

Statistika 2 z računalniško analizo podatkov

Bivariatne analize

1



V Statistične analize v SPSS-ju

V.4 Bivariatne analize

Analyze - Descriptive statistics - Crosstabs

Analyze - Correlate - Bivariate

Analyze - Regression - Linear

Analyze - Compare means

2

V Statistične analize v SPSS-ju

V.4 Bivariatne analize

Analyze - Descriptive statistics - Crosstabs

Analyze – Correlate – Bivariate

Analyze – Regression - Linear

Analyze – Compare means

3

V.4 Bivariatne analize = analize dveh spremenljivk

| Koraki bivariatne statistične analize | Procedura SPSS-u |
|--|--|
| Najprej naredimo univariatno analizo za vse spremenljivke | Npr. Analyze – Descriptive statistics - Frequences |
| Nominalni tip para spremenljivk | Analyze – Descriptive statistics - Crosstabs |
| Ordinalni tip para spremenljivk | Analyze – Correlate - Bivariate |
| Intervalni ali razmernostni tip para spremenljivk - povezanost | Analyze – Correlate - Bivariate |
| Intervalni ali razmernostni tip para spremenljivk - odvisnost | Analyze – Regression – Linear |
| Razlika povprečij po skupinah | Analyze – Compare means ... |

Procedura CROSSTABS – za nominalni tip para spremenljivk

Kako?

- V meniju izberemo: *Analyze – Descriptive statistics – Crosstabs*.

Kaj omogoča?

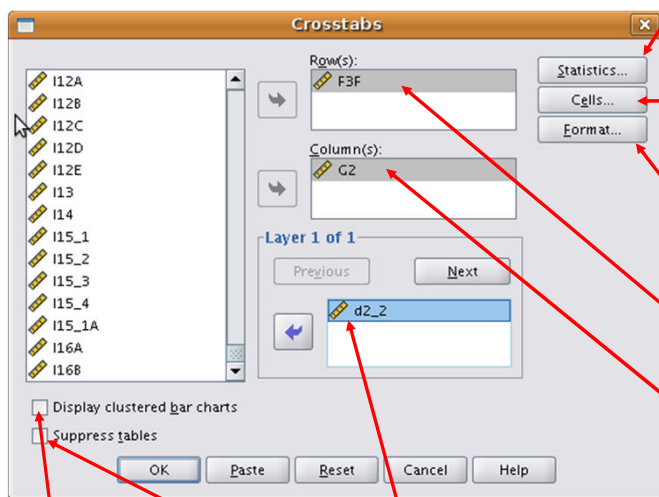
- Izpis kontingenčne tabele glede na 2 ali več spremenljivk (dvorazsežne ali večrazsežne kontingenčne tabele). Položaj spremenljivk v kontingenčni tabeli določamo s prenosom spremenljivk v ustrezna okna (*Row(s)*, *Column(s)*, *Layer*)
- Izpiše empirične frekvence, izračuna pa tudi teoretične frekvence ter strukturne odstotke po vrsticah, stolpcih, glede na celoto ter razlike med empiričnimi in teoretičnimi frekvencami (rezidualne) (gumb *Cells*).
- Izračuna statistike za ugotavljanje povezanosti, npr. χ^2 , kontingenčni koeficient, Cramerjev koeficient (gumb *Statistics*).
- Določamo lahko vrstni red izpisa vrednosti v kontingenčni tabeli (gumb *Format*).
- Izriše strukturne stolpce (označena opcija *Display clustered bar chart*).
- Lahko tudi izpustimo izpis kontingenčne tabele (označena opcija *Suppress tables*).
- Omogoča tudi dodatne izračune, ki jih zaenkrat ne bomo obravnavali.

Kdaj je to uporabno?

- Če nas zanima odnos med dvema ali več (3, 4) nominalnimi spremenljivkami.
- Včasih smiselno tudi za spremenljivke boljše merske lestvice, vendar z manjšim številom vrednosti.

5

Procedura CROSSTABS



Statistike, ki naj se izračunajo

Kaj naj se izpiše v celicah tabele

Vrstni red vrednosti v tabeli

Spremenljivka, ki naj bo v vrsticah

Spremenljivka, ki naj bo v stolpcih

Izpis strukturnih stolpcev

Ne izpiše konting. tabele.

Spremenljivka, ki določa "plasti" – vsaka zahteva (npr. izpis kontingenčne tabele, izračun koeficientov) se izvede za vsako "plast" (vrednost te tretje spremenljivke). Možno je dodati več spremenljivk. ⁶

Procedura CROSSTABS

Kontingenčni koeficient
Cramerjev koeficient

% po stolpcih, vrsticah, skupaj

Empirične frekvence

Teoretične frekvence

χ^2

Reziduali – razlike med empiričnimi in teoretičnimi frekvencami

The image shows two dialog boxes from SPSS. The 'Crosstabs: Cell Display' dialog has 'Observed' checked under 'Counts', and 'Unstandardized' checked under 'Residuals'. The 'Crosstabs: Statistics' dialog has 'Chi-square' checked, and 'Contingency coefficient' and 'Phi and Cramer's V' checked under 'Nominal'. 'Gamma' is checked under 'Ordinal'. 'Eta' is checked under 'Nominal by Interval'. 'Cochran's and Mantel-Haenszel statistics' is also checked. A text box explains that residuals are the differences between empirical and theoretical frequencies.

Ponovitev: χ^2 test za povezanost dveh nominalnih spremenljivk

- χ^2 test uporabljamo za preverjanje domneve (testiranje hipoteze) o povezanosti dveh nominalnih spremenljivk.
- Temelji na razliki med empiričnimi in teoretičnimi frekvencami. Slednje so frekvence, kakršne bi bile v tabeli, če ne bi bilo povezanosti med spremenljivkama. Kadar je razlika med empiričnimi in teoretičnimi frekvencami zadosti velika, zaključimo, da sta spremenljivki povezani.
- Domnevi:
 $H_0: \chi^2 = 0$ – spremenljivki na populaciji **nista** povezani (ni razlik med emp. in teor. frekv.)
 $H_1: \chi^2 > 0$ – spremenljivki **sta** na populaciji povezani (so razlike med emp. in teor. frekv.)
- Če ničelno hipotezo zavrnemo, lahko trdimo, da sta spremenljivki povezani.
- Predpostavka za χ^2 test: χ^2 test je veljaven le, če so vse teoretične frekvence večje od 0, večina pa jih je vsaj 5. In sicer, če je kakšna teoretična frekvenca enaka 0 ali če je teoretičnih frekvenc, ki so manjše od 5, več kot 20%, je predpostavka huje kršena.

Ponovitev: χ^2 test za povezanost dveh nominalnih spremenljivk

- Vrednost χ^2 nam ne poda moči povezanosti. V ta namen izračunamo Cramerjev koeficient in/ali popravljeni Pearsonov koeficient kontingence, ki imata vrednosti na intervalu [0, 1].

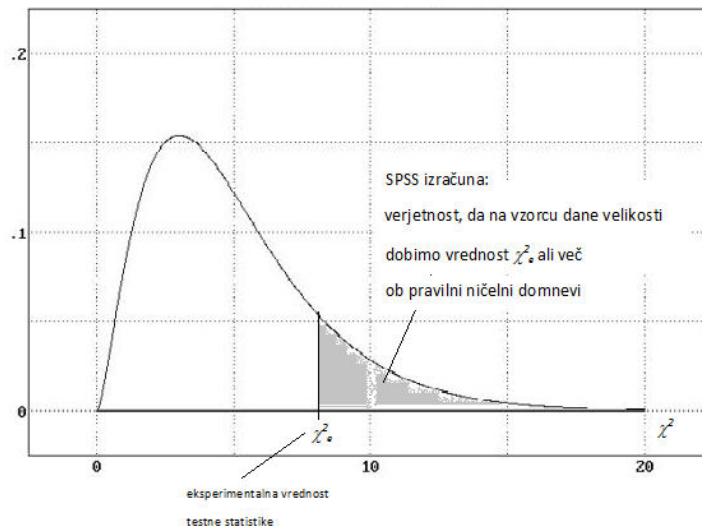
9

Testiranje hipotez z SPSS-om (ali kakim drugim računalniškim programom)

- Pri testiranju hipotez s SPSS-om (oz. računalniškimi programi na splošno) je običajno postopek malce drugačen, kot se ga učite pri Statistiki I.
- SPSS ne preverja, ali je eksperimentalna vrednost statistike (npr. χ_e^2) večja od testne vrednosti statistike pri izbrani stopnji značilnosti α (npr. $\chi_{1-\alpha}^2$), saj običajno stopnje značilnosti sploh ne vnesemo v program (navodil).
- SPSS izračuna eksperimentalno vrednost in ugotovi, kakšna je verjetnost, da bi dobili tako veliko vrednost eksperimentalne statistike, če bi ničelna hipoteza držala.

10

Testiranje hipotez z SPSS-om (ali kakim drugim računalniškim programom)



11

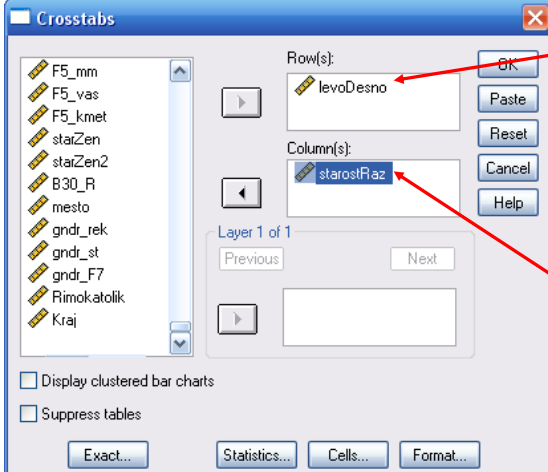
Testiranje hipotez z SPSS-om (ali kakim drugim računalniškim programom)

- Če je ta verjetnost majhna, potem podvomimo v pravilnost ničelne domneve.
- Ta verjetnost je enaka tveganju, da se zmotimo, če zavrnemo ničelno hipotezo.
- To verjetnost imenujemo tudi **natančna stopnja tveganja** in jo označimo s p . Imenujemo jo tudi **statistična značilnost** ali **signifikanca**.
- Ničelno domnevo torej zavrnemo, kadar je p , t.j. stopnja tveganja, da se zmotimo ob zavračanju ničelne domneve majhna oz. manjša od vnaprej določene stopnje značilnosti α (α si sami določimo pred preverjanjem domnev).
- Torej:
 - $p \leq \alpha$ ($p \leq 0.05$): Ničelno domnevo zavrnemo. → Sprejmemo alternativno domnevo.
 - $p > \alpha$ ($p > 0.05$): Ničelne domneve ne zavrnemo.
- Če bi izračunali vrednost statistike pri stopnji značilnosti $\alpha = p$, bi bila ta vrednost natanko enaka eksperimentalni vrednosti.

12

Primer: Ali sta starost in politična opredelitev povezana?

V meniju izberemo: *Analyze – Descriptive statistics – Crosstabs.*

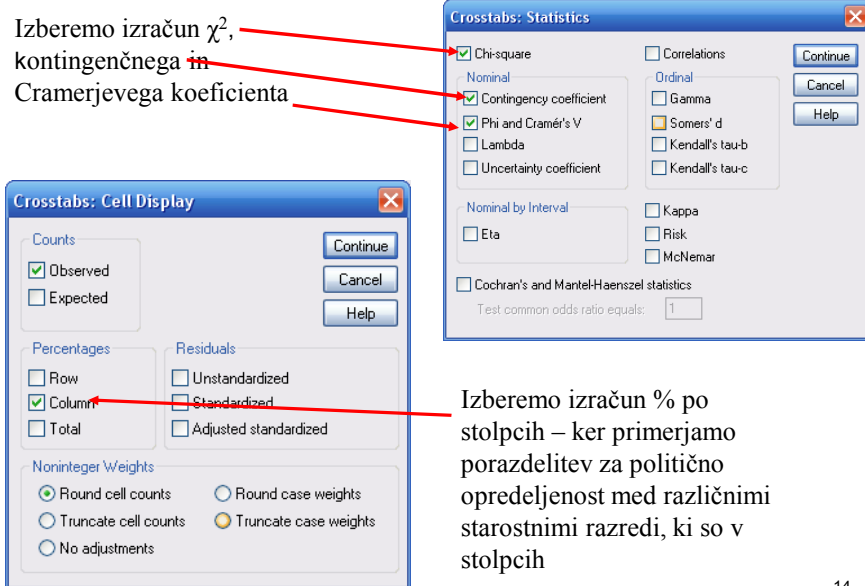


Spremenljivka v vrsticah
(v tem primeru politična opredelitev)

Spremenljivka v stolpcih
(v tem primeru starost, izražena s starostnimi razredi)

13

Izberemo izračun χ^2 , kontingenčnega in Cramerjevega koeficienta



Izberemo izračun % po stolpcih – ker primerjamo porazdelitev za politično opredeljenost med različnimi starostnimi razredi, ki so v stolpcih

14

Primer: Ali sta starost in politična opredelitev povezani?

Rezultat: Empirične frekvence in strukturni odstotki

levoDesno Politična opredelitev (levo, desno) * starostRaz Starost Crosstabulation

| | | starostRaz Starost | | | Total | |
|---|-----------------------------|-----------------------------|----------------|---------|--------|-------|
| | | 1 mladi | 2 srednja leta | 3 stari | | |
| levoDesno Politična opredelitev (levo, desno) | 1 Levo | Count | 28 | 145 | 38 | 211 |
| | | % within starostRaz Starost | 15,7% | 21,6% | 21,7% | 20,6% |
| | 2 Sredinsko | Count | 118 | 390 | 81 | 589 |
| | | % within starostRaz Starost | 66,3% | 58,1% | 46,3% | 57,5% |
| | 3 Desno | Count | 32 | 136 | 56 | 224 |
| | | % within starostRaz Starost | 18,0% | 20,3% | 32,0% | 21,9% |
| Total | Count | 178 | 671 | 75 | 1024 | |
| | % within starostRaz Starost | 100,0% | 100,0% | 100,0% | 100,0% | |

Odstotki se glede na starostni razred razlikujejo – obstaja neka povezanost (vsebinsko lahko govorimo tudi o odvisnosti)

Rezultat: χ^2 test povezanosti (ali na populaciji obstaja povezanost)

Chi-Square Tests

| | Value | df | Asymp. Sig. (2-sided) |
|------------------------------|---------------------|----|-----------------------|
| Pearson Chi-Square | 18,978 ^a | 4 | ,001 |
| Likelihood Ratio | 18,387 | 4 | ,001 |
| Linear-by-Linear Association | 1,313 | 1 | ,252 |
| N of Valid Cases | 1024 | | |

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 36,06.

Vrednost χ^2 statistike je 18,978.

Število prostostnih stopenj (df – degrees of freedom) je 4.

Natančna stopnja značilnosti je manjši od 0,001. Tveganje ob zavrnitvi ničelne domneve je manjše od 0,1%.

Predpostavke χ^2 testa so izpolnjene, saj nima nobena celica teoretične frekvence manj kot 5 (najmanjša teoretična frekvenca je 36,06).

Sklep:

- Ob stopnji značilnosti, manjši od 0,1%, lahko trdimo, da sta starost in politična opredelitev statistično značilnost povezani. (Vsebinsko lahko tudi rečemo, da starost vpliva na politično opredelitev.)
- Stopnja značilnosti manjša od 0,1%: tveganje, da sprejmemo napačen sklep, ko zavrnemo ničelno domnevo, je manjše od 0,1%, torej zelo majhno in jo lahko zavrnemo. Ali drugače, ob pravilni ničelni domnevi, ki pravi, da je $\chi^2 = 0$ (ni povezanosti), je verjetnost, da na vzorcu 1024 enot in prostostnih stopnjah 4 dobimo vrednost χ^2 18,978 ali več, le 0,1%, torej zelo malo verjetna. Zato lahko rečemo, da povezanost obstaja.

Rezultat: Moč povezanosti

Symmetric Measures

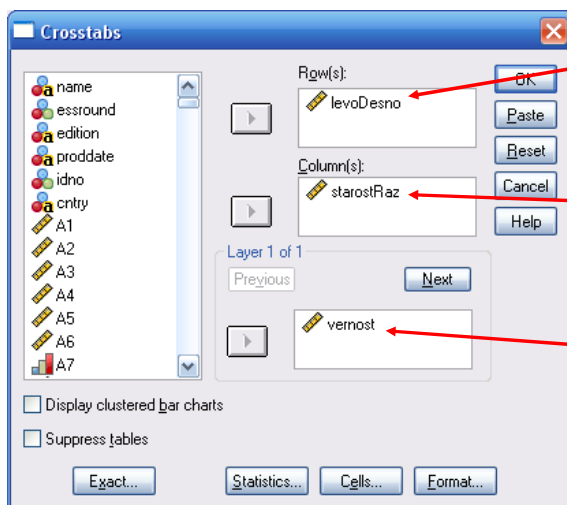
| | | Value | Approx. Sig. |
|-----------------------|-------------------------|-------|--------------|
| Nominal by Nominal | Phi | ,136 | ,001 |
| | Cramer's V | ,096 | ,001 |
| | Contingency Coefficient | ,135 | ,001 |
| N of Valid Cases | | 1024 | |

Cramerjev koeficient in kontingenčni koeficient kažeta na (zelo) šibko povezanost.

- a. Not assuming the null hypothesis.
- b. Using the asymptotic standard error assuming the null hypothesis.

17

Primer: Ali sta starost in politična opredelitev povezana za vse kategorije vernosti?



Spremenljivka v vrsticah (politična opredelitev)

Spremenljivka v stolpcih (starost, izražena v razredih)

Spremenljivka, katere vpliv želimo kontrolirati (izločiti) oz. ki določa skupine, znotraj katerih preučujemo povezanost zgornjih spremenljivk (vernost)

18

Rezultat: Trirazsežna kontingenčna tabela

levoDesno Politična opredelitev (levo, desno) * starostRaz Starost * vernost Kako verni ste Crosstabulation

| vernost Kako verni ste | levoDesno Politična opredelitev (levo, desno) | 1 Levo | Count | starostRaz Starost | | | Total |
|------------------------|---|-----------------------------|--------|-----------------------------|----------------|---------|-------|
| | | | | % within starostRaz Starost | 2 srednja leta | 3 stari | |
| 1,00 nič ali malo | 1 Levo | Count | 17 | 73 | 20 | 110 | |
| | | % within starostRaz Starost | 28,8% | 35,4% | 47,6% | 35,8% | |
| | | 2 Sredinsko | Count | 34 | 110 | 15 | 159 |
| | 2 Sredinsko | Count | 34 | 110 | 15 | 159 | |
| | | % within starostRaz Starost | 57,6% | 53,4% | 35,7% | 51,8% | |
| | | 3 Desno | Count | 8 | 23 | 7 | 38 |
| | 3 Desno | Count | 8 | 23 | 7 | 38 | |
| | | % within starostRaz Starost | 13,6% | 11,2% | 16,7% | 12,4% | |
| | | Total | Count | 59 | 206 | 42 | 307 |
| Total | Count | 59 | 206 | 42 | 307 | | |
| | % within starostRaz Starost | 100,0% | 100,0% | 100,0% | 100,0% | | |
| | 2,00 srednje | Count | 6 | 45 | 15 | 66 | |
| 1 Levo | Count | 6 | 45 | 15 | 66 | | |
| | % within starostRaz Starost | 9,4% | 16,7% | 18,5% | 15,9% | | |
| | 2 Sredinsko | Count | 48 | 172 | 45 | 265 | |
| 2 Sredinsko | Count | 48 | 172 | 45 | 265 | | |
| | % within starostRaz Starost | 75,0% | 63,7% | 55,6% | 63,9% | | |
| | 3 Desno | Count | 10 | 53 | 21 | 84 | |
| 3 Desno | Count | 10 | 53 | 21 | 84 | | |
| | % within starostRaz Starost | 15,6% | 19,6% | 25,9% | 20,2% | | |
| | Total | Count | 64 | 270 | 81 | 415 | |
| Total | Count | 64 | 270 | 81 | 415 | | |
| | % within starostRaz Starost | 100,0% | 100,0% | 100,0% | 100,0% | | |
| | 3,00 zelo | Count | 5 | 26 | 3 | 34 | |
| 1 Levo | Count | 5 | 26 | 3 | 34 | | |
| | % within starostRaz Starost | 9,1% | 13,7% | 5,8% | 11,4% | | |
| | 2 Sredinsko | Count | 36 | 105 | 21 | 162 | |
| 2 Sredinsko | Count | 36 | 105 | 21 | 162 | | |
| | % within starostRaz Starost | 65,5% | 55,3% | 40,4% | 54,5% | | |
| | 3 Desno | Count | 14 | 59 | 28 | 101 | |
| 3 Desno | Count | 14 | 59 | 28 | 101 | | |
| | % within starostRaz Starost | 25,5% | 31,1% | 53,8% | 34,0% | | |
| | Total | Count | 55 | 190 | 52 | 297 | |
| Total | Count | 55 | 190 | 52 | 297 | | |
| | % within starostRaz Starost | 100,0% | 100,0% | 100,0% | 100,0% | | |

Dobimo 3 take tabele kot v prejšnjem primeru, po eno za vsako vrednost spremenljivke *vernost*. Interpretacija posamezne tabele je enaka kot prej.

19

Rezultat: Trije χ^2 testi - za vsako kategorijo tretje spremenljivke

Chi-Square Tests

| vernost Kako verni ste | | Value | df | Asymp. Sig. (2-sided) |
|------------------------------|------------------------------|--------------------|---------------------|-----------------------|
| 1,00 nič ali malo | Pearson Chi-Square | 5,971 ^a | 4 | ,201 |
| | Likelihood Ratio | 6,064 | 4 | ,194 |
| | Linear-by-Linear Association | 1,488 | 1 | ,223 |
| | N of Valid Cases | 307 | | |
| | 2,00 srednje | Pearson Chi-Square | 6,299 ^b | 4 |
| Likelihood Ratio | | 6,505 | 4 | ,164 |
| Linear-by-Linear Association | | ,031 | 1 | ,861 |
| N of Valid Cases | | 415 | | |
| 3,00 zelo | | Pearson Chi-Square | 13,380 ^c | 4 |
| | Likelihood Ratio | 13,012 | 4 | ,011 |
| | Linear-by-Linear Association | 6,419 | 1 | ,011 |
| | N of Valid Cases | 297 | | |

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

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 10,18.

c. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5,95.

Predpostavka v zvezi z velikostjo vzorca je izpolnjena pri vseh treh skupinah – vrednostih spremenljivke *vernost*.

Zopet dobimo 3 tabele, za vsako vrednost spremenljivke *vernost*. Interpretacija je enaka kot v primeru dvorazsežnih kontingenčnih tabel.

Na podlagi rezultatov lahko pri 5% stopnji značilnosti trdimo, da starost vpliva na politično predelitev pri zelo vernih osebah. Za neverne oz. malo verne ter srednje verne pa tega pri 5% stopnji značilnosti ne moremo trditi.

20

Rezultat: Tri-krat izračunani koeficienti za moč povezanosti - za vsako kategorijo tretje spremenljivke

Symmetric Measures

| vernost | Kako verni ste | | Value | Approx. Sig. | |
|---------|----------------|--------------------|-------------------------|--------------|------|
| 1,00 | nič ali malo | Nominal by Nominal | Phi | ,139 | ,201 |
| | | | Cramer's V | ,099 | ,201 |
| | | | Contingency Coefficient | ,138 | ,201 |
| | | N of Valid Cases | 307 | | |
| 2,00 | srednje | Nominal by Nominal | Phi | ,123 | ,178 |
| | | | Cramer's V | ,087 | ,178 |
| | | | Contingency Coefficient | ,122 | ,178 |
| | | N of Valid Cases | 415 | | |
| 3,00 | zelo | Nominal by Nominal | Phi | ,212 | ,010 |
| | | | Cramer's V | ,150 | ,010 |
| | | | Contingency Coefficient | ,208 | ,010 |
| | | N of Valid Cases | 297 | | |

a. Not assuming the null hypothesis.

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

Zopet dobimo tri izračune kontingenčnega in Cramerjevega koeficienta, po enega za vsako vrednost (skupino) spremenljivke *vernost*.

Vidimo, da je povezanost med starostjo in politično opredelitvijo najmočnejša pri zelo vernih (a še vedno šibka), najbolj šibka pa pri srednje vernih.

21

Procedura CORRELATE – za ordinalni in za intervalni/razmernosti tip para spremenljivk

Kako?

- V meniju izberemo: *Analyze – Correlate - Bivariate*.

Kaj omogoča?

- Izpis Pearsonovega koeficienta korelacije za izbrane pare intervalnih ali razmernostnih spremenljivk (opcija *Correlation coefficients – Pearson*).
- Izpis Spearmanovega koeficienta korelacije rangov za izbrane pare ordinalnih spremenljivk (opcija *Correlation coefficients – Spearman*).
- Izračuna statistično značilnost za preverjanje domneve (opcija *Test of significance*) o zgornjih koeficientih za enostranske (*One-tailed*) in dvostranske teste (*Two-tailed*) ter z zvezdicami označi statistično značilne koeficiente (opcija *Flag significant correlations*).
- Izračuna nekatere opisne statistike in kovarianco (gumb *Options*).
- Omogoča še dodatne analize, ki jih zaenkrat ne bomo obravnavali.

22

Procedura CORRELATE – za ordinalni in za intervalni/razmernosti tip para spremenljivk

Kdaj je to uporabno?

- Če nas zanima povezanost za ordinalni ali intervalni/razmernostni tip para spremenljivk.
- Mogoče je izračunati koeficiente korelacije za več parov spremenljivk hkrati (v obliki t. im. korelacijske matrike).

Pozor:

- ✓ 2 ordinalni ali 1 ordinalna in ena boljša spremenljivka – Spearmanov koeficient
- ✓ 2 intervalni ali 2 razmernostni ali 1 intervalna in 1 razmernostna spremenljivka – Pearsonov koeficient

23

Ponovitev: test za linearno povezanost dveh intervalnih/razmernostnih spremenljivk

- Test uporabljamo za preverjanje domneve (testiranje hipoteze) o linearni povezanosti dveh intervalnih/razmernostnih spremenljivk
- Linearna povezanost – v razsevnem grafikonu predstavljene točke, določene s paroma vrednosti obeh spremenljivk, se gostijo okoli premice (t.j. linearne funkcije).
- Mera povezanosti: Pearsonov koeficient korelacije, r (na vzorcu) oz. ρ na populaciji.
- Interval vrednosti koeficienta: $[-1, 1]$.
 - ✓ Vrednost 0: ni linearne povezanosti.
 - ✓ Vrednost -1 : popolna negativna linearna povezanost (z naraščanjem vrednosti ene spremenljivke padajo vrednosti druge spremenljivke).
 - ✓ Vrednost $+1$: popolna pozitivna linearna povezanost (z naraščanjem vrednosti ene spremenljivke naraščajo vrednosti druge spremenljivke).

24

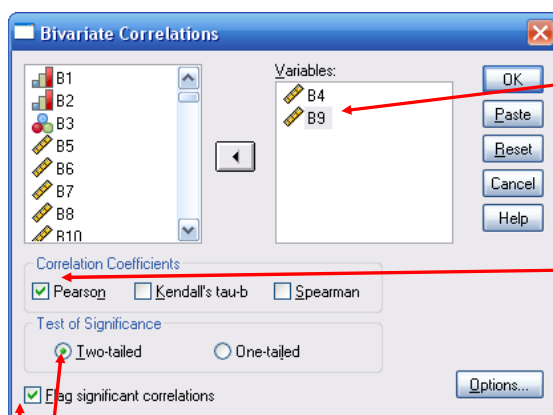
Ponovitev: test za linearno povezanost dveh intervalnih/razmernostnih spremenljivk

- Domnevi:
 - $H_0: \rho = 0$ – spremenljivki **nista** linearno povezani na populaciji
 - $H_1: \rho \neq 0$ – spremenljivki **sta** linearno povezani na populaciji
(Lahko tudi: $H_1: \rho < 0$ oz. $H_1: \rho > 0$, negativna oz. pozitivna linearna povezanost na populaciji.)
- Če ničelno hipotezo zavrnemo, lahko trdimo, da sta spremenljivki (pozitivno, negativno) linearno povezani.

25

Primer: Povezanost med zaupanjem slovenskemu in evropskemu parlamentu

V meniju izberemo: *Analyze – Correlate – Bivariate*.



Spremenljivke, med katerimi računamo korelacijo (povezanost). Če izberemo več spremenljivk, se korelacija izračuna za vse možne pare spremenljivk.

Ker sta obe spremenljivki intervalni in predpostavimo linearno povezanost, izberemo *Pearsonov* koeficient korelacije

izberemo, ali želimo enostranski test (ali vnaprej predpostavljamo smer oz. predznak povezanosti) ali dvostranski test (ne predpostavljamo smeri).

Želimo, da so statistično značilni koeficienti označeni z zvezdicami *.

26

Primer: Povezanost med zaupanjem slovenskemu in evropskemu parlamentu

Rezultat: Korelacijska matrika: Pearsonov koeficient korelacije za vse možne pare spremenljivk

| | | B4 Zaupanje v državni zbor | B9 Zaupanje v evropski parlament |
|----------------------------------|---------------------|----------------------------|----------------------------------|
| B4 Zaupanje v državni zbor | Pearson Correlation | 1 | ,618** |
| | Sig. (2-tailed) | | ,000 |
| | N | 1387 | 1231 |
| B9 Zaupanje v evropski parlament | Pearson Correlation | ,618** | 1 |
| | Sig. (2-tailed) | ,000 | |
| | N | 1231 | 1247 |

Vrednost (izbranega) koeficienta korelacije

Natančna stopnja značilnosti (dvostranski test)

Število enot, na katerih je bila korelacija izračunana

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

Korelacijska matrika je simetrična, zato se vsak izračun ponovi 2x.

Na diagonali je vrednost korelacijskega koeficienta vedno 1, saj je vsaka spremenljivka popolnoma korelirana sama s sabo. Število enot na diagonali predstavlja število veljavnih vrednosti izbrane spremenljivke.

27

Interpretacija

| | | B4 Zaupanje v državni zbor | B9 Zaupanje v evropski parlament |
|----------------------------------|---------------------|----------------------------|----------------------------------|
| B4 Zaupanje v državni zbor | Pearson Correlation | 1 | ,618** |
| | Sig. (2-tailed) | | ,000 |
| | N | 1387 | 1231 |
| B9 Zaupanje v evropski parlament | Pearson Correlation | ,618** | 1 |
| | Sig. (2-tailed) | ,000 | |
| | N | 1231 | 1247 |

Korelacija je značilna pri stopnji značilnosti 0,01 (verjetnost napake ob zavrnitvi ničelne domneve je manjša od 1%.)

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

Povezanost na vzorcu: Spremenljivki "zaupanje v državni zbor" in "zaupanje v evropski parlament" sta na vzorcu srednje močno povezani ($r = 0,618$, izračunan na 1231 enotah). Tisti, ki bolj zaupajo v slovenski parlament, tudi bolj zaupajo v evropski parlament.

Test o povezanosti na populaciji: Pri stopnji značilnosti $< 0,0005$ ($< 0,05\%$) lahko trdimo, linearna povezanost med spremenljivkama "zaupanje v državni zbor" in "zaupanje v evropski parlament" obstaja tudi na populaciji. (Ob pravilni domnevi, da ni linearne povezanosti med spremenljivkama, je verjetnost, da na vzorcu 1231 enot dobimo $r = 0.618$ ali več, skoraj ničelna ($< 0,05\%$), zato zavrnemo ničelno domnevo. Verjetnost, da smo s tem naredili napako, je manjša od $0,05\%$.)

28

Interpretacija

Correlations

| | | B4 Zaupanje v državni zbor | B9 Zaupanje v evropski parlament |
|--|---------------------|-------------------------------|--|
| B4 Zaupanje v državni zbor | Pearson Correlation | 1 | ,618** |
| | Sig. (2-tailed) | | ,000 |
| | N | 1387 | 1231 |
| B9 Zaupanje v evropski parlament | Pearson Correlation | ,618** | 1 |
| | Sig. (2-tailed) | ,000 | |
| | N | 1231 | 1247 |

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

29

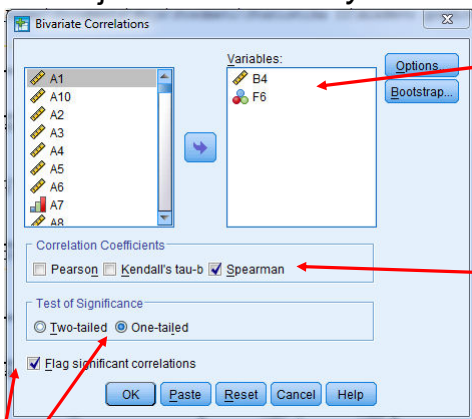
Ponovitev: test za linearno povezanost dveh ordinalnih spremenljivk

- Test uporabljamo za preverjanje domneve (testiranje hipoteze) o povezanosti dveh ordinalnih spremenljivk (ali ene ordinalne, druge intervalne).
- Mera povezanosti: Spearmanov koeficient korelacije rangov, r_S (na vzorcu) oz. ρ_S na populaciji.
- Interval vrednosti koeficienta: $[-1, 1]$.
 - ✓ Vrednost 0: ni povezanosti.
 - ✓ Vrednost -1 : popolna negativna povezanost (z naraščanjem rangov ene spremenljivke padajo rangi druge spremenljivke).
 - ✓ Vrednost $+1$: popolna pozitivna povezanost (z naraščanjem rangov ene spremenljivke naraščajo rangi druge spremenljivke).
- Domnevi:
 - H_0 : $\rho_S = 0$ – spremenljivki **nista** povezani na populaciji
 - H_1 : $\rho_S \neq 0$ – spremenljivki **sta** povezani na populaciji
(Lahko tudi: H_1 : $\rho < 0$ oz. H_1 : $\rho > 0$, negativna oz. pozitivna povezanost.)
- Če ničelno hipotezo zavrnemo, lahko trdimo, da sta spremenljivki (pozitivno, negativno) povezani.

30

Primer: Povezanost med stopnjo izobrazbe in zaupanjem v državni zbor

V meniju izberemo: *Analyze – Correlate – Bivariate*.



Spremenljivke, med katerimi računamo korelacijo (povezanost). Če izberemo več spremenljivk, se korelacija izračuna za vse možne pare spremenljivk.

Ker je vsaj ena spremenljivka ordinalna, izberemo Spearmanov koeficient korelacije rangov

Izberemo, ali želimo enostranski test (ali vnaprej predpostavljamo smer oz. predznak povezanosti) ali dvostranski test (ne predpostavljamo smeri).

Želimo, da so statistično značilni koeficienti označeni z zvezdicami *.

31

Primer: Povezanost med stopnjo izobrazbe in zaupanjem v državni zbor

Rezultat: Korelacijska matrika: Spearmanov koeficient korelacije za vse možne pare spremenljivk

| Correlations | | | B4 Zaupanje v državni zbor | F6 Highest level of education |
|----------------|-------------------------------|-------------------------|----------------------------|-------------------------------|
| Spearman's rho | B4 Zaupanje v državni zbor | Correlation Coefficient | 1,000 | ,197** |
| | | Sig. (1-tailed) | . | ,000 |
| | | N | 1387 | 1384 |
| | F6 Highest level of education | Correlation Coefficient | ,197** | 1,000 |
| | | Sig. (1-tailed) | ,000 | . |
| | | N | 1384 | 1439 |

Vrednost (izbranega) koeficienta korelacije

Natančna stopnja značilnosti (enostranski test)

Število enot

** . Correlation is significant at the 0.01 level (1-tailed).

Korelacija je značilna pri stopnji značilnosti 0,01 (verjetnost napake ob zavrnitvi ničelne domneve je manjša od 1%.)

32

Interpretacija

Correlations

| | | | B4 Zaupanje v državni zbor | F6 Highest level of education |
|----------------|-------------------------------|-------------------------|----------------------------|-------------------------------|
| Spearman's rho | B4 Zaupanje v državni zbor | Correlation Coefficient | 1,000 | ,197** |
| | | Sig. (1-tailed) | . | ,000 |
| | | N | 1387 | 1384 |
| | F6 Highest level of education | Correlation Coefficient | ,197** | 1,000 |
| | | Sig. (1-tailed) | ,000 | . |
| | | N | 1384 | 1439 |

** . Correlation is significant at the 0.01 level (1-tailed).

Povezanost na vzorcu: Spremenljivki "zaupanje v državni zbor" in "izobrazba" sta na vzorcu šibko pozitivno povezani ($r_S = 0,197$, izračunan na 1384 enotah). Bolj izobraženi tudi bolj zaupajo v državni zbor.

Test o povezanosti na populaciji: Pri stopnji značilnosti $< 0,0005$ ($< 0,05\%$) lahko trdimo, na populaciji obstaja pozitivna povezanost med spremenljivkama "zaupanje v državni zbor" in "izobrazba". (Ob pravilni domnevi, da ni pozitivne povezanosti med spremenljivkama, je verjetnost, da na vzorcu 1384 enot dobimo $r_S = 0,197$ ali več, skoraj ničelna ($< 0,05\%$), zato zavrnemo ničelno domnevo. Verjetnost, da smo s tem naredili napako, je manjša od $0,05\%$.)

33