z lego

EXCEL: MEDIAN (niz) ~ mediana

MODE (niz) ~ modus = pove number katerih je več

~ Razmiki variabilnosti

variacijski razmik predstavlja — Me.

- variacijski razmik: $VR = Y_{max} - Y_{min}$ EXCEL: MAX (niz) - MIN (niz)
- decilni razmik: $DR = D_9 - D_1$ EXCEL: PERCENTILE (niz; 0,9) - PERCENTILE (niz; 0,1)
- kvartilni razmik: $QR = Q_3 - Q_1$ kvartilni razmik predstavlja — mediane
- relativni kvartilni razmik: $QR_{Me} = \frac{Q_3 - Q_1}{Me}$ - EXCEL: $\frac{\text{PERCENTILE (niz; 0,75) - PERCENTILE (niz; 0,25)}}{\text{MEDIAN (niz)}}$

~ Povprečni absolutni odklon

- od aritmetične sredine: $AD_{My} = \frac{\sum_{i=1}^{N} |Y_i - M_Y|}{N}$ predstavlja
- od mediane: $AD_{Me} = \frac{\sum_{i=1}^{N} |Y_i - Me|}{N}$

~ Standardni odklon

- varianca: $\sigma_y^2 = \frac{\sum_{i=1}^{N} (Y_i - M_Y)^2}{N}$ EXCEL: VARP (niz)
- standardni odklon: $\sigma_y = \sqrt{\sigma_y^2}$ ~ korenimo varianco EXCEL: STDEVP (niz)

~ Koeficient variacije je relativni standardni odklon

- koeficient variacije: $KV = \frac{\sigma_Y}{M_Y}$

~ Enako široki razredi

① variacijski razmik: $VR = Y_{max} - Y_{min}$

② izračunamo širino razreda: $d = \frac{VR}{k} = \frac{(Y_{max} - Y_{min})}{k}$

- spodnja meja: $Y_{1,s} = Y_{min}$
- zgornja meja: $Y_{k,z} = Y_{max}$
- preostale meje: $Y_{i,z} = Y_{i,s} + d$; $Y_{i+1,s} = Y_{i,z}$
- sredina razreda: $Y_j = \frac{Y_{j,s} + Y_{j,z}}{2}$

~ Relativno enako široki razredi

① $7 \leq k \leq 14 = k$

② določimo spodnjo in zgornjo mejo stat. množice: Y_{max}, Y_{min}

③ zgornja in spodnja meja razredov (prilagodimo): $Y_{k,z}$, $Y_{1,s}$

④ določimo razmerje: $C = \sqrt[k]{\frac{Y_{k,z}}{Y_{1,s}}}$ k = št. razredov

⑤ izračunamo spodnje in zgornje meje razredov: $Y_{j,z} = Y_{j,s} + d$

EXCEL = C = POWER(max(niz) / min(niz); 1/7)

	$Y_{k,s}$	$Y_{k,z}$
1	150	= ROUND(150 · c; 0)
2	393	...
3	1025	...

~ Frekvenčna analiza

- število enot v populaciji: $N = \sum_{j=1}^{k} f_j$
- relativna frekvenca: $f_j^{\circ} = \frac{f_j}{N}$
- kumulativna frekvenca: $F_j = F_{j-1} + f_j$

f_i	F_n
6	$0+6=6$
29	$6+29=35$

- relativna kumulativna frekvenca: $F_j^{\circ} = F_{j-1}^{\circ} + f_j^{\circ}$

f_j ~ EXCEL: FREQUENCY (celoten niz, zgornje meje frekvenčne porazdelitve)
= zap: zgornje meje
2) označimo polje
3) F2
4) shift + ctrl + enter

!

~ Tehtana aritmetična sredina

- splošna formula: $M_y = \bar{Y} = \frac{Y}{N}$
- total Y za frekvenčne porazdelitve: $Y = \sum_{j=1}^{k} f_j \times Y_j$
- število enot N: $N = \sum_{j=1}^{k} f_j$
- tehtana aritmetična sredina: $M_y = \frac{\sum_{j=1}^{k} f_j \cdot y_j}{\sum_{j=1}^{k} f_j} =$

~ Aritmetična sredina aritmetičnih sredin

- najprej total in nato tehtano ar. sr.: $M_{M_y} = \frac{\sum_{j=1}^{k} f_j \cdot M_{y_j}}{\sum_{j=1}^{k} f_j}$

~ Aritmetična sredina strukture

$$M_{y_j^{\circ}} = \frac{\sum_{j=1}^{k} f_j \cdot y_j^{\circ}}{\sum_{j=1}^{k} \cdot f_j}$$

Suzi

1-STRUKTURA

② št/skupaj

OBČINA	$f_j x_j$	3. naloga (motorna k./skupaj) $f_j \cdot x_j$	Motorna kolesa (x)	Os. avtomobili	Avtobusi	Tovorna vozila	Traktorji	SKUPAJ	VRSTIČNA STRUKTURA	
Ajdovščina	17,62164	0,02078	848	11041	2	1085	1267	14 243	0,062 848/14243	11041/14243
Apače	0,0474416	0,001078	44	756	1	39	45	885	0,06 44/885	756/885
Beltinci	0,4466	0,003308	135	3834	0	262	853	5084	0,05 135/5084	
Benedikt	.	0,002303	94	1151	5	85	167	1502	0,027 94/167	
Bistrica ob S.	.	0,000662	27	707	0	51	229	1014		
Bled ...	...	0,005097 ...	... 208	... 5222	... 0	... 339	... 343	6112 ...		
SKUPAJ	836,47147		40 808	1 051 836	2378	88457	84316 =	1 267 775		
Bohinj			84	2586	0	151	366	3187		

1. Kako je določena struktura?

a) s katero spremenljivko je določena vrstična struktura? Katere vrednosti zavzema? Kaj je naslov vrstic, vrsta motornega vozila, ki zavzema vrednosti motorna kolesa, osebna vozila...

b) s katero spremenljivko je določena stolpčna struktura? Katere vrednosti zavzema? določajo vr. ki so zapisane v 1. stolpcu, v našem primeru so občine, zaseda vr. imena vseh občin

c) kaj predstavlja kotna struktura? Je vsota vseh sredstev, razmerje 2:1 več kot vsak drugi slovenec ima avtomobil

Vrednost, ki se nanaša na posamezni vrednosti, kotne in stolpčne in vrstične strukture, kar pomeni za eno vrednost stolpčne strukture.

Predstavlja vrednost glede na posamezno vr. spremenljivke, ki določa vrstično strukturo glede na izbrano vr. spremenljivke

2. Določite vrstično strukturo za izbrano občino. Izbrana občina je BOHINJ

a) določite vrstično strukturo za občino Bohinj.

b) kaj nam vrstična struktura pomeni v danem primeru? Dobili smo deleže vrst motornih vozil med vsemi danimi motornimi vozili v dani občini, glede na skupno št. vseh motornih koles.

c) katera vozila predstavljajo največji delež vozil v občini, katera najmanjši?

a) delež / celota

Določimo skupno vsoto vseh prevoznih sredstev v občini Bohinj = 3187

Stolpčna struktura: 84/3187 = $\boxed{0,026357}$ 2586/3187 = $\boxed{0,8114214}$ 0/0 = $\boxed{0}$ 151/3187 = $\boxed{0,047379981}$ 366/3187 $\boxed{0,114842}$

b) Vrstična struktura nam predstavlja delež prevoznih sredstev v občini Bohinj

c) Največji delež vozil predstavljajo: OSEBNI AVTOMOBILI, najmanjši: AVTOBUSI

3. Določite stolpčno strukturo za izbrano vrsto vozil

a) določite stolpčno strukturo za motorna kolesa.

b) kaj nam stolpčna struktura pove v danem primeru?

c) katera občina ima največji delež motornih koles v Slo. in katera najmanjši?

a) motorna kolesa / motorna kolesa skupaj $$

CELJE: $\boxed{0,021246}$ motorna kolesa = 867 skupaj = 40808

b) 2% vseh motornih koles v RS se nahaja v Celju → delež, ki predstavlja št. motornih koles v pos. občini

c) Občina z največjim deležem motornih koles v RS. = $\boxed{\text{MARIBOR}}$ = INDEX (A8:A217 ; MATCH(MAX (C8:C217) ; C8:C217 ; 0))

ALI max in min

= INDEX(občine ; MATCH (MAX (vrednosti 3.nal.) ; vr. 3. nal ; 0))

4. Določite aritmetično sredino za strukturo št. motornih koles v slovenskih občinah.

a) za izračun uporabite strukturo stolpca (tehtana aritmetična sredina)

b) kako bi lahko hitreje izračunali povprečno strukturo motornih vozil v slovenskih občinah?

c) kolikšno je število občin z manjšim deležem motornih koles v Slo od aritmetične sredine?

a) f_j = delež motornih koles v posameznih občinah $y = f_j \times x_j$ = 3. naloga × motorna kolesa

x_j = število motornih koles v —||—

b) Vrednost tehtane aritmetične sredine: = 836,47147 / 40808 = SUMPRODUCT (D8:D217 ; C8:C217) / SUM (D8:D217)

= 0,020498 → delež (3. naloga)

c) =COUNTIF (niz ; "<0,020498") → 3. naloga pogoj napišeš s števila! = 205

= SUMPRODUCT (št. mot. kol ; delež mot. kol.) / SUM (št. mot. kol)

* 4. Določite aritmetično sredino za delež motornih koles v Slo. občinah.

a) za izračun uporabite vrstično strukturo, izhajajte iz deleža motornih koles v Slo. občinah

b) po stolpcu izračunajte tehtano aritmetično sredino za delež (delež motornih koles × vsa vozila v občini)

c) kako bi lahko hitreje izračunali povp. strukturo motornih vozil v Slo. občinah?

d) Kolikšno je št. občin z manjšim deležem motornih koles v Slo od aritmetične sredine?

NPR: Občina Ajdovščina ima

5. Krajevni indeksi

a) Določite in komentirajte krajevne indekse za št. motornih koles glede na največje št. motornih koles.

b) —||— —||— za št. osebnih avtomobilov glede na $\boxed{\text{najmanjše}}$ št. os. avtomobilov.

c) —||— —||— za št. avtobusov glede na aritmetično sredino avtobusov.

d) —||— —||— za št. tovornih glede na mediano spremenljivke tovornih vozil.

a) Največji delež motornih koles ima Maribor, največji indeks glede na maribor ima LJ. kar pomeni da predstavlja št. motornih koles 99,1% števila motornih koles v Mariboru

KRAJEVNI INDEKS

MAX: motorna kolesa = 3253

$\boxed{\text{MIN:}}$ osebni avtomobili = 133

3253	133	(5b)	(5c) = AVERAGE (avtobusi) = 11,3238095 (aritmetična sredina BUS)
848 (motorna kolo) / 3253 $$ 44/3253 = 1% 135/3253 = 4% ⋮	11041 (os. avto) / 133 $$ 756/133 = 5,68 3834/133 = 28,83 ⋮	5 x več 28 x več	= tovorna vozila / 11,3238095 $$

2-ARITMETIČNA SREDINA KOEFICIENTOV

OBČINA	ŠT. AVTOMOBILOV	OS. AVTO na 100~~0~~ preb	ŠT. PREBIVALCEV
Ajdovščina	11 004	59,0	18 651
Apače	751	20,7	3629
Beltinci	3827	45,9	8337
Benedikt	1146	48,7	2354
Bistrica ob Sotli	707	50,2	1408
Bled	5207	63,9	8148
...	...	...	...

1- Primerjajte št. avtomobilov po občinah.

a) v kateri občini imajo največ avtomobilov?

v kateri občini imajo največ avto. na 100~~0~~ preb?

b) v kateri občini imajo najmanj avtomobilov?

v -11- -11- -11- na 100~~0~~ preb?

a) K = 100 x št. avtomobilov / št. prebivalcev

število avtomobilov = K / 100 x št. prebivalcev

LJUBLJANA

TREBNJE = INDEKS (OBČINA ; MATCH (MAX(na 100preb) ; na 100 preb ; 0))

b) MIN

2) Izračunajte aritmetično sredino koeficienta št. avtomobilov na 100~~0~~ preb.

a) izračunano vrednost komentirajte

b) kolikšno je št. občin, ki imajo manjše št. avtomobilov na 100~~0~~ preb. od aritmetične sredine?

a) SUMPRODUCT (os. auto na 100 preb ; št. prebivalcev) = 51,42873 11 = AVERAGE (os. auto. na 100 preb) = 48,9

b) = COUNTIF (avt. na 100 preb ; "< 51,4287 311") = COUNTIF (auto na 100 preb ; "<48,9")

KOMENTAR!

3) Naredi kratko analizo variabilnosti v statistični množici slovenskih občin.

a) za koeficient št. avtomobilov na 100 preb. določite variacijski razmik in ga komentirajte.

b) za koeficient št. avtomobilov na 100 preb. določite kvartilni razmik in ga komentirajte.

c) za koeficient št. avtomobilov na 100 preb. določite decilni razmik in ga komentirajte.

d) za št. avtomobilov v slovenskih občinah določite standardni odklon in ga komentirajte.

e) Določite koeficient variacije in ga komentirajte.

a) VARIACIJSKI RAZMIK: = MAX (os. avt. na 100 preb) - MIN (os avt. na 100 preb) = 53,0

Razlika med max in minimalno vrednostjo 53%

Celoten variacijski razmik v opazovani statistični množici znaša 53.

b) KVANTILNI RAZMIK: = PERCENTILE (os. auto na 100 preb ; 0,75) - PERCENTILE (os. auto. na 100 preb ; 0,25) = 5,5

Kvartilni razmik znaša 5,5. 50% občin se nahaja na intervalu z dolžino 5,5.

c) DECILNI RAZMIK: = PERCENTILE (os. auto na 100 preb ; 0,9) - PERCENTILE (os. auto. na 100 preb ; 0,1) = 14,4

Decilni razmik znaša 14,4. 80% občin (0,9-0,1) ima koeficient št. osebnih avtom. med 57,14 in 42,7

0,9 = PERCENTILE (auto na 100 preb ; 0,9) = 57,14

0,1 = PERCENTILE (auto na 100 preb ; 0,1) = 42,7

Suri

3-KVANTILI, SV, M (srednje vrednosti, mere variabilnosti)

	tovorna mot. vozila	traktorji	(3.A) RANK tovorna vozila	traktorji	(3B) PERCENTRANK traktorji
SKUPAJ	88 437	84 316			
Ajdovščina	1085	1267	=rank(1085;tov) F4	=rank(1267;traktorji) $F4$	=PERCENTRANK(traktorji; 1267) B
Apače	39	45	=rank(39;tov) F4		
Beltinci	262	853	⋮		
Benedikt	85	167			

1. Opazovanje statistične množice
a) koliko statističnih enot imamo v opazovani statistični množici? 210 (tovorna vozila, traktorji)
b) katera občina ima največ tovornih vozil, katera najmanj?. MAX, MIN ali INDEX
c) katera občina ima največ traktorjev, katera najmanj?

2. Za vsako od spremenljivk izračunajte zahtevane kvantile (namig: uporabite funkcijo PERCENTILE)

	TOVORNA VOZILA	TRAKTORJI
a) PRVI DECIL (D1)	=PERCENTILE(tov;0,1)=35,9	69,4
PRVI KVARTIL (Q1)	74	147,25
MEDIANA (M)	177	293,5
TRETJI KVARTIL (Q3)	417	485
DEVETI DECIL (D9)	836,4	853,6

D1 = 10% občin v stat. množici je takih, ki imajo manj kot 35,9 tovornih vozil
Q1 = pomeni da ima 74% občin večje število od 74 ali 25% občin ima št. tovornih vozil nižje od 74.

3. Izračun ranga in kvantilnega ranga
a) za vse občine določite rang (funkcija RANK) = Rank občine Bled je 68.
b) za vse občine določite kvantilni rang (funkcija PERCENTRANK)
c) kolikšen je delež občin, kjer je manj kot 500 traktorjev? → = COUNTIF(traktorji;"<500")/210 ← DELEŽ! = 0,752381
d) kolikšno je število občin, kjer je manj kot 500 traktorjev? = COUNTIF(traktorji;"<500") = 158
e) kolikšen je delež občin, kjer je več kot 500 traktorjev. = COUNTIF(traktorji;">500")/210 → DELEŽ!
f) -//- število -//- manj ... = COUNTIF(traktorji;">500")
g) določite aritmetični sredini za število tovornih vozil in trakt.
h) kolikšen je delež občin, kjer je št. tovornih vozil manjše od aritmetične sredine? = COUNTIF(tov;"<421,1286")/210
i) -//- delež -//- traktorjev večje -//- ? = COUNTIF(traktorji;">401,5048")/210
j) koliko je št. občin, kjer je št. traktorjev večje od aritmetične sredine?. = COUNTIF(traktorji;">401,5048")

a) 67 občin ima št. tovornih vozil večje od občine Bled, ostale pa imajo nižje.
c) Delež občin z manj kot 500 prebi je 75,24%
g) ARITMETIČNA SREDINA

tovorna vozila	traktorji
=AVERAGE(tov) = 421,1286	=AVERAGE(traktorji) = 401,5048

4. Nekaj vprašanj v zvezi z izračunom kvantilov
a) katero je tisto število tovornih vozil od katerega ima polovica občin manjše št. tovornih vozil?
b) kako imenujemo ta kvantil? mediana
c) katero je tisto število traktorjev od katerega ima četrtina občin večje število traktorjev?
d) kako imenujemo ta kvantil? TRETJI kvartil
e) katero je tisto št. tovornih vozil od katerega ima desetina občin večje št. tovornih vozil?
f) kako imenujemo ta kvantil?, deveti decil

5. Izračun mer variabilnosti. Za število tovornih vozil določite zahtevane parametre in komentirajte:
a) VARIACIJSKI RAZMIK: = MAX(tov) - MIN(tov) = 14582 kar pomeni, da je razlika med največjo in najmanjšo občino 14582
b) DECILNI razmik: = PERCENTILE(tov;0,9) - PERCENTILE(tov;0,1) = 800,5
c) KVARTILNI razmik: = PERCENTILE(tov;0,75) - PERCENTILE(tov;0,25) = 346 → 50% občin se nahaja na intervalu 5,5
d) VARIANCA: = VARP(tov) = 1254300,074
e) STANDARDNI odklon: = STDEVP(tov) = 1137,673056 glej 2-ARITMETIČNA SREDINA
f) KOEFICIENT variacije: = standardni odklon / AVERAGE(tov) = 2,70148637

d) kar pomeni da je povp. vr. kvadrata odmika vrednosti spremenljivke aritmetične sredine enaka 1254300,074
e) kar pomeni da se št. tov. povprečno odklanja od aritmetične sredine.
f) povp. odklon je 2x večji od aritmetične sredine. Standardni odklon predstavlja 270% aritmetične sredine.

Q1 = pomeni da ima 74% občin večje število od 74 ali 25% občin ima št. tovornih vozil nižje od 74.

5. Izračun ranga in kvantilnega ranga

✓ a) za vse občine določite rang (funkcija RANK) = Rank občine Bled je 68.

✓ b) za vse občine določite kvantilni rang (funkcija PERCENTRANK)

c) kolikšen je delež občin, kjer je manj kot 500 traktorjev? → = COUNTIF (traktorji ; "<500") / 210 ← DELEŽ! = 0,752381

d) kolikšno je število občin, kjer je manj kot 500 traktorjev? = COUNTIF (traktorji ; "<500") = 158

e) kolikšen je delež občin, kjer je več kot 500 traktorjev. = COUNTIF (traktorji ; ">500") / 210 → DELEŽ!

f) —||— število —||— manj) . . . = COUNTIF (traktorji ; ">500")

g) določite aritmetični sredini za število tovornih vozil in trakt.

h) kolikšen je delež občin, kjer je št. tovornih vozil manjše od aritmetične sredine? = COUNTIF (tov ; "<421,1286") / 210

i) —||—, delež —||— traktorjev večje —||— ? = COUNTIF (traktorji ; ">401,5048") / 210

j) koliko je št. občin, kjer je št. traktorjev večje od aritmetične sredine?. = COUNTIF (traktorji ; ">401,5048")

a) 67 občin ima št. tovornih vozil večje od občine Bled, ostale pa imajo nižje.

c) Delež občin z manj kot 500 preb je 75,24%

g) ARITMETIČNA SREDINA

tovorna vozila: =AVERAGE (tov) = 421,1286

traktorji: =AVERAGE (traktorji) = 401,5048

1. Za izbrane občine prikažite časovno vrsto živorojenih otrok za izbrane občine

		1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
3	BELTINCI	84	99	88	76	83	79	79	74	79	79	86	53	81	89
13	BREZOVICA	94	90	98	113	97	95	37	83	82	96	96	116	109	142
150	SEŽANA	93	107	102	78	86	86	97	84	91	77	104	109	110	143

a) Graf komentiraj, opišite gibanje št. živorojenih otrok v danem obdobju za izbrane občine. LINIJSKI GRAFIKON
b) kdaj je bilo št. živorojenih otrok največje, najmanjše? SKUPAJ leto. MAX in MIN
c) ali med izbranimi občinami opazite kakšne bistvene razlike? dve odstopanji

2. Za izbrane občine določite indekse s stalno osnovo v izbranem letu.

~~1995~~ 1			
84/84 = 100 %	99/84 = 118%	88/84 = 105%	osnova je leto 1995
94/94 = 100 %	90/94 = 96%	98/94 = 104%	
93/93 = 100 %	107/93 = 115%	102/93 = 110%	

3. Za izbrane občine določite verižne indekse.

1995	1996	1997	1998
–	99/84 = 118%	88/99 = 89%	76/88 = 86%
–	90/94 = 96%	98/90 = 109%	113/98 = 115%
–	107/93 = 115%	102/107 = 95%	78/102 = 76%

→ AVERAGE

NETO PLAČA v 1000 SIT	Št. del. (f_i)	Sredina razr. (y_i)	$(y_i \cdot f_i)$	$y_i - \bar{Y}$	$f_i \times \lvert y_i - \bar{Y} \rvert$	$(y_i - \bar{Y})^2$	$f_i \times (y_i - \bar{Y})^2$
od 20 do pod 30	4	25	100	-40	160	1600	6400
-"- 30 -"- 40	5	35	175	-30	150	900	4500
-"- 40 -"- 50	6	45	270	-20	120	400	2400
-"- 50 -"- 60	9	55	495	-10	90	100	900
-"- 60 -"- 70	16	65	1040	0	0	0	0
-"- 70 -"- 80	28	75	2100	10	280	100	2800
-"- 80 -"- 90	12	85	1020	20	240	400	4800
Σ SKUPAJ	80		5200		1040		21800

Aritmetična sredina

$\bar{Y} = \Sigma(y_i \cdot f_i) / \Sigma f_i = 5200/80 = 65$ tisoč SIT

Če bi vsi delavci v podjetju januarja imeli enako plačo bi ta bila 65 000 SIT

Povprečni absolutni odklon od aritmetične sredine ($AD\bar{Y}$)

$AD\bar{Y} = \Sigma f_i \times |y_i - \bar{Y}| / \Sigma f_i = 1040/80 = 13000$ SIT

Plača 80 zaposlenih delavcev v podjetju se je v januarju 2000 odklanjala od aritmetične sr. v povp. 13 000 SIT navzgor in navzdol

Varianca in standardni odklon

$\sigma_v^2 = \Sigma f_i \times (y_i - \bar{Y})^2 / \Sigma f_i = 21800/80 = 272,5$

V primeru da imamo podatke razporejene v razrede, dobimo le približek $= \sigma_{v\,pop}^2 = \sigma_v^2 - d_0^2/12 = 272,5 - 10^2 = 264,2$

Standardni odklon

$\sigma_v = \sqrt{\sigma_v^2} = \sqrt{264,2} = 16,3$ tisoč SIT

Standardni odklon za podatke o plačah 80 zaposlenih delavcev v podjetju je bil znan 16,3 tisoč SIT

Relativni variacijski razmik (VR)

$VR = Y_{max} - Y_{min} / Me \times 100$ variacijski razmik predstavlja — Me.

Absolutni povp. odklon predstavlja — aritmetične sredine
od sredine

Standardni odklon predstavlja — ar. sr.

Decilni razmik predstavlja — Me.

Mere ~~asimetrije~~ sploščenosti

$KS = 1,9 \times (Q_3 - Q_1) / (D_9 - D_1) = 1,9 \times (21,2 / 44,8) = 0,90 \rightarrow KS < 1$

Koeficient sploščenosti je manjši od 1, zato lahko postavimo trditev, da je bila oblika porazdelitve za plače KONIČASTA

$KS = 1 \rightarrow$ normalna $KS > 1 \rightarrow$ sploščena $KS < 1 \rightarrow$ koničasta

Mere asimetrije

$KA_{Me} = 3 \times (\bar{Y} - Me) / \sigma_v = 3 \times (65 - 70,2) / 16,3 = -1$

$KA_Q = Q_1 + Q_3 - 2 \times Me / Q_3 + Q_1 = -0,3$

Oba koeficienta as. sta negativna, oblika asimetrična v levo.

Aritmetična sredina

$\bar{Y} = \Sigma(y_i \cdot f_i) / \Sigma f_i = 5200/80 = 65$ tisoč SIT

Če bi vsi delavci v podjetju januarja imeli enako plačo bi ta bila 65 000 SIT

Povprečni absolutni odklon od aritmetične sredine ($AD\bar{Y}$)

$AD\bar{Y} = \Sigma f_i \times |y_i - \bar{Y}| / \Sigma f_i = 1040/80 = 13000$ SIT

Plača 80 zaposlenih delavcev v podjetju se je v januarju 2000 odklanjala od aritmetične sr. v povp. 13 000 SIT navzgor in navzdol.

Varianca in standardni odklon

$\sigma_v^2 = \Sigma f_i \times (y_i - \bar{Y})^2 / \Sigma f_i = 21800/80 = 272,5$

V primeru da imamo podatke razporejene v razrede, dobimo le približek $= \sigma_{v\,pop}^2 = \sigma_v^2 - d_0^2/12 = 272,5 - 10^2 = 264,2$

Standardni odklon

$\sigma_v = \sqrt{\sigma_v^2} = \sqrt{264,2} = 16,3$ ~~tisoč~~ SIT

Standardni odklon za podatke o plačah 80 zaposlenih delavcev v podjetju je bil znan 16,3 tisoč SIT

Relativni variacijski razmik (VR)

$VR = Y_{max} - Y_{min} / Me \times 100$ variacijski razmik predstavlja — Me.

Absolutni povp. odklon predstavlja — aritmetične sredine
od sredine

Standardni odklon predstavlja — ar. sr.

Decilni razmik predstavlja — Me.

Mere ~~asimetrije~~ sploščenosti

$KS = 1,9 \times (Q_3 - Q_1) / (D_9 - D_1) = 1,9 \times (21,2 / 44,8) = 0,90 \rightarrow KS < 1$

Koeficient sploščenosti je manjši od 1, zato lahko postavimo trditev, da je bila oblika porazdelitve za plače KONIČASTA

$KS = 1 \rightarrow$ normalna $KS > 1 \rightarrow$ sploščena $KS < 1 \rightarrow$ koničasta

Mere asimetrije

$KA_{Me} = 3 \times (\bar{Y} - Me) / \sigma_v = 3 \times (65 - 70,2) / 16,3 = -1$

$KA_Q = Q_1 + Q_3 - 2 \times Me / Q_3 + Q_1 = -0,3$

Oba koeficienta as. sta negativna, oblika asimetrična v levo.

Suzi

2. kolokvij

1. KORELACIJA

	SLOVENIJA	POVRŠINA	STAN SKUPAJ STANOVANJA	ZAP ZAPOSLITVENI STATUS	R_x	R_y	$(R_y - R_x)^2$
1	Ajdovščina	m^2	15.017	8.326	=RANK(15.017;STAN)= = 16	=RANK(8326;ZAP)= = 22	=POWER(16-22;2)=36
2	Apače	m^2	570	1.957	164	116	2304
3	Beltinci	m^2	5.434	4.144	55	58	9
4	Benedikt	m^2	3.617	1.017	72	176	10816
5	Bistrica	m^2	797	630	148	197	2401
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

1) V tabeli predstavljeni pojav grafično predstavite! (razsevni diagram)

2) Opišite graf. Kakšne so značilnosti, posebnosti? Ali je primeren za opazovanje pojava? Večina vrednosti je zbranih blizu izhodišča koordinatnega sistema, ena od vrednosti odstopa. Zaradi velikega odstopanja vrednosti spremenljivk za enoto westna občina LJ graf ni informativen.

3) Kaj bi bilo treba storiti, da bi bil graf bolj informativen? Izločiti občino LJ

5) Opiši graf brez LJ! Točke so v grafu nekoliko razpršene, vendar lahko razberemo pojav, da je v občini z večjimi stanovanjskimi površinami več delovno aktivnega prebivalstva.

6) S pomočjo grafa ocenite povezanost med spremenljivkama in jo komentirajte! Na osnovi grafa lahko ocenimo, da je povezanost med spremenljivkama POZITIVNA in ZMERNA.
V občinah z večjimi stanovanjskimi površinami je več delovno aktivnih prebivalcev.

7) Izračunaj SPEARMANOV koeficient korelacije ranga.

TABELA

= 1-6*SUM($\boxed{(R_y-R_x)^2}$) / (COUNT($\boxed{(R_y-R_x)^2}$) *
F4 F4

* (POWER(COUNT($\boxed{(R_y-R_x)^2}$);2)-1)) = 0,79
F4

VREDNOST KOEFICIENTA	MOČ POVEZANOSTI
0 → ± 20	NI
± 20 → ± 40	ŠIBKA
± 40 → ± 70	ZMERNA
± 70 → ± 1	MOČNA

8) Dobljeni rezultat komentiraj in primerjaj z oprosnim ocenami!
Povezanost med spremenljivkama je pozitivna in MOČNA.

9) Izračunaj PEARSONOV koeficient (korelacijski)
= PEARSON (stan; zap) = 0,88
= CORREL (stan; zap) = 0,88
→ ENAKA!

10) Dobljeni rezultat komentirajte. Primerjajte izračunani koeficient s koeficientom korelacije ranga in oceno povezanosti, ki ste jo naredili s pomočjo grafične predstavitve.
S pomočjo grafa smo ocenili, da je povezanost zmerna, saj so bile vrednosti nekoliko razpršene po grafu.
Spearmanov koeficient je nekoliko večji od Pearsonovega, to pa zato, ker je pri korelacijskem koeficientu upoštevan tudi vpliv velikih vrednosti, pri koeficientu korelacije ranga pa ne.

11) Ali bomo v korelacijski analizi uporabljali agregatni podatek (total) za celotno Slo?
Pogosto najdemo v tabelah tudi agregat (vsoto vrednosti vseh spremenljivk), ki ga pri korelacijskem računanju NE upoštevamo

12) Razmislite o razmerju med obema spr. Od katere spremenljivke sta odvisna?
Obe dani spremenljivki sta odvisni od št. prebivalcev. Zato absolutni vrednosti spremenljivk nista primerni za primerjavo med občinami. V ta namen bi morali uporabiti koeficienta na prebivalca.

13) Kaj lahko opazujemo z danima spremenljivkama in česa ne moremo?
Ne moremo primerjati občin, lahko pa ugotavljamo, katere posamezne enote odstopajo od osnovnega pravila. Katere enote imajo več delovno akt. preb. pri manjši stanovanjski površini

$y_i = \alpha + \beta \times X$

2. REGRESIJA

	SLOVENIJA	POVRŠINA	STANOVANJA (y, STAN)	ZAPOSLITVENI STATUS (x, ZAP)	Y_i (7. naloga)	$(y'-M_y)^2$	$(y-y')^2$
1	Ajdovščina	m^2	15.017	8.326	$= 698{,}7(\alpha) + 1{,}014(\beta) \times 8326$ = 9139	= 22 217 016	= POWER(15017 − 9139;2) = 34 551 475
2	Apače	m^2	570	1957	2683	= POWER(9139 − AVERAGE(STAN);2) = 3 037 628	= 4 462 984
3	Beltinci	m^2	5.434	4.144	4300	= 224 793	= 285 605
4	Benedikt	m^2	3.617	1.017	1730	= 7 767 199	= 3 501 968
5	Bistrica	m^2	797	630	1337	= 9 536 259	= 292 003
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

1) Izberite odvisno in neodvisno spremenljivko (razmislik o vzroku in posledici)

Št. delovno aktivnih vpliva na stanovanjsko površino v občinah

Št. delovno akt ~ neodvisna spremenljivka ~ x

stanovanja ~ odvisna spremenljivka ~ y

2) Ali je narava vzročno posledične povezave naključna ali vsebinska? $y_i = \alpha + \beta x_i + \varepsilon_i$

Št. akt. preb ni edina spr., ki vpliva na stanovanjsko površino

3) Pojav grafično prikažite. Določite regresijsko premico in enačbo regresijske premice.

4) Graf komentiraje. Opiši razmerje med regresijsko premico in vrednostmi statistične množice. S pomočjo grafične predstavitve ocenite ali je regresijska premica primerna za predstavitev pojava. Vsebinsko opredelite pomen položaja enot nad in pod regresijsko premico.

Delež enot, ki so blizu regresijske premice je kar velik. So pa tudi številne tudi enote, ki zelo odstopajo od regresijske premice. Zato regresijska premica ne more dobro predstaviti pojava.

- Enote NAD regresijsko premico predstavljajo občine, kjer je pri manjšem številu del. akt. preb stanovanjska pov. večja
- Statistično množico tako lahko razdelimo v tri razrede občin
  - enote, ki so blizu regresijske premice, kjer pravilo določa enačba premice
  - enote, NAD (POD) regresijsko premico (večja (MANJŠA) stan. površina pri manjšem (VEČJEM) št. delovno akt. preb)

5) Opredelite regresijsko funkcijo z njenimi parametrovma:

- regresijska konstanta (INTERCEPT): α = INTERCEPT(stan; zap) = 698,7
- regresijski koeficient (SLOPE): β = SLOPE(stan; zap) = 1,014

KOMENTAR:

V kolikor se št. del. akt. preb poveča za eno enoto, se stanovanjska površina v povp. poveča za vr. β

Začetna vrednost regresijske premice je α (vrednost y pri vrednosti $x = 0$)

Stanovanjska površina se za vsakega delovno aktivnega prebivalca poveča z približno $1 m^2$. Občina brez delovno aktivnih prebivalcev bi imela približno $700 m^2$ stanovanjskih površin

6) Ali pridobljene inf. zadostujejo za pravilno interpretacijo vzročno posledične povezave? Kaj je vsebinsko izhodišče vzročno posledičnega pojava?

Št. del. akt. preb predstavlja del vzroka, nekaj vpliva ta spremenljivka gotovo ima na stan. površino. V tem kontekstu moramo obravnavati še št. preb, BDP. Da bi lahko bolje opisali celoten pojav, bi morali uporabiti več neodvisnih spr., uporabiti bi morali metode multivariatne analize.

7) Izračunajte vrednost odvisne spremenljivke s pomočjo pridobljene regresijske funkcije.

Uporabiš enačbo premice in vanjo vstavite vrednost x za vsako enoto. Tako narediš celoten stolpec TREND Y_i

8) Določite celotno varianco, nepojasnjeno in pojasnjeno varianco:

- CELOTNA: =VARP(stan) = 38 212 759
- POJASNJENA: =SUM($(y'-M_y)^2$) / COUNT($(y'-M_y)^2$) ALI VARP(y') = 16 776 531

  Pojasnjena v. je aritmetična sredina vsote kvadratov razlik vrednosti na regresijski pr. y' od arit. sredine
- NEPOJASNJENA: =SUM($(y-y')^2$) / COUNT($(y-y')^2$) ALI celotna − pojasnjena = 21 436 228

  Nepojasnjena v. je vsota kvadratov razlik pravih vrednosti od vrednosti na regresijski premici.

9) Komentiraj!

- Celotna varianca predstavlja variabilnost v originalni statistični množici, variabilnost nad danimi vr. spr. y.
- Pojasnjena v. predstavlja variabilnost, ki jo določajo točke na regresijski premici. To je pojav, kakor ga pojasnimo s pomočjo reg. premice
- Nepojasnjena varianca je razlika med celotno in pojasnjeno v., to je variabilnost, ki je regresijska premica ne pojasnjuje, to so vrednosti, ki jih ne moremo pojasniti z zapisano regresijsko funkcijo.

10) Kako dobimo standardno napako? Izračunajte jo s pomočjo nepojasnjene variance in s pomočjo EXCELA!

Standardno napako dobimo kot standardni odklon iz nepojasnjene v. Izračunano nep. in jo $\sqrt{\ }$.

= SQRT (nepojasnjena)[21430228] = 4630

EXCEL: = STEYX (stan; zap) = 4652

11) Rezultata primerjaj in pojasni razliko

Standardna napaka, izračunana s funkcijo STEYX je nekoliko večja kot tista, izračunana iz nepojasnjene v.

Excelova funkcija upošteva formule za izračun med vzorci. VAR?

12) Izračunaj koeficient variacije in ga komentirajte.

KV_{ey} = standardna napaka 4630/ AVERAGE (stan) = 0,48

Koeficient variacije za standardno napako je 0,48 kar pomeni, da predstavlja standardna napaka 48% aritmetične sredine. To pomeni, da regresijska premica pojava ne predstavlja najbolje

13) S pomočjo ustrezne funkcije v EXCELU izračunaj determinacijski koeficient.

Lahko ga dobimo z izpisom ob regresijski premici

= RSQ (stan; zap) = 0,44

Determinacijski koeficient je tudi kvadrat korelacijskega koef., zato ga dobimo tudi s CORREL², PEARSON²

14) KOMENTAR: Determinacijski koef. je 0,44 kar pomeni, da predstavlja za izbrano regresijsko funkcijo določena pojasnjena varianca 44% celotne variabilnosti, ostala variabilnost pa ima druge vzroke.

15) Kaj nam pove standardna napaka in determinacijski koeficient o primernosti reg. prem.

Na osnovi standardne napake in determinacijskega koeficienta lahko ugotovimo, da nam regresijska funkcija le delno pojasnjuje dani pojav.

Odmiki pravih vr. od regresijske premice so relativno veliki, reg. f. pojasnjuje manj kot 50% celotne variabilnosti

16) Ali lahko v danem primeru govorimo o vzročno posledični povezanosti?

O pravi vzročni posledici ne moremo govoriti. Standardna napaka je prevelika, determinacijski koef. pa premajhen.

Na pojav vpliva več spremenljivk, ki bi jih morali vključiti v opazovanje.

Pojav bi morali obravnavati z multivariatno analizo.

17) S pomočjo funkcije LINEST določi parametre reg. analize za dani primer.

Je tabelarična funkcija, ki za primer bivariatne analize (ena neodvisna spr.) vrača rezultate v 2 stolpca in 5 vrstic.

18) Zapišite in komentirajte

- regresijski koeficient
- regresijsko konstanto
- determinacijski koef.
- standardno napako
- vsoto kvadratov odmikov vr. reg. fun. od pravih vrednosti in od aritmetične sredine

=LINEST (stan; zap; 1; 1)

= 1,044

označi

F2

shift + ctrl + enter

β =	1,044	698,7	$= \alpha$
	0,080	436,1	
determinacijski koef.: ρ^2_{xy} =	0,439	4652,3	= σ_{ey}: standardna napaka
	161,2	206	
	3 489 518 346	4 458 735 432	= vsota kvadratov odmikov vr. spr. od vrednosti na reg. premici

σ = vsota kvadratov razmikov odmikov vr. na regresijski funkciji od aritmetične sredine

19) Pojasnjena, nepojasnjena, celotna varianca: na regresijski prem.

- Pojasnjeno varianco dobimo kot aritmetično sredino vsote kvadratov odmikov vr. od aritmetične sredine:

  = 3489 518 346 / COUNT(y')

- Nepojasnjena v.:

  = 4458 735 432 / COUNT (stan)

- Celotna:

  = pojasnjena + nepojasnjena

  ALI

  = VARP (stan)

10) Kako dobimo standardno napako? Izračunajte jo s pomočjo nepojasnjene variance ... ?
in s pomočjo EXCELA!

Standardno napako dobimo kot standardni odklon iz nepojasnjene v. Izračunamo nep. in jo $\sqrt{\ }$.

$= SQRT(\text{nepojasnjena})^{21430228} = 4630$

EXCEL: = STEYX (stan; zap) = 4652

11) Rezultata primerjaj in pojasni razliko

Standardna napaka, izračunana s funkcijo STEYX je nekoliko večja kot tista, izračunana iz nepojasnjene v.
Excelova funkcija upošteva formulo za izračun med vzorci   VAR?

12) Izračunaj koeficient variacije in ga komentiraj.

KV_{Ey} = standardna napaka 4630 / AVERAGE (stan) = 0,48

Koeficient variacije za standardno napako je 0,48 kar pomeni, da predstavlja standardna napaka 48% aritmetične sredine. To pomeni, da regresijska premica pojava ne predstavlja najbolje

13) S pomočjo ustrezne funkcije v EXCELU izračunaj determinacijski koeficient.

Lahko ga dobimo z izpisom ob regresijski premici

= RSQ (stan; zap) = 0,44

Determinacijski koeficient je tudi kvadrat korelacijskega koef, zato ga dobimo tudi s CORREL^2; PEARSON^2

14) KOMENTAR: Determinacijski koef. je 0,44 kar pomeni, da predstavlja z izbrano regresijsko funkcijo določena pojasnjena varianca 44% celotne variabilnosti, ostala variabilnost pa ima druge vzroke.

15) Kaj nam pove standardna napaka in determinacijski koeficient o primernosti reg. prem.

Na osnovi standardne napake in determinacijskega koeficienta lahko ugotovimo, da nam regresijska funkcija le delno pojasnjuje dani pojav.

Odkloni pravih vr. od regresijske premice so relativno veliki, reg. f. pojasnjuje manj kot 50% celotne variabilnosti

16) Ali lahko v danem primeru govorimo o vzročno posledični povezanosti?

O pravi vzročni posledici ne moremo govoriti. Standardna napaka je prevelika, determinacijski koef. pa premajhen.

Na pojav vpliva več spremenljivk, ki jih moramo vključiti v opazovanje.

Pojav bi morali obravnavati z multivariatno analizo.

17) S pomočjo funkcije LINEST določi parametre reg. analize za dani primer.

Je tabelarična funkcija, ki za primer bivariatne analize (ena neodvisna spr.) vrača rezultate v 2 stolpca in 5 vrstic.

18) Zapišite in komentirajte

- regresijski koeficient
- regresijsko konstanto
- determinacijski koef.
- standardno napako
- vsota kvadratov odklonov vr. reg. fun. od pravih vrednosti in od aritmetične sredine

=LINEST (stan; zap; 1; 1)

= 1,044   označiš
F2
shift + ctrl + enter

β =	1,044	698,7 = α
	0,080	436,1
determinacijski koef.: ρ^2_{xy} =	0,439	4652,3 = σ_{Ey}: standardna napaka
	161,2	206
	3489 518 346	4458 735 432 = vsota kvadratov odklonov vr. spr. od vrednosti na reg. premici

σ = vsota kvadratov razlikov odklonov vr. na regresijski funkciji od aritmetične sredine

19) Pojasnjena, nepojasnjena, celotna varianca: na regresijski prem.

- Pojasnjeno varianco dobimo kot aritmetično sredino vsote kvadratov odklonov vr. od aritmetične sredine:

  = 3489 518 346 / COUNT(y')

- Nepojasnjena v.:

  = 4458 735 432 / COUNT (stan)

- Celotna:

  = pojasnjena + nepojasnjena

  ALI

  = VARP (stan)

Suzi

2. kolokvij

3. ČASOVNE VRSTE

prvo vrsto pustimo prazno

ID	DATUM	NFD VREDNOST	SBITOP	3. naloga $Y_t - Y_0$ NFD (in še za SBITOP isto)	5. naloga V_t (isto za SBITOP)	6. naloga Y_t / Y_0	12. naloga Y_i
1	1.10.	53 268,97	1067,76	0,00 = 53268,97 - 53268,97 F4	=ROUND(—/—;2)=/	100%	=TREND(NFD)
2	2.10.	108 192,22	1056,36	= 108192,22-53268,97	203,0%	= 108 192,22 / 53268,97 F4	+ označiš + ctrl + shift + L (čez stolpec da) dobiš še za SBITOP
3	3.10.	16 443,10	1055,84	⋮	15,0%	⋮	
4	⋮	⋮	⋮	⋮	⋮		

$(Y_i - Y_i')^2$

=POWER(53268,97 - Y' ; 2) potenca

1) max

2) min

3) Absolutne razlike glede na začetek (prvi) v nizu.

= $\Delta_{t/0} = Y_t - Y_0$

Absolutna razlika na NFD glede na 1.10 je večinoma pozitivna, največja negativna razlika je □, največja pozitivna Δ

4) Povprečna absolutna razlika v danem obdobju.

= ROUND ((zadnji NFD F4 - prvi NFD F4) / COUNT (NFD) - 1) ; 2) = 16 670,64 če je - se je znižalo!

V povp. se je vrednost delniško investicijskega sklada NFD v danem obdobju povišala za 16 670,64

5) Verižne indekse v danem obdobju

Verižni indeksi NFD se precej spreminjajo, kar odraža velike dnevne spremembe vrednosti NFD.

Največji skok navzdol —, največji skok navzgor —

SBITOP se ne spreminja veliko, kar odraža majhne dnevne spremembe vrednosti.

6) Indeksi s stalno osnovo na dan 1.10.

$\cdot Y_t / Y_0^{F4} \times 100$

Indeksi s stalno osnovo NFD ukazujejo gibanje vrednosti NFD. Vrednost se je v določenem obdobju povečala za več kot 7 x

7) Povprečno stopnjo rasti in povprečni koeficient dinamike → da lahko navzdol povlečeš

- Povp. stopnja rasti: =POWER ($ zadni NFD F4 / $ prvi NFD F4 ; 1 / (zadni datum - prvi datum)) - 1

  = 0,072 → Povp. stopnja rasti v danem obdobju je 7,2%. Vrednost se je povp. povečala za 7,2%

- Povp. koeficient dinamike: = isto kot POVP. STOPNJA RASTI razen -1

  = 1,072 → Povp. koef. dinamike je 1,072. Vrednost se je v danem obdobju povečala za 1,072 krat.

8) Obravnavana pojava grafično prikažite.

a) originalne vrednosti

b) verižne indekse

c) verižne indekse s stalno osnovo

R^2 0,44

9) V grafični prikaz gibanja obeh spr. vrisi funkcijo trenda, enačbo funkcije in determinacijski koef.

10) Komentiraj funkcijo trenda:

Y = 9272x + 10359

Začetna vrednost funkcije za NFD na dano obdobje je 10359. Vrednost se v naslednjem obdobju poveča za 9272 €

Y = 0,78x + 1087 → na SBITOP — — — — 1087. — — — zmanjša za 0,78 €

11) Komentiraj determinacijski koeficient:

Determinacijski koef. NFD je 0,44 kar pomeni, da predstavlja s funkcijo trenda pojasnjena varianca 44% celotne variance, 56% variance pa je posledica drugih vplivov.

12) Ocenite kakovost trenda (linearna metoda) TREND! (niz Y)

a) celotna varianca: NFD ~ =VARP(NFD F4) = 7900 147 341 SBITOP = 591,51

b) pojasnjena v.: NFD ~ =VARP(Y' F4) = 3460 807 432 SBITOP = 24,52

c) nepojasnjena v.: NFD ~ = celotna - pojasnjena = 4439 339 909 SBITOP = 566,99

d) standardna napaka: NFD ~ =SQRT(nepoj) = 66 628 SBITOP = 23,81

e) kakovost trenda: NFD ~ = napaka / AVERAGE(NFD) = 0,57 SBITOP = 0,02

- Kakovost trenda NFD je 0,57, kar pomeni da predstavlja standardna napaka 57% vr. aritmetične sredine.
- Determinacijski koeficient je 0,44 kar pomeni, da s funkcijo trenda pojasnimo 44% variance pojava.
- S funkcijo trenda pojasnimo le del gibanja NFD, nihanja so bila prevelika, da bi lahko funkcija trenda predstavljala kratkoročno predstavitev pojava. Dolgoročno pa bi lahko sledil
- Kakovost trenda SBITOP je 0, kar pomeni da je standardna napaka glede na vrednost spremenljivke zelo majhna. S pomočjo funkcije trenda jim ni moč slediti

15) Metode napovedovanj

		NFD	SBITOP
a) brez spremembe:		0	0
b) zadnja sprememba:	= 30.10 − 29.10 = (st. zadnji, st. predz.)	209096,68	15,88
c) povp. sprememba:	= (30.10 − 29.10) / (ID zadnji − ID prvi) (F4, F4)	16670,64	−0,41
f) metoda novejših podatkov:	= SLOPE(NFD; ID F4)	= 5273	−0,78
d) zadnji verižni indeks:	= zadnji NFD − predz. NFD =	2,076	1,015
e) povp. verižni indeks:	= POWER(zadnji NFD / prvi NFD; 1 / (ID zadnji − ID prvi)) (F4, F4)		
g) ekstrapolacija linearnega trenda:			

TRENUTNO
30.10.2009

MESEC kasneje: 30.11.2009

= 30.11.2009 − 30.10.2009
= 31

IZRAČUN

a) = NFD zadnji (F4) + 0 × (30.11.2009 (F4) − 30.10.2009 (F4))
b) = NFD$ zadnji + zadnja spr. × (30.11.2009 − 30.10.2009)
c) = −||− + povp. spr. × −||−
f) = −||− + ~~▬~~ novejši podatki × −||−
d) = NFD$ zadnji × POWER(2,076; (30.11.09 (F4) − 30.10.09 (F4)))
e) = −||− −||− povp. ver. ind −||−
g) = FORECAST(31 + (ID zadnji − ID prvi); NFD$; ID (F4))

Suzi

4. VZORČENJE

Delodajalec je meril mnenje delojemalcev s pomočjo ankete. Podatki predstavljajo 1 posamezne odgovore na 2 vprašanja

1 - Ali mislite, da trdo in uspešno delo prinaša velike odgovornosti?

2 - Ali menite, da je napredovanje odvisno predvsem od doseženih rezultatov?

ID	ugodnosti	napredovanje
1	3	2
2	4	5
3	3	1
⋮	⋮	⋮
50	2	1

LIKARTOVA LESTVICA

1 - popolnoma se ne strinjam

2 - se ne strinjam

3 - komaj se strinjam

4 - se strinjam

5 - popolnoma se strinjam

1) Tvorite dve frekvenčni porazdelitvi nad slučajnima spr. ugodnosti in napredovanje!

zg. meja	$f_j(U)$		$f_j(N)$
1	= FREQUENCY (U ; zgornje meje) =	6	= FREQUENCY (N ; zg. meje) = 21
2	→ označiš	12	12
3	→ F2	19	10
4	→ shift + ctrl + enter	11	5
5		2	2

12 → 12 anketirancev je na vprašanje 2 odgovorilo da se ne strinja

2) Podatkom frekvenčne porazdelitve določite kumulativne frekvence!

zg. meje	$f_j(U)$	$f_j(N)$	$F_j(U)$	$F_j(N)$
1	6	21	0 + 6 = 6	0 + 21 = 21
2	12	12	6 + 12 = 18	21 + 12 = 33
3	19	10	18 + 19 = 37	33 + 10 = 43
4	11	5	37 + 11 = 48	43 + 5 = 48
5	2	2	48 + 2 = 50	48 + 2 = 50

$F_j = F_{j-1} + f_j$

3) 37 anketirancev je odgovorilo na vpr 1 z odgovori 1, 2 in 3

4) Izračunajte relativne frekvence in relativne kumulativne frekvence.

	$f_j(U)$	$f_j(N)$	$F_j(U)$	$F_j(N)$	$f_j(U)^\circ$	$f_j(N)^\circ$
1	6	21	6	21	= fj(U) / 50 F4 = 0,12	= fj(N) / 50 F4 = 0,42
2	12	12	8	33	0,24	0,24
3	19	10	57	43	0,38	0,2
4	11	5	48	48	0,22	0,1
5	2	2	50	50	0,04	0,04

8

22% anketirancev je na 1. vpr. odgovorilo s trditvijo se strinjam

$F_j(U)^\circ$

= 0 + 0,12 = 0,12

0,12 + 0,24 = 0,36

0,36 + 0,38 = 0,74

0,74 + 0,22 = 0,96

0,96 + 0,04 = 1

$F_j(N)^\circ$

= 0 + 0,42 = 0,42

0,42 + 0,24 = 0,66

0,66 + 0,2 = 0,86

0,86 + 0,1 = 0,96

0,96 + 0,04 = 1

8

VSI

1, 2 → NE

3, 4, 5 → DA

5) 36% anketirancev se ni strinjalo s trditvijo

6) Grafični prikaz

→ stolpični graf (ordinalna spr.) OPAZIMO: napredovanje neodvisno od dela

7) Izračunaj točkovno oceno povp. mnenja dolgoinalcev nad "ugodnosti so velike" in "napredovanje odvisno".

Točkovna ocena povp. mnenja = G

Aritmetična sredina vzorca $\bar{Y} = \frac{\Sigma Y_i}{N}$

Povp. vr. ugodnosti je večja od napredovanja

- ugodnosti so velike: AVERAGE (ugodnosti) = 2,82
- napredovanje odvisno: AVERAGE (napr) = 2,10

8) Določite pozitivni delež mnenj "ugodnosti" (vrednosti 3+)

- ugodnosti: (3,4 in 5) = SUM (19+11+2)/50 = $\boxed{0,64}$

64% anketirancev je na vpr. dalo pozitivno mnenje

9) • napredovanje: (1 in 2) = SUM (6+12)/50 = $\boxed{0,66}$ ALI 0,66

10) Izračunajte in komentirajte intervalne ocene pri stopnji tveganja 0,05!

a) točkovno oceno povp. mnenja (U)

= CONFIDENCE (0,05 ; STDEVP (ugodnosti) ; 50) = 0,289 304

2,82 - 0,289304 = 2,530696 < y < 2,1 - 0,289304 = 3,109304

S 95% gotovostjo trdimo, da se dejansko povp. mnenje dolgoinalcev nad vpr. 1 nahaja med 2,5 in 3,1.

b) delež pozitivnega mnenja (U)

- odklon zaupanja: SQRT (($\boxed{0,64}$ x (1 - 0,64) / 50) = 0,067882
- Z-alpha: 1,96

0,64 - 0,067882 x 1,96 = 0,506951 < p < 0,66 - 0,067882 x 1,96 = 0,773049

S 95% gotovostjo trdimo, da se dejansko delež populacije s pozitivnim mnenjem na vpr. 1 nahaja med 0,5 in 0,77