UNIVERZA V LJUBLJANI
PEDAGOŠKA FAKULTETA
ODDELEK ZA SOCIALNO PEDAGOGIKO

PA NE SPET SPSS!!!

GRADIVO ZA UČENJE OSNOV UPORABE SPSS-A

(za interno uporabo)

Avtorica: dr. Mija M. Klemenčič Rozman
Pregled: dr. Bojan Dekleva

Ljubljana, oktober 2015

Pa ne spet SPSS: gradivo za učenje osnov uporabe SPSS-a
(za interno uporabo)

Avtorica: Mija M. Klemenčič Rozman
Strokovni pregled: Bojan Dekleva

Izdala in založila: Pedagoška fakulteta Univerze v Ljubljani
Za Pedagoško fakulteto: Janez Krek, dekan
Oblikovanje in tehnično urejanje: avtorica

Dosegljivo na: https://www.pef.uni-lj.si/fileadmin/Datoteke/Zalozba/e-publikacije/uvod-v-spss_2015.pdf

CIP – Kataložni zapis o publikaciji
Narodna in univerzitetna knjižnica Ljubljana

004.42:311.1(0.034.2)

KLEMENČIČ Rozman, Mija Marija
Pa ne spet SPSS!!! [Elektronski vir] : gradivo za učenje osnov uporabe SPSS-a : (za interno uporabo) / avtorica Mija M. Klemenčič Rozman. - El. knjiga. - Ljubljana : Pedagoška fakulteta, 2015

Način dostopa (URL): https://www.pef.uni-lj.si/fileadmin/Datoteke/Zalozba/e-publikacije/uvod-v-spss_2015.pdf

ISBN 978-961-253-185-0 (pdf)

282116352

KAZALO

NAMESTO UVODA: PA SPET SPSS ☺

Pričujoče gradivo sem napisala leta 2005 kot spodbudo za delo pri vajah pri predmetu Preddiplomski seminar na takrat še starem študijskem programu Socialne pedagogike na Oddelku za socialno pedagogiko Pedagoške fakultete Univerze v Ljubljani. Nastalo je na osnovi opažanj, da so za študentke in študente prvi koraki v program SPSS lahko frustrirajoči, kar sem izkusila tudi na lastni koži in kasneje spoznala, da je bila vsa frustracija odveč. Vsa ta leta je gradivo romalo po e-naslovih kot interno gradivo pri predmetu, kar je bila posledica ideje, da bo nekoč dopolnjeno in prenovljeno ugledalo luč sveta. V tem trenutku mi je jasno, da se to ne bo nikoli zgodilo, saj se s tem področjem nisem in se verjetno tudi ne bom nikoli poglobljeno ukvarjala. So bile pa povratne informacije tistih, ki so skripto dobili v roke, zame toliko opogumljajoče, da posodabljam ta uvod in da v elektronski verziji postane gradivo tudi širše dostopno.

Celotno gradivo je zasnovano na statistični analizi podatkov pridobljenih s pomočjo *Vprašalnika o nekaterih značilnostih življenja študentk in študentov 4. letnika socialne pedagogike*, ki so ga skozi pretekla študijska leta izpolnjevale_i študentke in študenti pri predmetu Preddiplomski seminar. Slednji je bil v prenovljenem študijskem programu spremenjen v Raziskovalni projekt. Ker gradivo predstavlja podporo pri izvedbi vaj pri tem predmetu, so zajete zgolj izbrane teme in zgolj nekateri vidiki uporabe SPSS-a pri izvedbi posameznih statističnih metod. Zaradi interne rabe gradiva to ostalo nelektorirano, kar z drugimi besedami pomeni, da boste pri prebiranju zagotovo našli kakšno napako. Če bom s tem nehote prizadela vašo skrb za lep slovenski jezik, se že vnaprej opravičujem.

V besedilu sem poskušala prikazati analize takih primerov, ki bi bili vsaj zanimivi, morda celo pikantni, da bi bralce in bralke, ki se šele spoznavajo s tem računalniškim programom, zadržala v poziciji radovednosti. Priznam, nekaj manipulacije je v tem pristopu ☺, a izhaja zgolj iz želje, da pokažem, da uporaba tega programa v raziskovalnem procesu ni noben bavbav. Obenem sem se pri pisanju držala sistema receptov oz. načina »če to, potem to«, kar pomeni, da je gradivo napisano izjemno poenostavljeno oz. preprosto. Prosim vas, drage bralke in dragi bralci, da ga kot takega tudi vzamete v roke. Gre zgolj za prve korake v svet uporabe SPSS-a. Zato zelo toplo priporočam, da si v zahtevnejši in metodološko bolj korektni literaturi pregledate obširnejše razlage posameznih teoretskih predpostavk, testov, izračunov… Osebno mi je na tej poti zagotovo najbolj koristila knjiga avtorja Andyja Fileda z naslovom Discovering Statistics

using SPSS[1] (mimogrede, kratica SPSS stoji za Statistical Package for the Social Sciences).

Omenjeni avtor tudi fino opomni na najizrazitejšo značilnost uporabe SPSS-a skritem v akronimu GIGO, kar stoji za sintagmo »Garbage in, garbage out«. Če sem se o čem prepričala v času spoznavanja z SPSS-om, je to, da GIGO še kako drži. Zato naj ne bo odveč vabilo, da je potrebno poznavanje statistike in pogojev za uporabo posamezne metode ter vsebinski razmislek o preučevanem pojavu, sicer lahko izračunamo neverjetne veleume, kot je denimo povezava med pogostostjo pregretij teflonske posode in številom pasjih kakcev na ulici. Ti veleumi so (pazite!) na prvi pogled toliko bolj verodostojni, ker za njimi stojijo številke.

Druga past, ki se sicer neposredno ne tiče uporabe SPSS-a, še kako pa se tiče moje osuplosti, pa je poročanje o dobljenih raziskovalnih rezultatih v splošno dostopnih medijskih objavah. V mislih imam začetek povedi z »raziskave kažejo«, zraven pa ni navedene niti institucije, ki je raziskavo opravila, letnice in naslova raziskave in numerusa z načinom vzorčenja. Kot da besedna zveza »raziskave kažejo« podarja avtorju ali avtorici neomejeno verodostojnost zapisa, nam bralkam in bralcem pa imperativ, da moramo zapisanemu slepo verjeti. Meni ta igra ni všeč, ker odvzame osrednjo značilnost znanstvenega delovanja, to je preverljivost dobljenih rezultatov, s tem pa odvzame tudi priložnost za dvom. Brez dvoma in preizpraševanja pa ni znanosti. Opisani pasti sta tudi dva od razlogov, zakaj si želim, da bi se (bodoče_i) strokovnjakinje_i psihosocialnih strok, manj bale_i kvantitativne metodologije.

Gradivo je že ob njegovem nastanku strokovno pregledal prof. dr. Bojan Dekleva, kateremu gre zahvala, da marsikateri lapsus ni ugledal luči sveta. Hkrati pa je bil Bojan tudi tisti, ki je prepoznal mojo »raziskovalno žilico« in me spremljal na poti spoznavanja z raziskovalnimi metodologijami, a mi je puščal dovolj prostora za avtonomno izbiro smeri na razpotjih. Bojan, hvala!

Naj za konec rečem le še, da močno upam, da bo pričujoče gradivo zmanjšalo kakšno začetno frustracijo pri vstopanju v svet SPSS-a. Še bolj kot to pa vam želim obilo raziskovalne radovednosti!

[1] Navedena v literaturi na koncu.

DILEME V ZVEZI Z VPRAŠALNIKOM

1. Kakšen je dober vprašalnik?

Dober vprašalnik mora imeti tri značilnosti. Biti mora:

- veljaven, kar pomeni, da res meri tisto značilnost, za katero mislimo, da jo meri;
- objektiven, kar pomeni, da je rezultat merjenja odvisen samo od lastnosti, ki jo merimo, ne pa tudi od tistega, ki merjenje izvaja;
- zanesljiv, kar pomeni, da daje pri ponovljenih merjenjih istih lastnosti pri istih osebah enake rezultate, torej vedno meri isto lastnost (Mesec, 1997). Več o merjenju zanesljivosti v poglavju Zanesljivost vprašalnika.

Standardiziran vprašalnik ima vse te značilnosti, ker je bil uporabljen na več raziskavah z velikim vzorcem, predhodno pilotsko preverjen, popravljen ter analiziran tako, da resnično dobro meri parametre, ki naj bi jih meril. Vendar je v praksi malo težje priti do teh vprašalnikov, ko se odločimo, da bi jih uporabili pri seminarski nalogi.

2. Kako naj sam_a sestavim vprašalnik?

Čeprav to ni ravno najboljša pot pri raziskovalni nalogi, saj je *ad hoc* sestavljen vprašalnik lahko pomanjkljiv, površen in premalo zanesljiv, se ga mnogi še vedno poslužujejo, kar za namene učenja načeloma ni narobe, je pa potrebno biti nekoliko bolj pazljiv na nekatere pogoste napake. Sestavljen vprašalnik najprej pilotsko preverimo, npr. na manjšem vzorcu, ki ustreza značilnostim pravega vzorca. Ko sestavljamo vprašalnik, je priporočljivo, da si ogledamo različna poročila raziskav na našo raziskovalno temo, saj ta ponavadi vključujejo tudi npr. uporabljene lestvice v vprašalniku in ponavadi tako ugotovimo, da smo na kako področje pozabili, da smo nekatera vprašanja zastavili nekoliko okorno itd.

3. Katere so najbolj splošne napake pri sestavljanju vprašalnika?

- Manjka motivacijski uvod in predstavitev.
- Nejasna in površna navodila v uvodu vprašalnika, kjer udeležence prosimo za sodelovanje.
- Nejasno in okorno zastavljena vprašanja.
- Jezik vprašanj ni prilagojen starosti ali pa kaki drugi značilnosti vprašanih.
- Nepregleden in oblikovno prenasičen vprašalnik.
- Nejasno uporabljeni pojmi (npr. živim na: a) vasi, b) manjšem mestu, c) večjem mestu. Tu je bolje, da pripišete število prebivalcev, da se bodo udeleženci raziskave lažje opredelili. Kaj je večje in kaj manjše mesto, je namreč odvisno tudi od subjektivnega pogleda).

- Izpuščena zaporedna številka pred vprašanjem. Npr. Vprašanja so oštevilčena 1., 2., 3., 4., 6., 7., 8. To povzroči predvsem zmedo pri vnašanju podatkov.
- Če je v vprašalniku zahtevano rangiranje podatkov, t.j. da jih posamezniki razvrstijo od 1-7 po prioriteti kot velja za njih, je pomembno, da je navodilo podano jasno. Zgodi se, da posamezniki ne rangirajo postavk vprašalnika, temveč jim dva ali trikrat pripišejo iste številke.
- Nepremišljeno postavljeni možni odgovori – da npr. kak odgovor sploh ni zajet med možnimi odgovori. Pa še priporočilo: dobro je, če okence pred odgovorom oštevilčimo z malo številko, ki nam kasneje služi pri vnašanju podatkov v SPSS. Npr. okence pred odgovorom »ženska« oštevilčite z 1, okence pred odgovorom »moški« pa z 2.
- Pri označevanju odgovorov na lestvicah stališč, ki imajo npr. 4 stopnje: se popolnoma strinjam, se deloma strinjam, se ne strinjam in se sploh ne strinjam, naj bodo odgovori postavljeni opisno in ne s številkami od 1 – 4, saj številke nekako implicirajo, da je 4 več vredna od 1 itd. (Mesec, 1997).

4. Kaj naj vsebuje uvod vprašalnika?

- Vašo predstavitev; kdo ste in od kod prihajate (institucija)
- pojasnilo s kakšnim namenom izvajate raziskavo
- zagotovitev anonimnosti (če je vprašalnik res anonimen)
- navodila, kako naj posamezniki rešujejo vprašalnik
- vašo zahvalo za sodelovanje.

Če boste vprašalnik sami razdeljevali skupinam (npr. v razredu), naj tudi vaša predstavitev pred razdelitvijo vprašalnika vsebuje enake informacije kot v uvodu, s tem da nikar ne pozabite na koncu dati možnosti, da lahko udeleženci vprašajo, če česa niso razumeli ali pa če jih zanima še kaj več v zvezi z raziskavo.

5. Kaj je bolje razdeliti vprašalnik neposredno ljudem in počakati, da ga izpolnijo ali pa jih poslati po pošti?

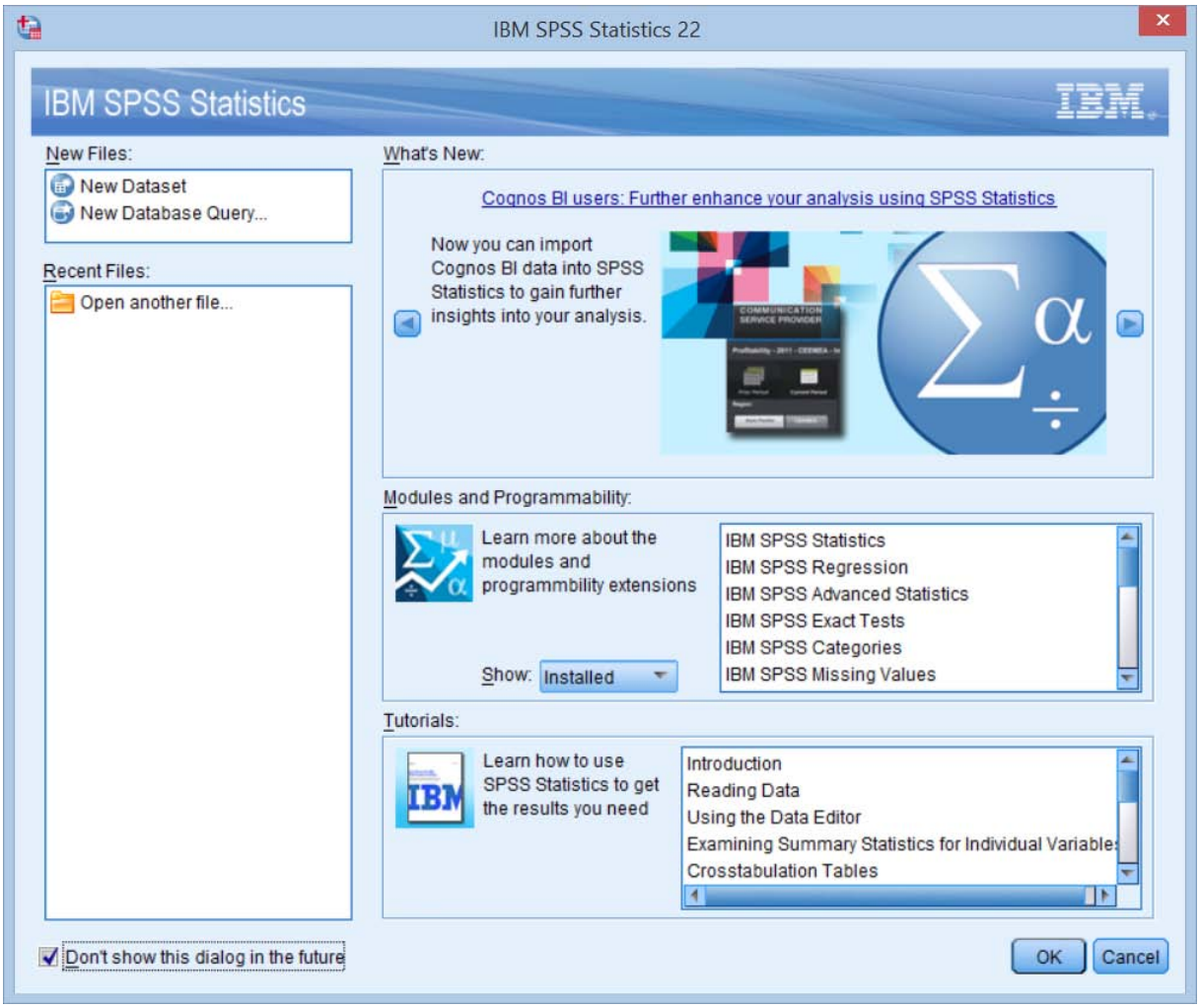
Tu nekako ni pravilnega odgovora. Za kakšen način razdeljevanja vprašalnikov se odločite, je odvisno od načina vzorčenja, saj nekaterih skupin ljudi ne boste mogli doseči na enem kupu. Prednost neposrednega razdeljevanja vprašalnikov je vsekakor ta, da jih (skoraj) vsi posamezniki izpolnijo in jih takoj dobite nazaj. Poleg tega lahko pojasnite kako nejasnost in obrazložite navodila na začetku. Če vprašalnike pošiljate po pošti, lahko računate na približno četrtino do tretjino vrnjenih vprašalnikov. Če se denimo odločite za metodo snežne kepe, je odvisno, kako dobro socialno mrežo boste dobili. Morda vam bo uspelo dokaj hitro dobiti nazaj (skoraj) vse vprašalnike. Tako da na kratko lahko zaključimo takole: za kakšen način razdeljevanja vprašalnikov se odločite, je odvisno od metode vzorčenja, ta pa je odvisna od populacije, ki jo želite preučevati, ta pa od fenomena, ki ga raziskujete. Zato si vzemite nekoliko časa za razmislek pred tem korakom.

6. Kaj je prvi korak, ko dobim nazaj vprašalnike?

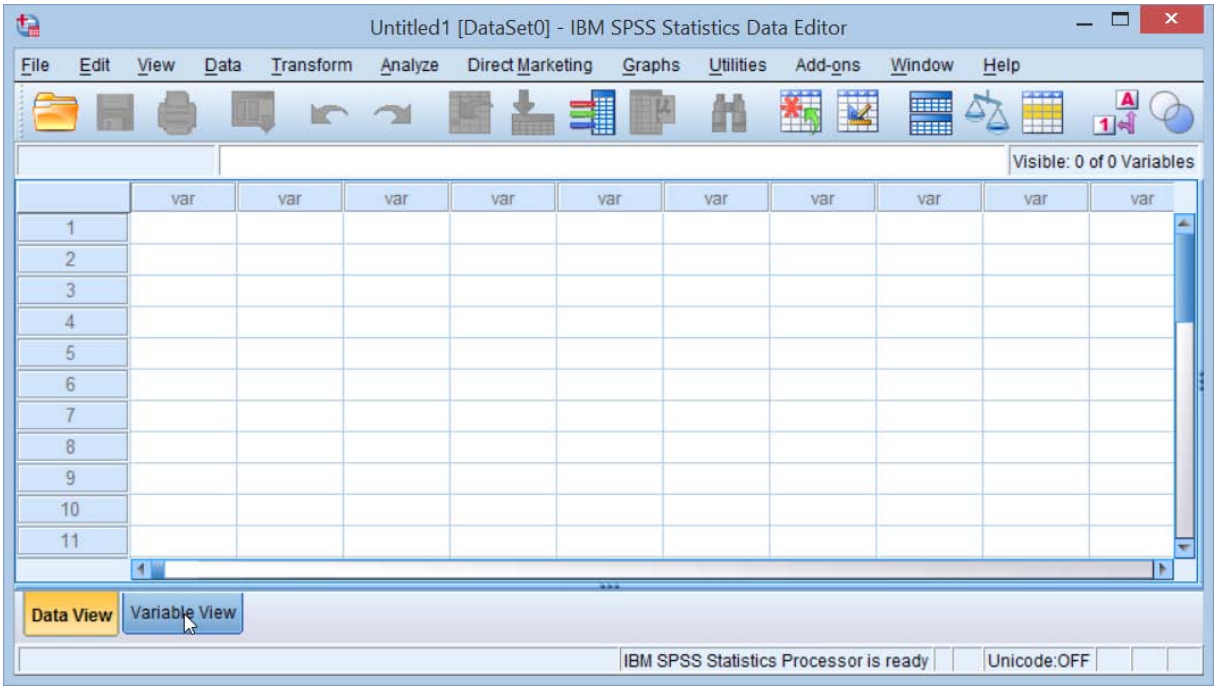
No, najprej zahvala ljudem, da so se odzvali na vašo prošnjo in si vzeli čas za izpolnjevanje vprašalnika. Potem pa je resnično prvi korak ta, da vse vprašalnike po vrsti ročno oštevilčimo. Začnemo z 01, 02… Oštevilčenje je zelo pomembno za kasnejši vnos podatkov. Ko ugotovimo, da so nekateri vnosi podatkov v SPSS nekonsistentni ali pa ugotovimo, da smo se pri vnosu zmotili, lahko zlahka ugotovimo, kje smo naredili napako, saj hitro najdemo npr. 121-i vprašalnik, za katerega smo pri preverbi ugotovili, da je prišlo do napačnega vnosa.

VNAŠANJE PODATKOV V SPSS

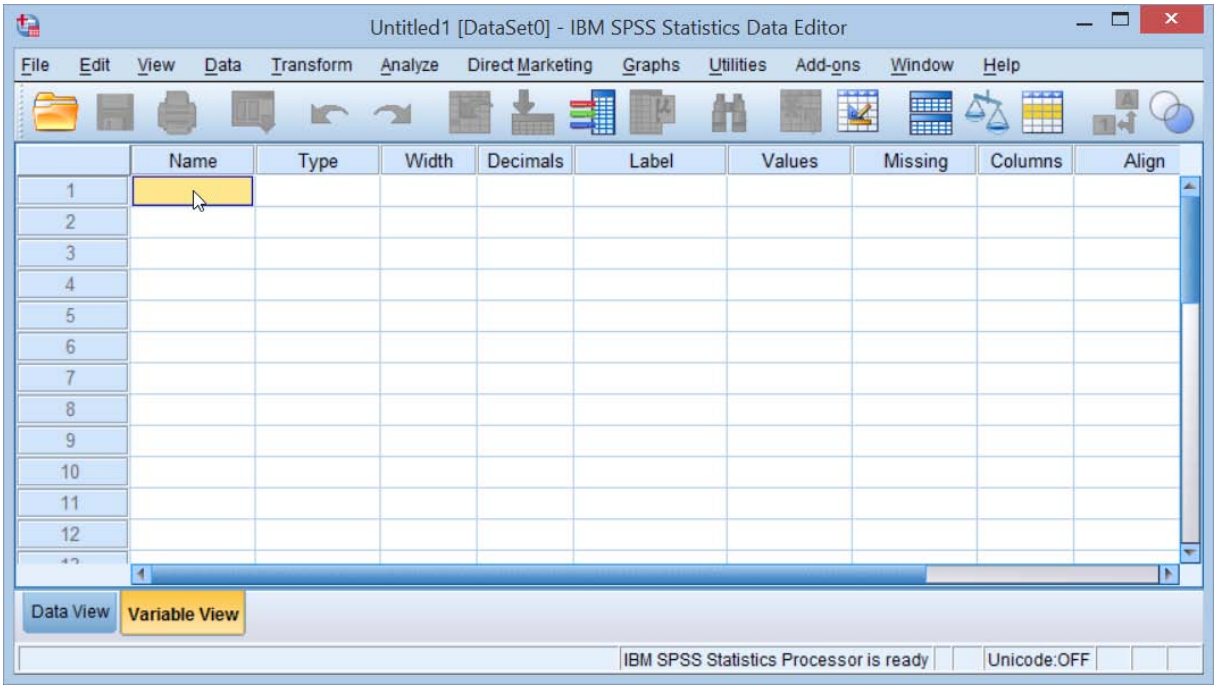
Odprite program SPSS. Pojavi se vam začetno okno. V njem kliknite Don't show this dialog in the future (tako se vam naslednjič to okno ne bo več odprlo in boste imeli dostop neposredno do podatkov).



Ko ste zgornje okno zaprli, se vam prikaže naslednje okno. V njem vnašamo podatke ter delamo različne izračune.

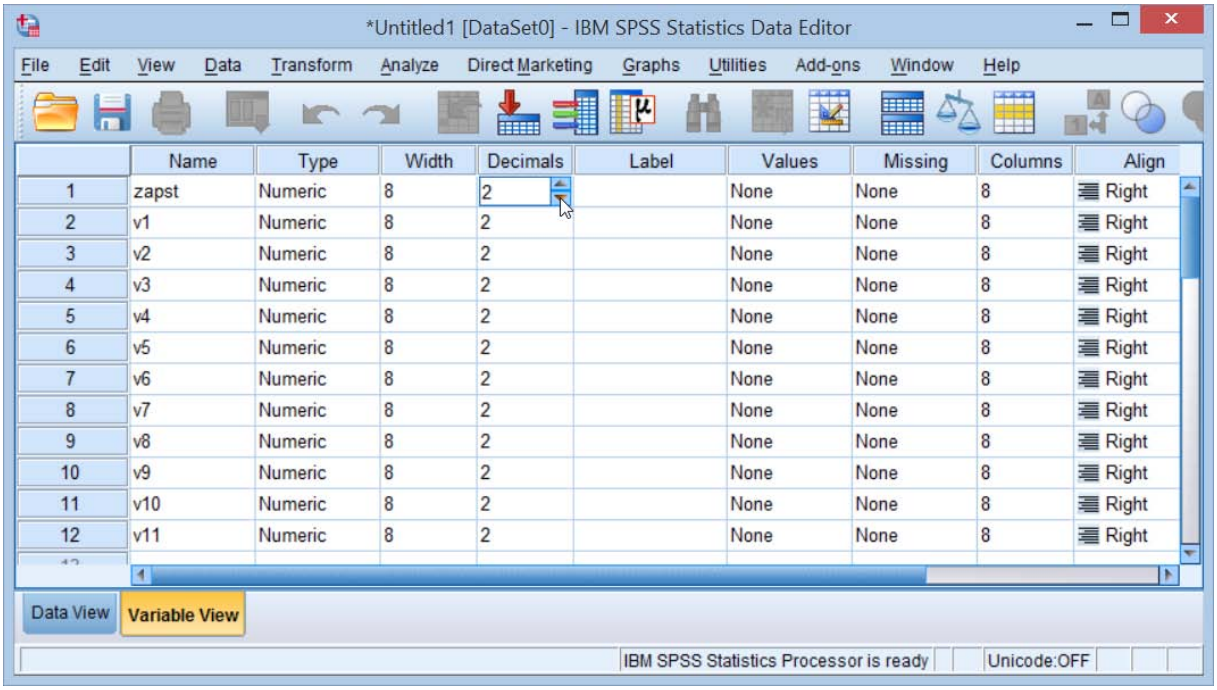


Še preden bomo začeli vnašati podatke iz vprašalnikov v SPSS, bomo pripravili okolje za vnos podatkov. Kliknite na Variable View v levem spodnjem kotu okna. Prikaže se vam tako okno:

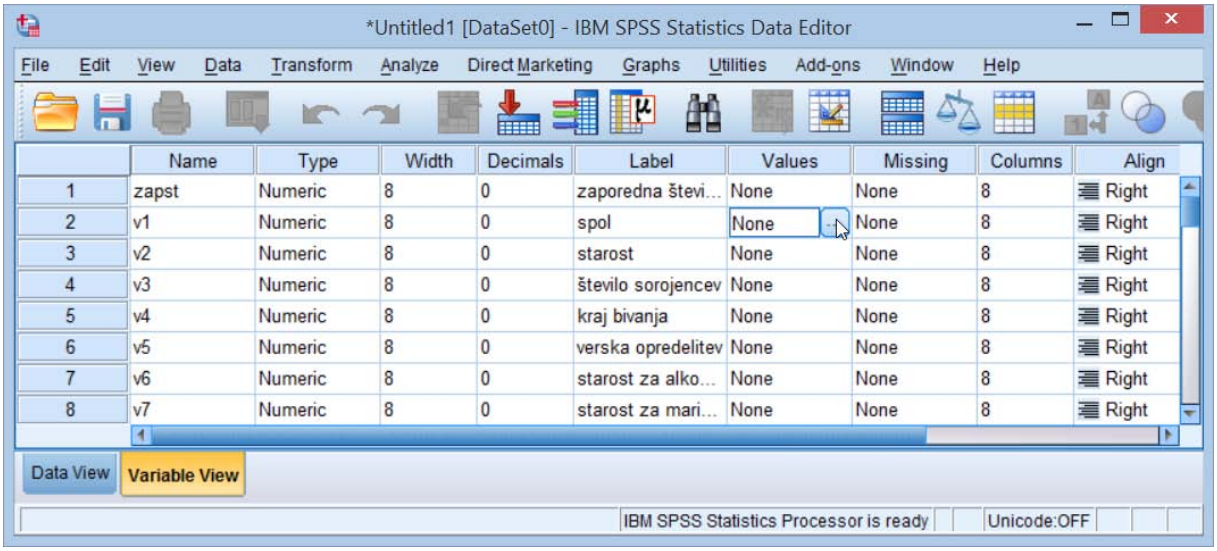


V polje Name vnašamo zaporedne številke vprašanj tako kot si sledijo v vprašalniku. Le v prvo okence vnesemo zaporedno številko vprašalnika (vsak vprašalnik mora biti oštevilčen). Tako v stolpec Name začnemo vnašati: zapst, v01 (pomeni vprašanje 1), v02, v03, v04, v05 itd. – če imamo največ do 99 spremenljivk/vprašanj, oz. v001, v002, v003, v004, če imamo

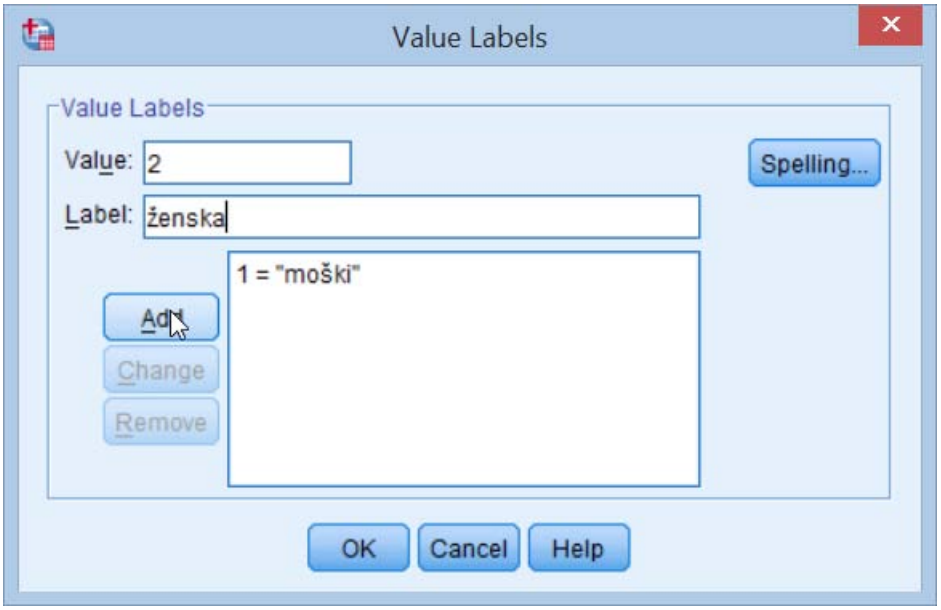
več kot sto spremenljvk/vprašanj. Če v polje Name pišete besede s sičniki ali šumniki pri starih različicah tega računalniškega programa lahko naletite na težave. Stolpca Type ne odpiramo, če imamo se odgovore oštevilčene in zato je tip spremenljivke označen kot Numeric. Field (2005) pravi, da le v primeru, ko bi želeli vpisati v SPSS nek komentar o vprašanem ali kako drugo informacijo, ki je ne želimo definirati kot skupinsko spremenljivko (denimo priimek vprašane osebe), kliknemo na desno stran okenca z besedo Numeric v želeni vrstici in ko se nam odpre okno Variable Type, kliknemo na String. String pomeni zapis podatkov z besedo (kot smo rekli, npr. priimek). V praksi začetniki tega skorajda ne potrebujejo. Prav tako ne spreminjamo stolpca Width, ki določa, koliko je največje mestno število za zapis podatka. Stolpec Decimals pa zaradi preglednosti zmanjšamo na 0 kot prikazuje slika spodaj, če vnašamo podatke, ki so zapisani kot cela števila.



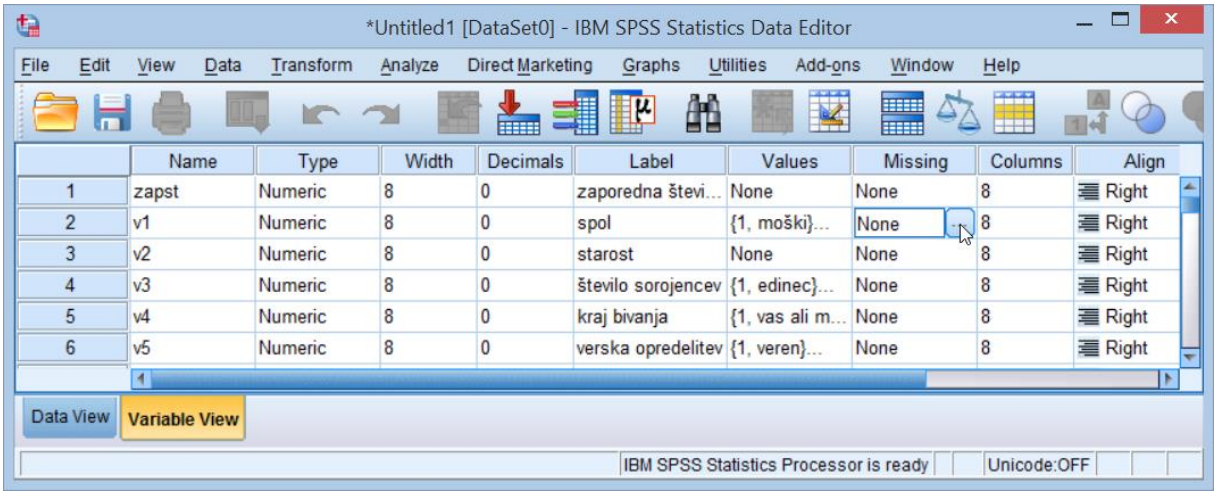
V stolpcu Label zapišemo pri vsakem vprašanju ime spremenljivke (npr. pri v1 zapišemo spol, ker to vprašanje sprašuje po tej spremenljivki). Zapis v stolpcu Label sedaj zgleda takole:



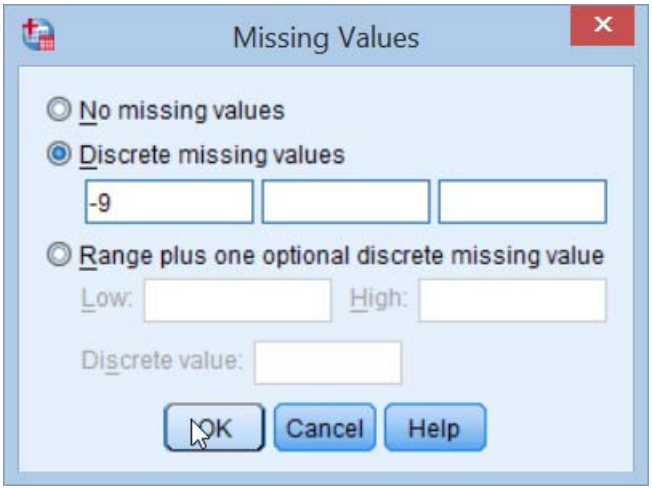
Sedaj v vsaki vrstici (razen prvi oz. tisti, kjer je podatek numeričen, npr. starost) kliknemo v stolpcu Values na desno stran in odpre se nam okence:



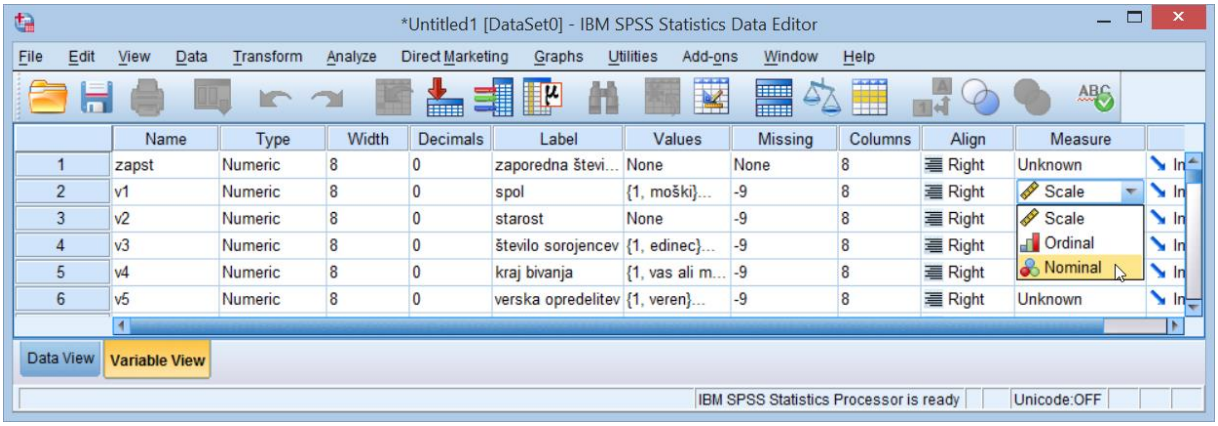
V vrstico Value vnesemo numerično vrednost za odgovor ter ga v vrstici Value Label označimo z besedo, nato kliknemo Add, ko končamo z oznako vseh možnih vrednosti odgovorov pa še OK. Ko to naredimo za vse spremenljivke, zgleda takole:



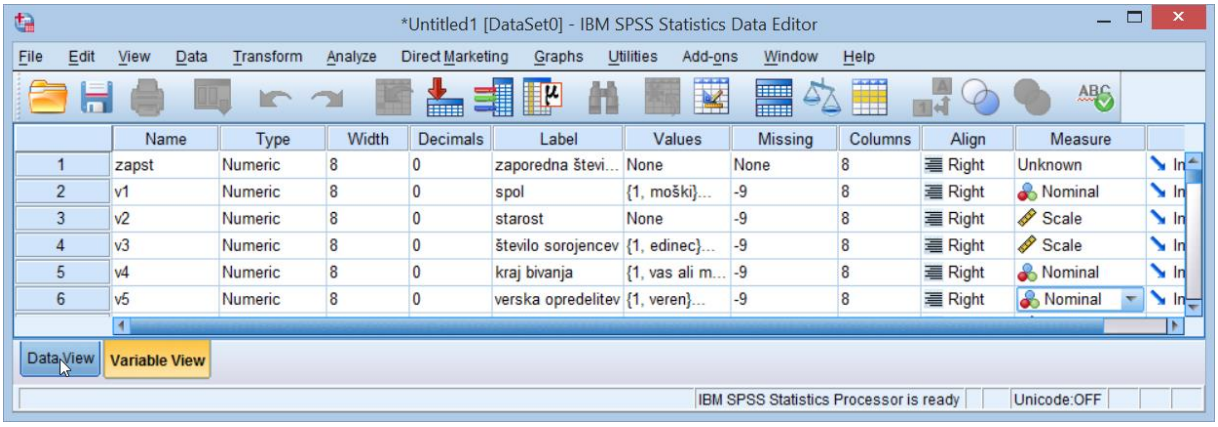
Sedaj v stolpcu Missing določimo manjkajoče vrednosti, kar pomeni vrednost, ki bo v programu pomenila, da ni bilo odgovora. Manjkajoče vrednosti določimo tako, da kliknemo na desno stran okenca None v stolpcu Missing. Odpre se nam naslednje pogovorno okno:



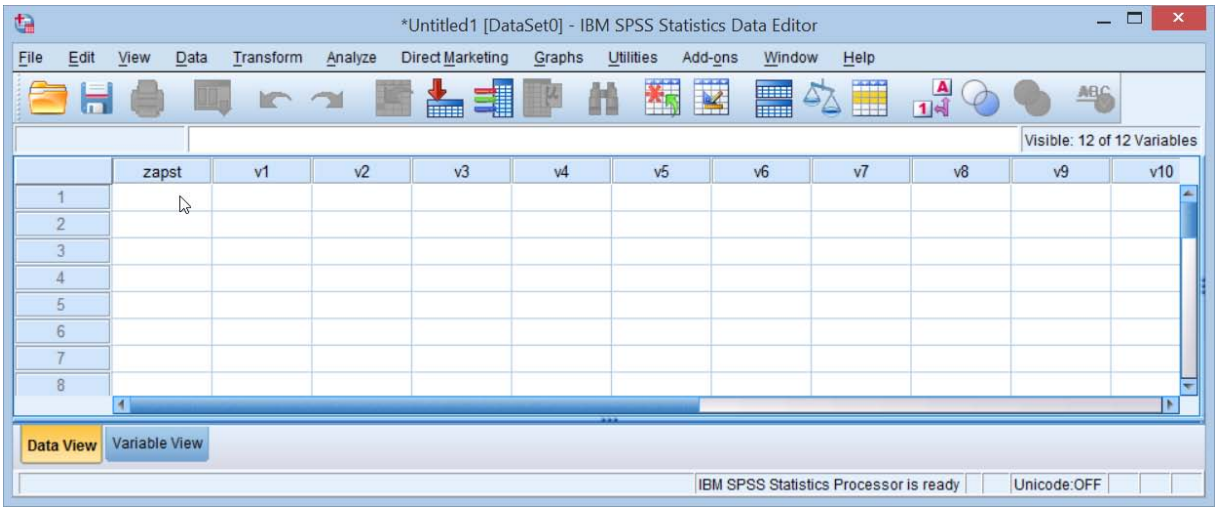
Kliknite Discrete missing values in v okence vnesite -9 (lahko pa tudi 9 ali 99, vendar le, če ti dve vrednosti nista možni vrednosti odgovora). In nato kliknite ok. To skopirajte v vsako vrstico tega stolpca (razen prve, ki označuje le zaporedno številko vprašalnika in zato ni manjkajoče vrednosti). Tako smo za vrednost -9 v programu SPSS določili, da jo vnesemo vedno, ko ni bilo odgovora pri posameznem vprašalniku na določeno vprašanje in program bo te informacije upošteval pri računanju parametrov in testov. Sedaj okno Variable View zgleda takole:



Stolpca Columns, ki določa število znakov v vrstici ter Align, ki določa poravnavo številk in teksta, ne spreminjamo. Ostane nam samo še stolpec Measure. SPSS ima nastavljene spremenljivke kot Unknown, zato jih moramo še določiti. V primeru, da je naša spremenljivka npr. nominalna (glej v1 – spol), kliknemo na desno stran okenca Unknown in iz spustnega seznama izberemo opcijo Nominal. Tako naredimo za vse nominalne spremenljivke. Če so naše spremenljivke ordinalne, kliknemo opcijo Ordinal, za numerične pa Scale. Ko končamo z nastavitvijo vseh teh stolpcov, dobimo:

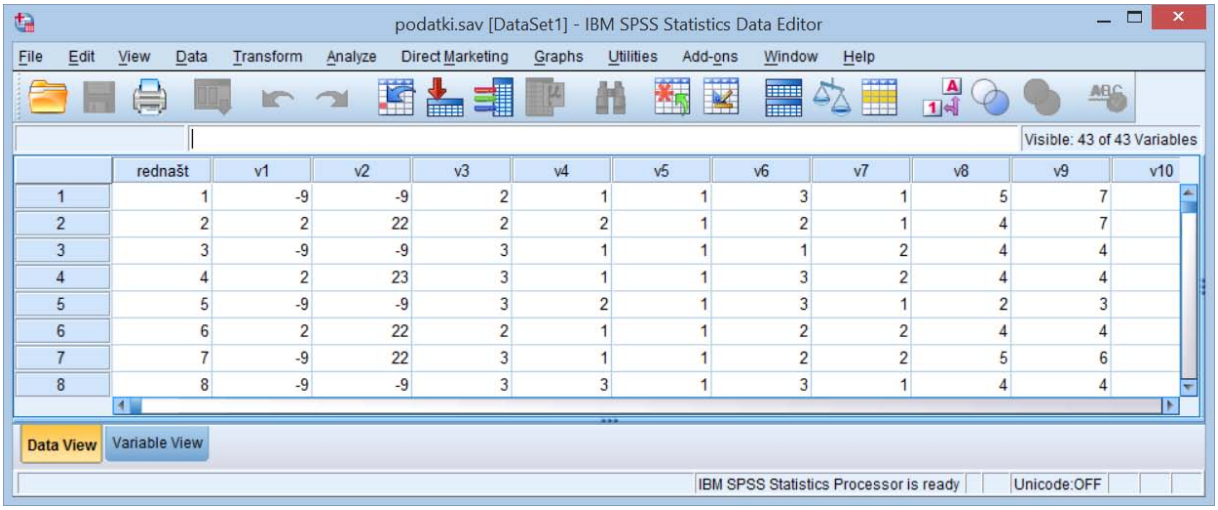


Sedaj kliknemo nazaj na okno Data View in začnemo z vnosom podatkov. Pazite: v Variable View vnašamo vprašanja v stolpcu Name in jih beremo vertikalno. Če pa kliknemo nazaj na Data View, si vprašanja sledijo horizontalno v prvi vrstici tabele. To zgleda takole:



Preden začnete vnašati podatke, si shranite nov dokument z ukazom File/Save as. Naše podatke bomo shranili kot podatki.sav. Še bolje je, če dokument shranite že na začetku, če pa ste ga takrat pozabili, ga sedaj nikar ne pozabite.

Sedaj pričnemo vnašati podatke za vsak posamezen vprašalnik vodoravno v vrstico (zapis v eni vrstici predstavlja vse podatke za enega posameznika). Ko končamo je v oknu za naš primer zapisano:



	rednašt	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10
1	1	-9	-9	2	1	1	3	1	5	7	
2	2	2	22	2	2	1	2	1	4	7	
3	3	-9	-9	3	1	1	1	2	4	4	
4	4	2	23	3	1	1	3	2	4	4	
5	5	-9	-9	3	2	1	3	1	2	3	
6	6	2	22	2	1	1	2	2	4	4	
7	7	-9	22	3	1	1	2	2	5	6	
8	8	-9	-9	3	3	1	3	1	4	4	

Sedaj, ko imamo vnešene vse podatke, lahko začnemo z njihovo obdelavo. Kako izračunamo posamezne parametre ter teste, je prikazano v naslednjih poglavjih.

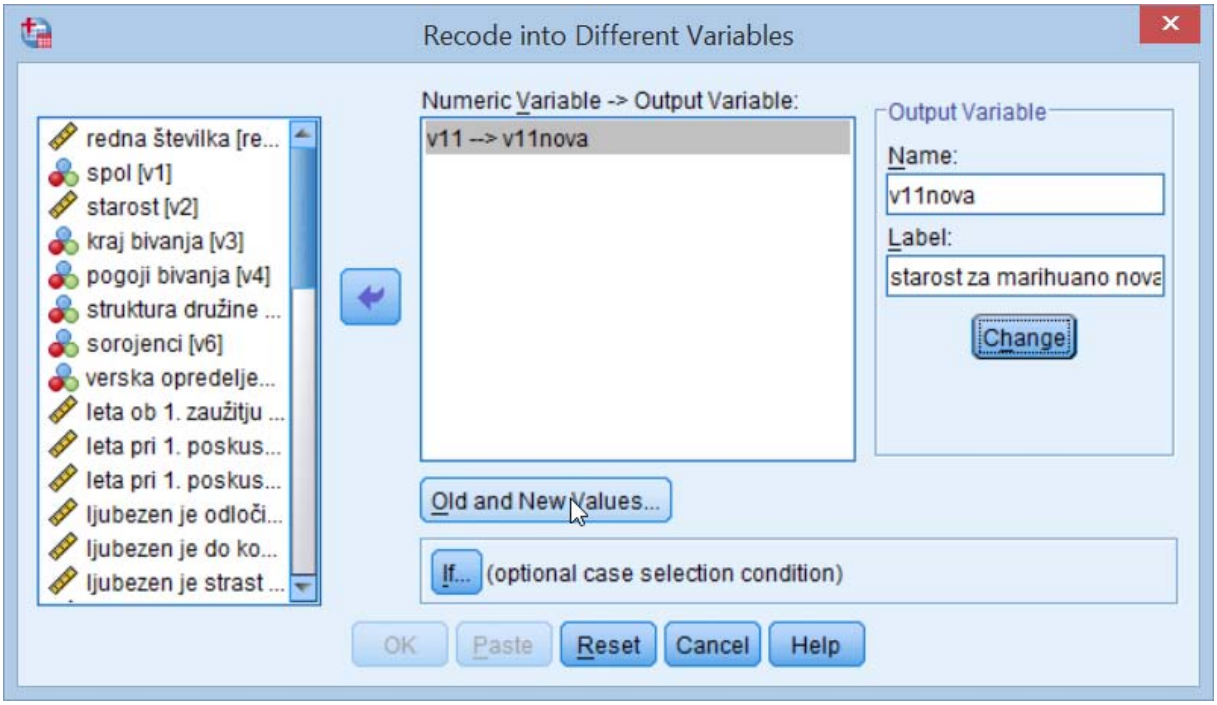
PRILAGAJANJE PODATKOV IN OBLIKOVANJE NOVIH SPREMENLJIVK

1. Zakaj naj prilagodim podatke in oblikujem nove spremenljivke?

- Zgodi se, da želimo izračunati nek parameter, a ugotovimo, da zaradi premajhnega numerusa nimamo izpolnjenih pogojev. To se nam npr. lahko velikokrat zgodi pri izračunu χ^2 – testa. Takrat je ena od možnosti, da kategorije posameznih spremenljivk združimo. Uporabimo funkcijo Recode.
- Včasih potrebujemo novo spremenljivko, ki jo lahko izračunamo iz danih spremenljivk. V ta namen uporabimo funkcijo Compute.
- Včasih pa želimo delati analizo le na določenem delu podatkov. Takrat uporabimo funkcijo Select Cases.

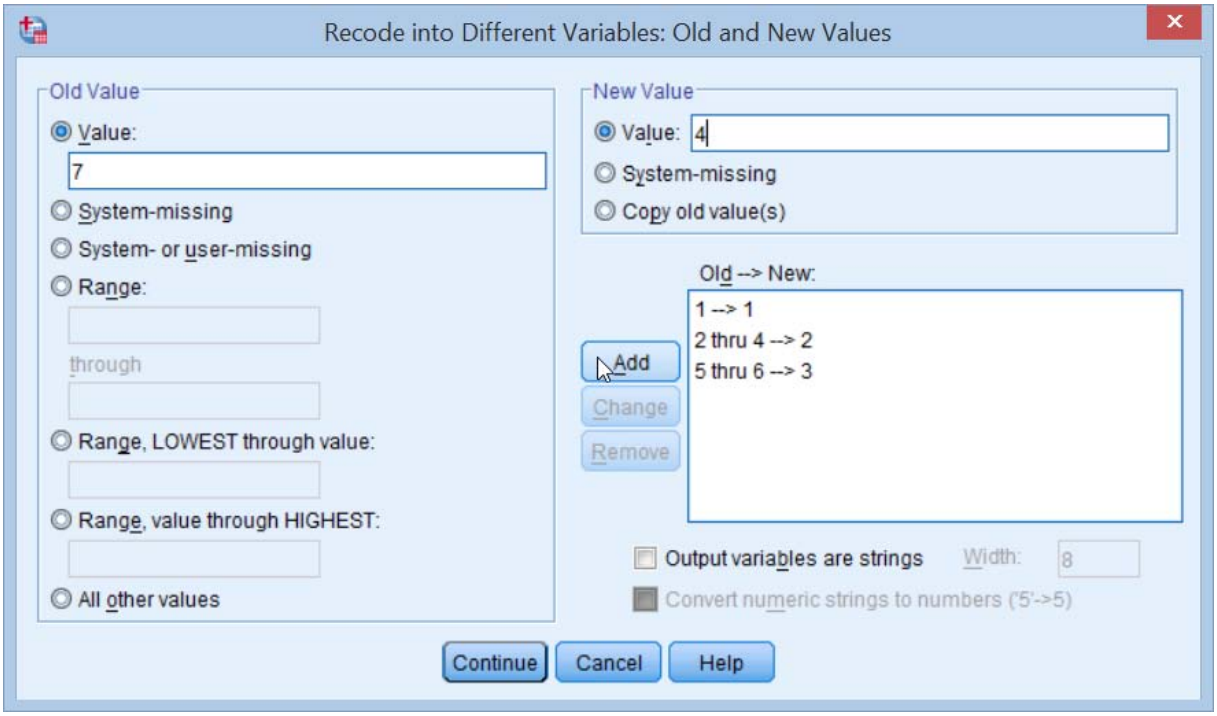
2. Kje najdem funkcijo Recode v SPSS-u?

Odprite pogovorno okno Transform → Recode Into Different Variables (tako ustvarite novo spremenljivko ob tem, da staro obdržite). Če pa želite podatke rekodirati v isto spremenljivko, uporabite Into Same Variables (ne priporočam za začetnike). Nato vnesite oznako in ime nove variable ter kliknite Change. Nato kliknite okence Old & New values.



Zgornja slika prikazuje primer za vprašanje Starost ob prvem poskusu marihuane, kjer imamo na voljo 7 kategorij: 1 – nikoli, 2 – 12 let ali manj, 3 – 13 let, 4 – 14 let, 5 – 15 let, 6 – 16 let ter 7 – 17 let ali več. Denimo, da želimo spremenljivko izraziti z drugačnimi kategorijami in sicer: 1 – nikoli, 2 – 14 let ali manj, 3 – 15 do 16 let ter 4 – 17 let ali več. Vrednosti prekodiramo takole.

V okence Old Value vnesemo staro vrednost, v okence New Value pa novo, nato kliknemo Add. Tako naredimo za vse vrednosti. Če želimo npr. vrednosti 2, 3 in 4 prekodirati v 2, potem izberemo opcijo Range (četrti krogec) in napišemo v zgornje okence 2, v spodnje pa 4, nato določimo novo vrednost 2 ter kliknemo Add.

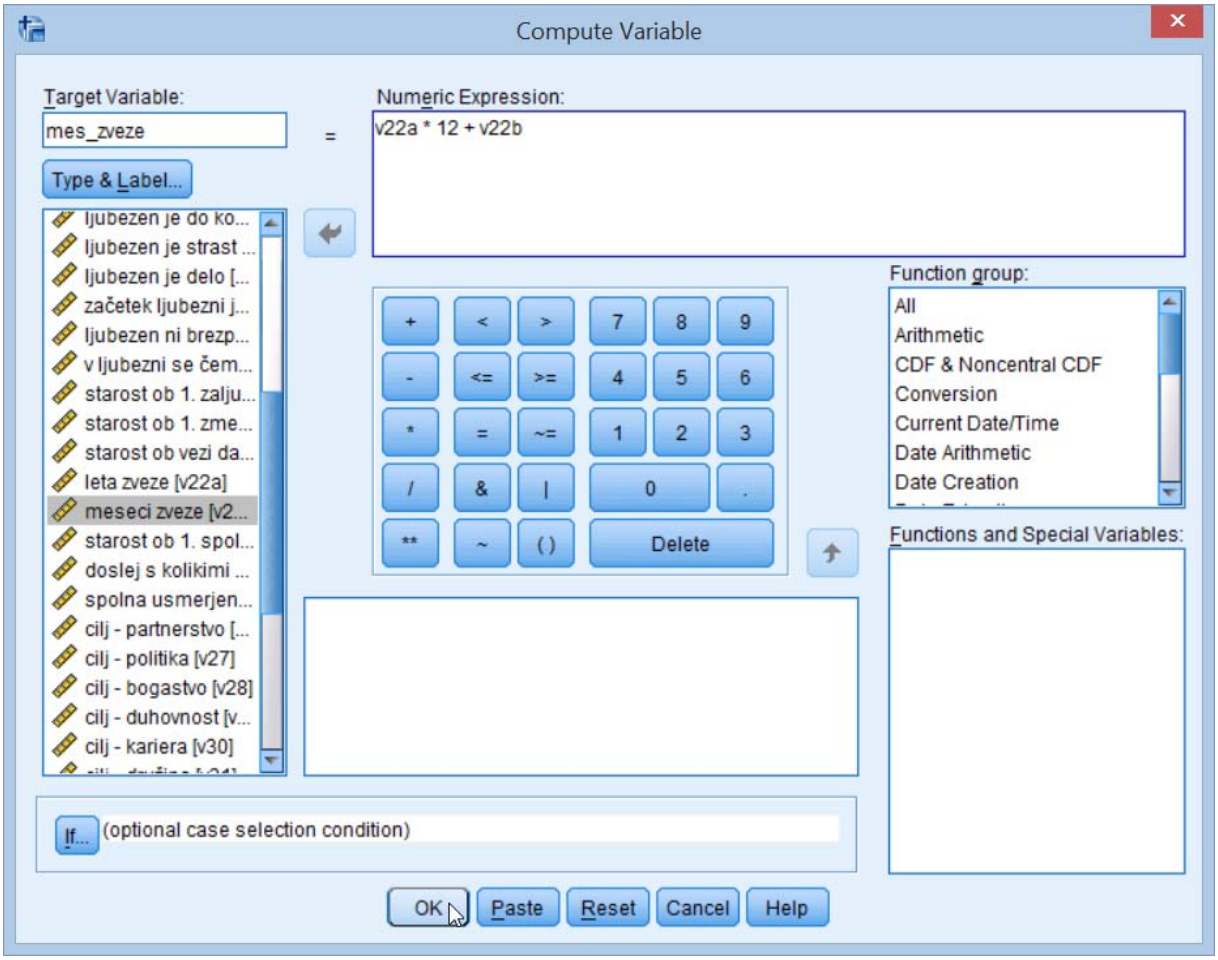


Ko končamo, kliknemo še Continue in OK. Nova spremenljivka ima tako 4 kategorije starosti ob poskusu marihuane: 1 – nikoli, 2 – 14 let ali manj, 3 – 15 do 16 let ter 4 – 17 let ali več. Tako kot smo na začetku določili Label, Value in Missing, tudi za te nove spremenljivke v pogovornem oknu Variable View določimo te značilnosti spremenljivk.

Če se sprašujete, ali funkcija Recode pomeni spreminjanje odgovorov posamezne osebe, je odgovor ne. S tem nič ne spreminjamo podatkov, samo zapišemo jih v drugačni obliki.

3. Kje najdem funkcijo Compute v SPSS-u?

Odprite pogovorno okno Transform → Compute. V okence Target Variable vnesete ime nove spremenljivke, ki jo boste izračunali, v okence Numeric Expression pa računski izraz za izračun nove spremenljivke.

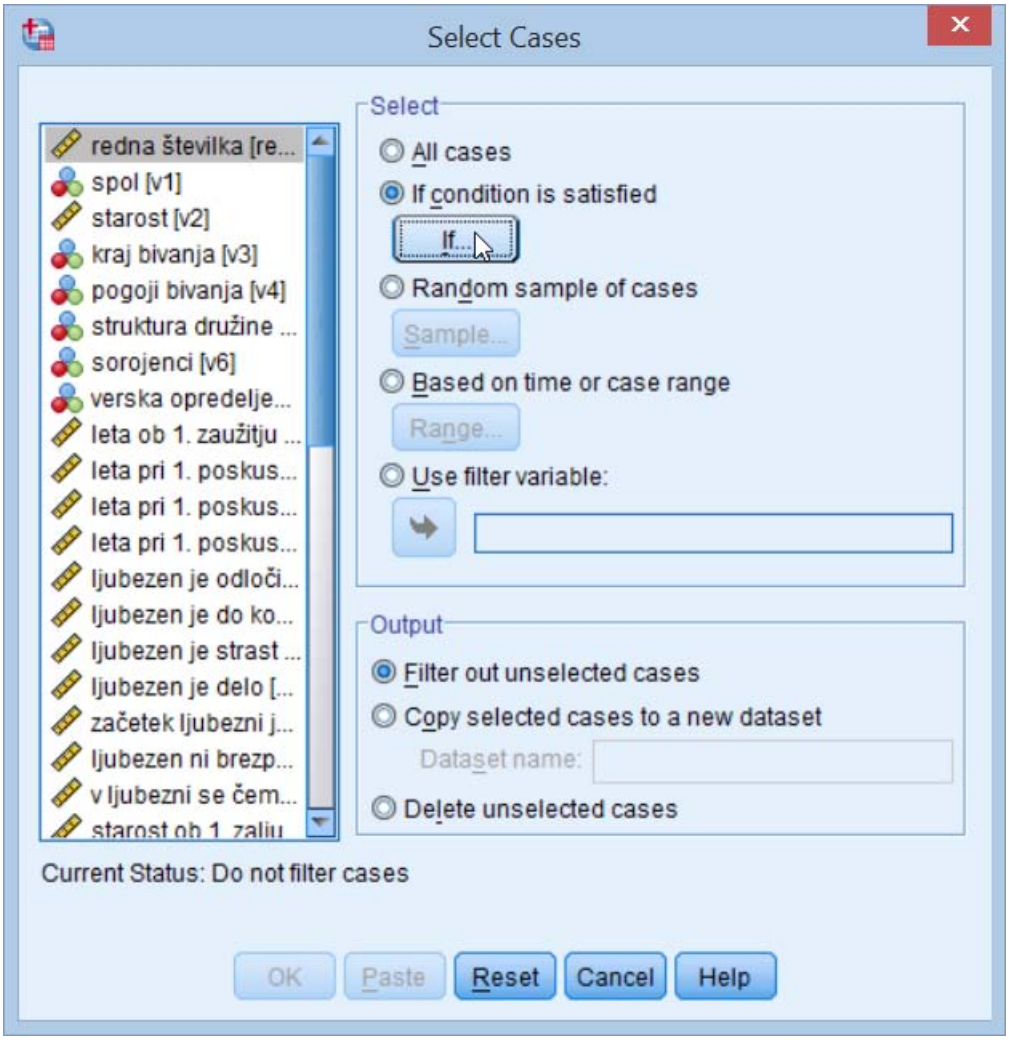


Zgornja slika prikazuje primer za vprašanje Trajanje moje najdaljše partnerske zveze, kjer je odgovor zapisan tako, da vprašani na črto zapišejo leta ter mesece, npr. 2 leti in 3 mesece. Sedaj pa bi radi trajanje najdaljše zveze vprašanih izrazili zgolj v mesecih. V tem primeru bi morali zapisati računski izraz za izračun nove spremenljivke, ki bi jo poimenovali Zveza v mesecih kot *leta x 12 mesecev + meseci*. Na zgornji sliki je v22a leta zveze, v22b pa meseci zveze.

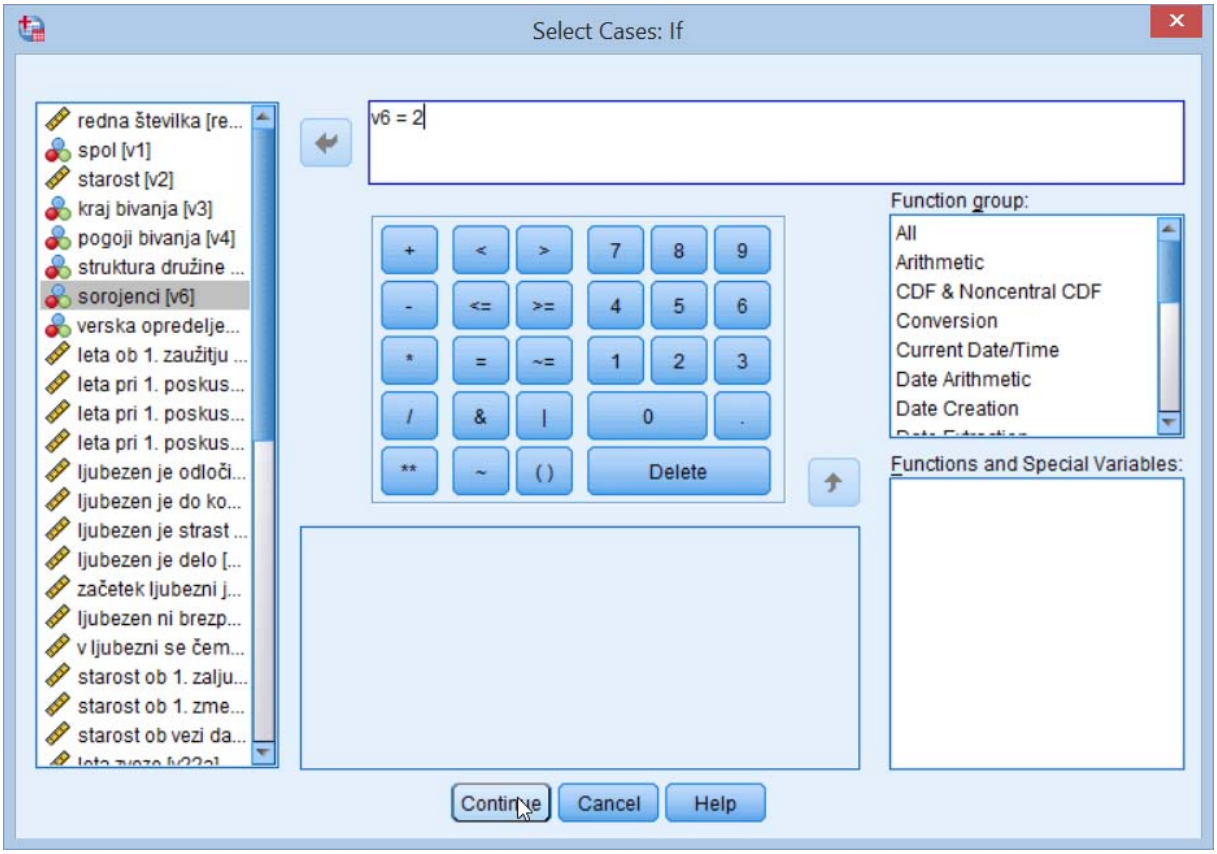
Kliknite še OK. Tako dobimo novo spremenljivko, ki meri zvezo v mesecih. Zopet ji v oknu Variable View določimo Label, Values in Missing.

3. Kje najdem funkcijo Select Cases v SPSS-u?

Odprite pogovorno okno Data → Select Cases (predzadnja po vrsti), kliknite na opcijo If condition is satisfied in potem na okence If.



Spodnja slika prikazuje primer, da želimo delati samo s podatki vprašanih, ki imajo 1 brata ali sestro. Uporabili bomo torej vprašanje Moji sorojenci, ki ima možne kategorije: 1 – sem edinka, 2 – imam 1 brata/sestro, 3 – imam 2 brata/sestri ter 4 - imam 3 ali več bratov/sester. Računski izraz za zgornji pogoj napišemo takole: v prazno okence prenesete spremenljivke in jih zapišete v računskem izrazu tako, da dobite želeni pogoj.



Kliknemo še Continue in OK. Ob tako postavljenem pogoju bomo delali samo na podatkih, ki smo jih dobili od vprašanih z 1 bratom ali sestro. Če želimo zopet delati na vseh podatkih, še enkrat odpremo pogovorno okno Data → Select Cases in kliknemo na opcijo All cases.

DESKRIPTIVNA STATISTIKA

1. Kaj je deskriptivna statistika?

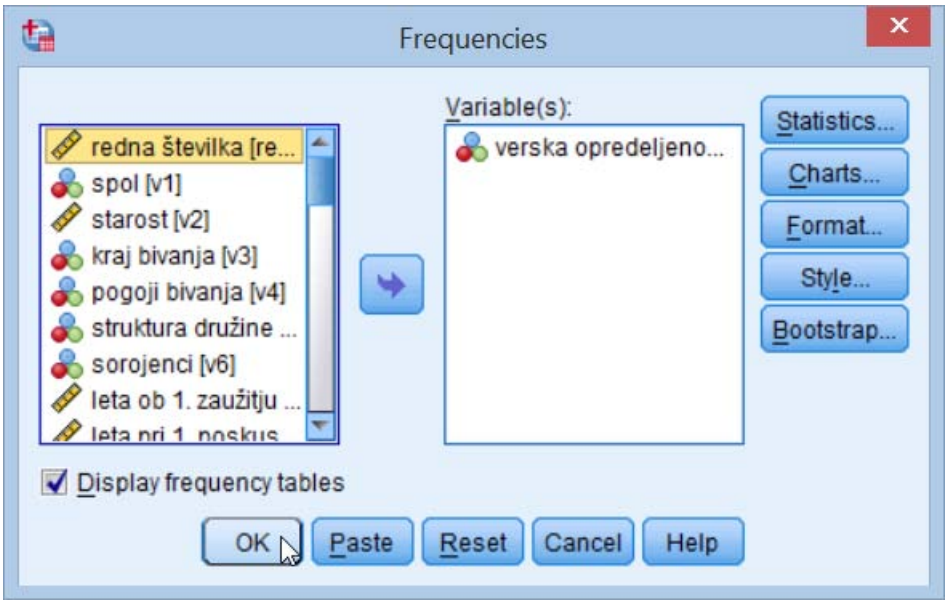
Deskriptivna statistika oz. opisna statistika je nabor metod, s katerimi opišemo značilnosti vzorca. Pri:

- nominalnih podatkih prikažemo frekvence ter odstotke;
- pri intervalnih in razmernostnih pa lahko poleg teh izračunamo osnovno statistiko (Ambrožič in Leskošek, 1999).

Te rezultate lahko prikažemo tudi ilustrativno (glej naslednje poglavje Ilustrativno prikazovanje podatkov).

2. Kje najdem izračun frekvenc v SPSS-u?

Odprite pogovorno okno Analyze → Descriptive Statistics→ Frequencies. V okence na desno prenesite želene spremenljivke. (Če želite poleg frekvenc izračunati še aritmetično sredino in ostale mere variabilnosti, odprite podokno Statistics in odkljukajte želene parametre; če želite grafe, kliknite na podokno Charts.)



Zgornja slika prikazuje primer za isti vzorec kot zgoraj tabelo frekvenčne distribucije za spremenljivko verska opredeljenost, ki vsebuje dve kategoriji: a) sem verna, b) sem neverna.

V tabeli 1 vidimo, da prikaže frekvence, odstotke, veljavne ter kumulativne odstotke. Ker nekateri posamezniki niso odgovorili na vprašanje, se stolpec Valid Percent razlikuje od stolpca Percent, saj stolpec Percent prikaže tudi manjkajoče vrednosti, stolpec Valid Percent pa zgolj odstotke dobljene z odgovori oseb, ki so odgovorili na vprašanje. Stolpec Cumulative Percent prikaže kumulativne odstotke (vsak naslednji odstotek prišteje k predhodnemu).

V tem gradivu bodo prikazane tabele, ki jih dobimo v Outputu SPSS-a v svoji osnovni obliki, torej tudi z angleškim zapisom parametrov, da vam bo lažje slediti pri delu z SPSS-om. Ko pa boste pisali seminarsko ali diplomsko nalogo ipd., morajo biti tabele v celoti poslovenjene (tj. namesto percent – odstotek)!!!

Tabela 1: Frekvenčna porazdelitev spremenljivke verska opredeljenost

verska opredeljenost

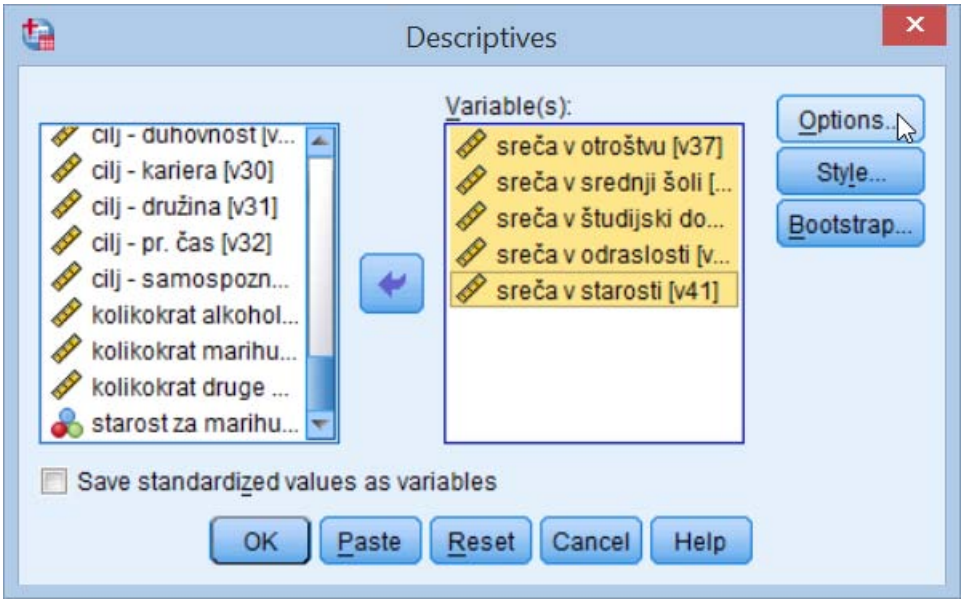
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	veren	30	60,0	63,8	63,8
	neveren	17	34,0	36,2	100,0
	Total	47	94,0	100,0	
Missing	-9	3	6,0		
Total		50	100,0		

3. Kako opišem tabelo Frequences?

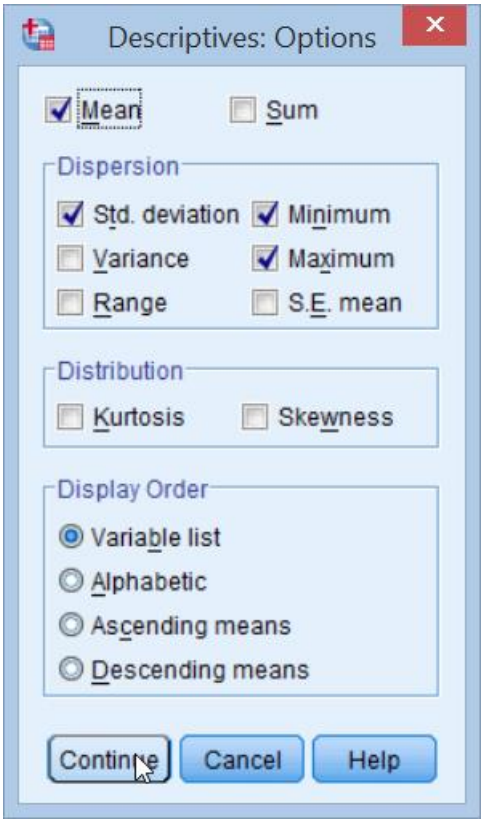
Zgornjo tabelo bi lahko opisali takole: 63,8% vprašanih, ki je odgovorilo na vprašanje (N = 47), se izreka za verne, 36,2% pa za neverne. Pazi: pri opisovanju tabel se moramo držati reda, da ves čas navajamo samo veljavne odstotke ali odstotke ali pa oboje, vendar to mora biti konsistentno, ne pa da bi pri opisu ene spremenljivke navedli odstotke, pri opisu druge pa veljavne odstotke (Hinton et al., 2004).

4. Kje najdem izračun mer srednje vrednosti in variabilnosti podatkov v SPSS-u?

Odprite pogovorno okno Analyze → Descriptive Statistics→ Descriptives. V okence na desno prenesite želene spremenljivke.



V podoknu Options odkljukajte želene parametre.



Zgornji sliki prikazujeta nastavitev za tabelo osnovne statistike za vprašanja o doživljanju sreče v posameznih življenjskih obdobjih kot so jo oz. predvidevajo, da jo bodo doživljale študentke in študenti istega vzorca kot zgoraj. Odgovori so bili merjeni s štiristopenjsko lestvico, kjer je 1 – zelo nesrečna, 2 – precej nesrečna, 3 – precej srečna ter 4 – zelo srečna. Tabela zgleda takole:

Tabela 2: Mere opisne statistike za spremenljivko doživljanje sreče

Descriptive Statistics

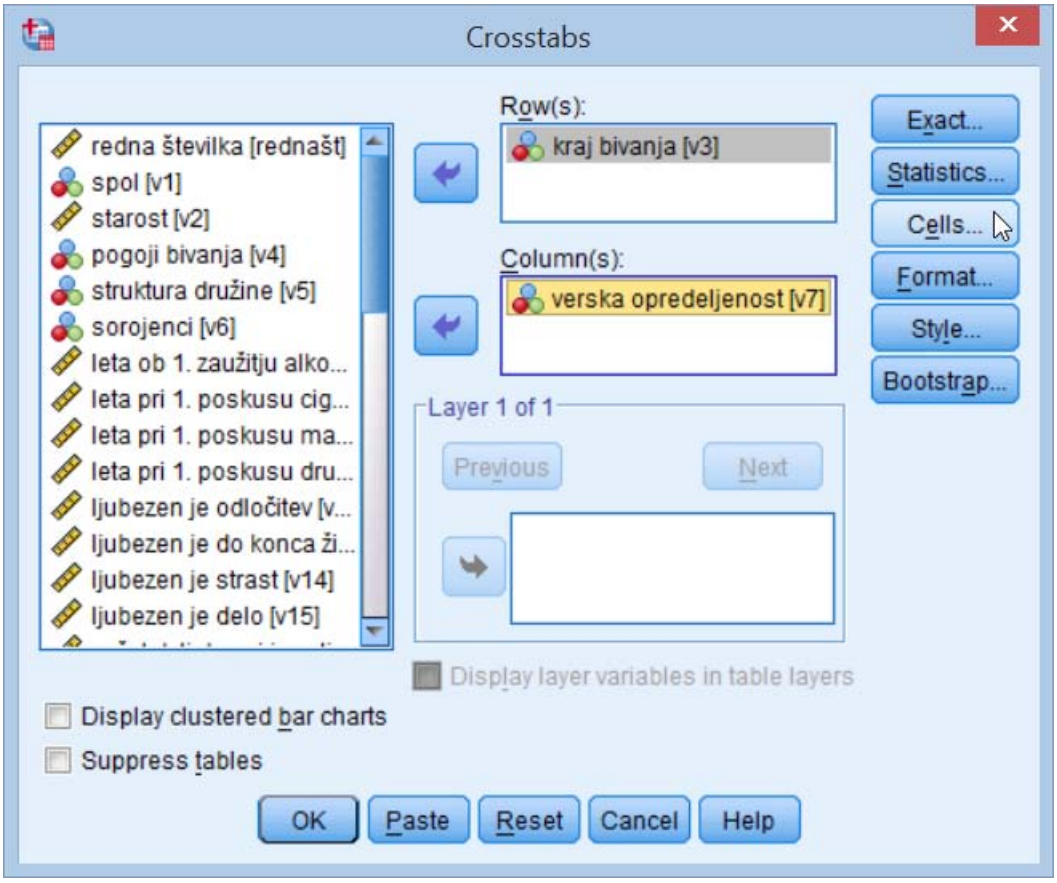
	N	Minimum	Maximum	Mean	Std. Deviation
sreča v otroštvu	50	1	4	3,20	,728
sreča v srednji šoli	50	1	4	2,66	,823
sreča v študijski dobi	50	2	4	3,28	,607
sreča v odraslosti	50	2	4	3,24	,476
sreča v starosti	50	1	4	3,14	,606
Valid N (listwise)	50				

5. Kako opišem podatke v tabeli Descriptive Statistics?

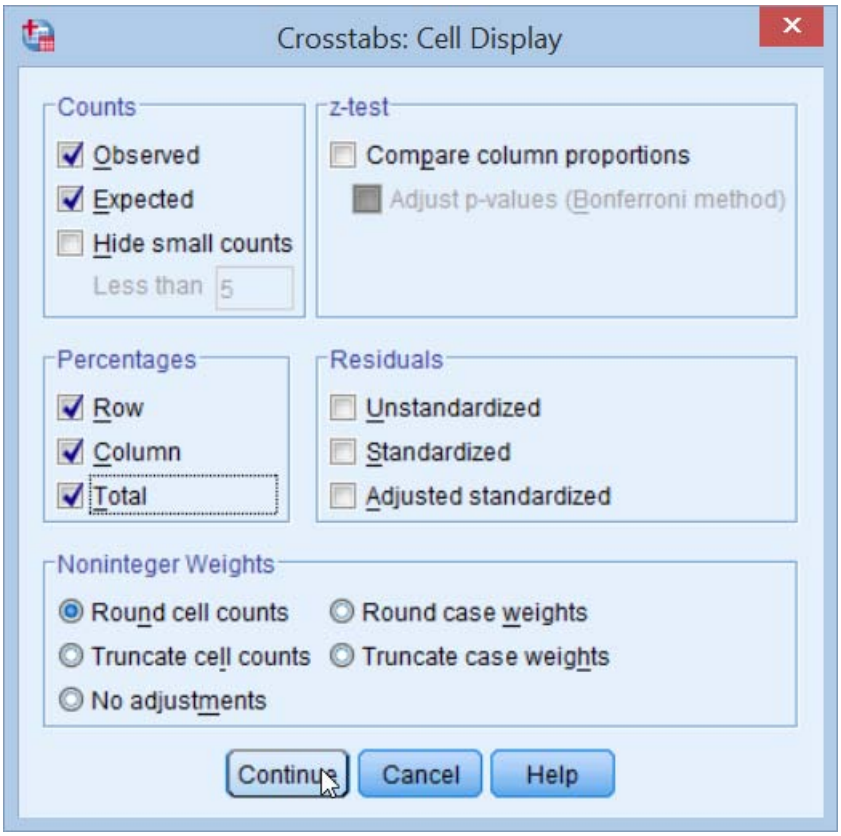
Zgornjo tabelo bi lahko opisali takole:
Iz zgornje tabele lahko razberemo, da so na vprašanja o sreči odgovorile vse študentke ter da je razen pri vprašanjih o sreči v študijski dobi in odraslosti minimalna vrednost izbranega odgovora znašala 1, maksimalna pa pri vseh vprašanjih 4. Samo pri vprašanjih o sreči v študijski dobi in odraslosti je minimalna vrednost izbranih odgovorov znašala 2, torej nobena študentka za ti dve obdobji ne ocenjuje, da je/bo zelo nesrečna. Najvišja srednja vrednost ocene sreče nastopa pri vprašanju sreče v študijski dobi in sicer znaša 3,28, torej se študentke v splošnem doživljajo trenutno kot precej srečne. Nekoliko manjši sta srednji vrednosti ocen sreče v odraslosti ter otroštvu. V povprečju pa študentke ocenjujejo, da so bile najmanj srečne v obdobju srednje šole, aritmetična sredina v tem primeru znaša 2,66 s standardnim odklonom .823. Odgovori na vprašanje o doživljanju sreče v obdobju srednje šole, so bili torej najbolj razpršeni (so najbolj variirali), najmanj pa so bili razpršeni pri vprašanju sreče v odraslosti in sicer .476.

6. Kje najdem izračun frekvenc kategorij ene spremenljivke po kategorijah druge spremenljivke v SPSS-u?

Odprite pogovorno okno Analyze → Descriptive Statistics→ Crosstabs. V okence Row prenesete neodvisno spremenljivko, v okence Column pa odvisno. Nato kliknite okence Cells.



V okencu Cells odkljukajte še pri Counts: Observed ter Expected in pri Percentages še: Row, Column, Total.



Zgornja slika kaže nastavitev za kontingenčno tabelo za dve vprašanji; kraj bivanja, ki je imelo dve kategoriji: vas ali manjše mesto in večje mesto ter moja verska opredeljenost, tudi dve kategoriji: sem verna, nisem verna.

Tabela 3 z naslovom Case Processing Summary kaže numerus ter frekvenčne odstotke za obe kategoriji skupaj. Vidimo, da je bilo vseh vprašanih študentk 50, od tega imamo podatke za 47 študentk, kar je 94% ter da tri študentke na eno ali drugo ali obe vprašanji niso odgovorile, kar je 6%. Te tabele ponavadi ne opisujemo. Tu je opis naveden zgolj zaradi ilustracije.

Tabela 3: Povzetek značilnosti statistične analize

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
kraj bivanja novi * verska opredeljenost	47	94,0%	3	6,0%	50	100,0%

Tabela 4 z naslovom kraj bivanja * verska opredeljenost Crosstabulation je kontingenčna tabela, ki kaže:

- empirične frekvence (Count),
- pričakovane frekvence (Expected Count) – pomembne za izračun χ^2 – testa (glej poglavje Uporaba χ^2 – testa)
- frekvenčne odstotke verske opredeljenosti glede na kraja bivanja (% within kraj bivanja),
- frekvenčne odstotke pripadanju določenemu kraju bivanja glede na versko opredeljenost (% within verska opredeljenost)
- frekvenčne odstotke celot (Total).

Kako torej beremo tako tabelo?
Poglejmo si najprej vrstico vas ali manjše mesto.
Vidimo, da je vseh vprašanih, ki živijo na vasi ali v manjšem mestu 29 (vrstica Count, stolpec Total), kar je 61,7% vseh vprašanih (vrstica % of Total, stolpec Total). Od vseh 29-ih, ki živijo na vasi ali v manjšem mestu, je 18 vernih (vrstica Count, stolpec veren), kar je 62,1 % vseh, ki živijo na vasi ali v manjšem mestu (vrstica % within kraj bivanja, stolpec veren) ter 11 nevernih (vrstica Count, stolpec neveren), kar je 37,9 % vseh, ki živijo na vasi ali v manjšem mestu.
Od vseh vernih je 60% teh, ki živijo na vasi oz. v manjšem mestu (vrstica % within verska opredeljenost, stolpec veren). Od vseh nevernih, je 64,7% teh, ki živijo na vasi oz. v manjšem mestu (vrstica % within verska opredeljenost, stolpec neveren).
Od vseh vprašanih je 38,3 % takih, ki živijo na vasi ali v manjšem mestu in so verni (vrstica % of Total, stolpec veren) ter 23,4 % takih, ki živijo na vasi ali v manjšem mestu in niso verni (vrstica % of Total, stolpec neveren).
Po enakem ključu beremo naslednjo vrstico večje mesto.
Vidimo, da je vseh vprašanih, ki živijo v večjem mestu 18, kar je 38,3 % vseh vprašanih. Od vseh 18-ih, ki živijo v večjem mestu, je 12 vernih, kar je 66,7 % vseh, ki živijo v večjem mestu ter 6 nevernih, kar je 33,3 % vseh, ki živijo v večjem mestu.

Od vseh vernih je 40 % teh, ki živijo v večjem mestu. Od vseh nevernih je 35,3 % teh, ki živijo v večjem mestu.
Od vseh vprašanih je 25,5 % takih, ki živijo v večjem mestu in so verni ter 12,8 % takih, ki živijo v večjem mestu in niso verni.
In sedaj še vrstica Total.
Vseh vprašanih je bilo 47 (vrstica Count, stolpec Total), od tega 30 vernih (vrstica Count, stolpec veren), kar je 63,8 % (vrstica % of Total, stolpec veren) vseh vprašanih in 17 nevernih (vrstica Count, stolpec neveren), kar je 36,2 % vseh vprašanih (vrstica % of Total, stolpec neveren).

Tabela 4: Kontingenčna tabela za kraj bivanja anketiranih glede na njihovo versko opredeljenost

kraj bivanja * verska opredeljenost Crosstabulation

			verska opredeljenost		Total
			veren	neveren	
kraj bivanja	vas ali manjše mesto	Count	18	11	29
		Expected Count	18,5	10,5	29,0
		% within kraj bivanja	62,1%	37,9%	100,0%
		% within verska opredeljenost	60,0%	64,7%	61,7%
		% of Total	38,3%	23,4%	61,7%
	večje mesto	Count	12	6	18
		Expected Count	11,5	6,5	18,0
		% within kraj bivanja	66,7%	33,3%	100,0%
		% within verska opredeljenost	40,0%	35,3%	38,3%
		% of Total	25,5%	12,8%	38,3%
Total		Count	30	17	47
		Expected Count	30,0	17,0	47,0
		% within kraj bivanja	63,8%	36,2%	100,0%
		% within verska opredeljenost	100,0%	100,0%	100,0%
		% of Total	63,8%	36,2%	100,0%

7. Kako opišem tabelo Crosstabulation?

Pri opisu kontingenčne tabele seveda ne pišemo tako obširne razlage kot zgoraj, ta je navedena zgolj zaradi lažjega razumevanja. Za naš primer bi bila navedba lahko taka:
Iz tabele razberemo, da je bilo vseh vprašanih 47, od tega 63,8 % vernih 36,2 % nevernih. Od tistih, ki živijo na vasi ali v manjšem mestu (ti predstavljajo 61,7% celotnega vzorca), je vernih 62,1 % ter 37,9 % nevernih. Vprašani, ki živijo v večjem mestu, predstavljajo 38,3 % celotnega vzorca. Od teh je 66,7 % vernih ter 33,3 % nevernih.

ILUSTRATIVNO PRIKAZOVANJE PODATKOV

1. Kdaj naj se odločim za ilustrativno prikazovanje podatkov?

Ilustrativno prikazovanje podatkov najpogosteje uporabljamo pri poglavju opis vzorca oz. opis osnovnih značilnosti odgovorov, kjer skozi deskriptivno statistiko pokažemo osnovne lastnosti vzorca oz. neodvisne spremenljivke. Ambrožič in Leskošek (1999) in Field (2005) pravijo, da pri:
a) Nominalnih podatkih (včasih tudi pri ordinalnih in pogojno intervalnih) prikažemo frekvence ter odstotke v strukturnem grafikonu (krog oz. pita – Pie Charts). Frekvence lahko prikažemo tudi v stolpčnem grafikonu (Bar Charts);

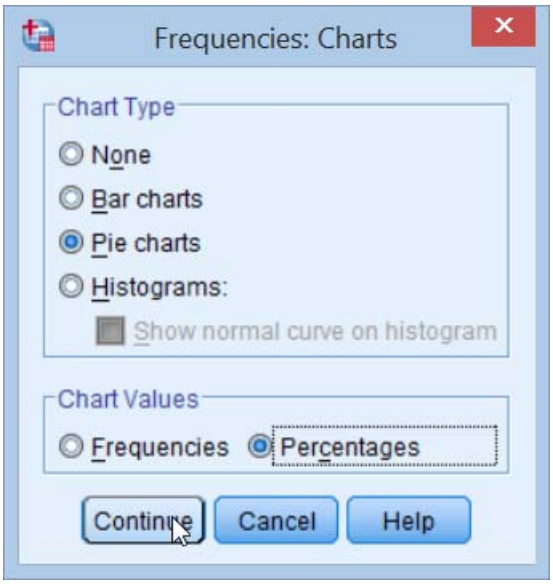
b) Pri intervalnih in razmernostnih pa izračunamo osnovno statistiko ter jih prikažemo s histogramom. Uporabimo pa lahko tudi stolpčni ali črtni grafikon.

- Histogram prikaže frekvence dobljenih točk spremenljivke in je uporaben za preverbo distribucije spremenljivke oz. točk.
- Stolpčni grafikon (Bar Charts) se običajno uporablja za prikaz aritmetičnih sredin različnih skupin ljudi (Summaries for groups of cases) ali za prikaz aritmetičnih sredin različnih spremenljivk (Summaries of separate variables).
- Črtni grafikon (Line) se prav tako uporablja za prikaz aritmetičnih sredin.

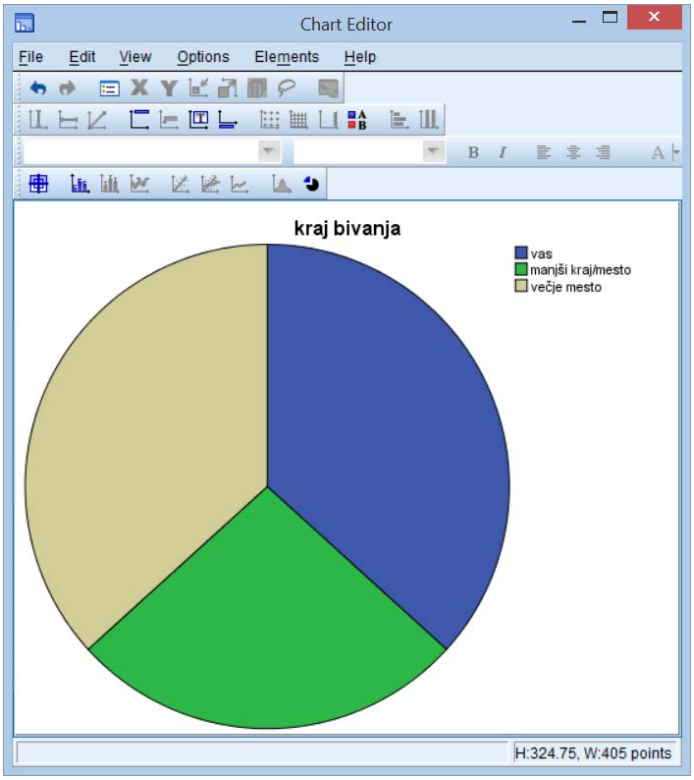
2. Kako najdem strukturni grafikon oz. pito v SPSS-u?

Odprite pogovorno okno Analyze → Descriptive Statistics → Frequences (druga možnost pa je: Graphs → Legacy Dialogs → Pie)

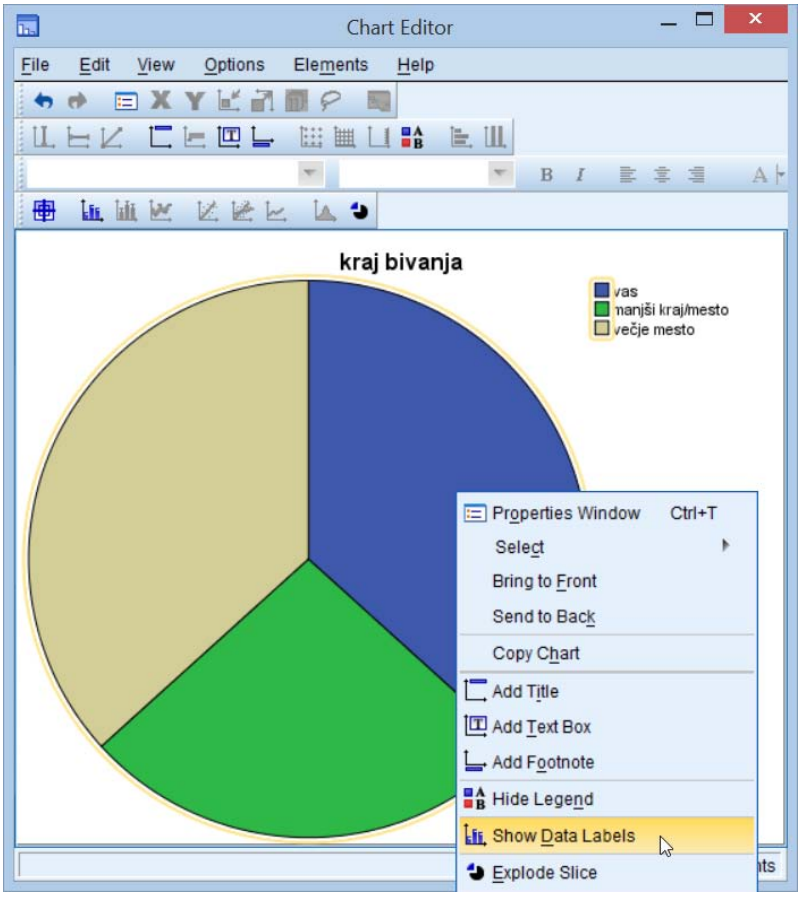
V odprtem oknu Frequences kliknite okence Charts in ko se vam odpre, kliknite Pie charts (in izberite Percentages – če želite prikazati odstotke in ne frekvenc). Druga možnost je, da odprete pogovorno okno Graphs → Legacy Dialogs → Pie, odpre se okno, v katerem pustite Summaries for groups of Cases, nato pa v oknu Define Pie vnesete spremenljivko).



Zgornja slika prikazuje nastavitev za strukturni grafikon oz. pito za spremenljivko kraj bivanja. Če strukturnega kroga nič ne uredite, potem vam SPSS izriše zgolj strukturni krog z legendo. Če pa želite, da bo na vsakem deležu strukturnega kroga zapisan odstotek, potem morate v SPSS-u v Outputu z dvoklikom klikniti na strukturni krog. Odpre se vam pogovorno okno Chart Editor.



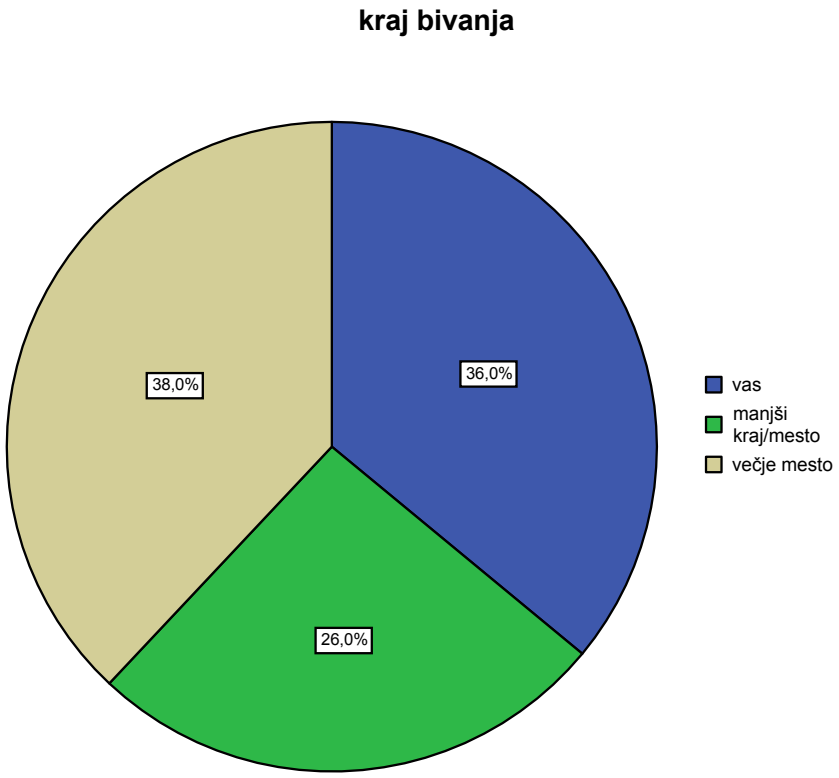
Z klikom na desni gumb miške na katerega koli od kosov na piti se vam odpre spustni meni, na katerem izberite Show data Labels.



Odpre se vam pogovorno okno Properties/ Data Value Labels. Ker je že nastavljena možnost Percent, kliknete le še Close. Zdaj bo vaš strukturni krog dopolnjen s temi informacijami.

Zaprite okno Chart Editor. Strukturni krog s prikazom odstotkov za spremenljivko kraj bivanja zgleda takole.

Slika 1: Strukturni krog s prikazom odstotkov za spremenljivko kraj bivanja



Mimogrede še nasvet: ko skopirate sliko iz SPSS-a v Word, kliknite nanjo z desnim gumbom in izberite funkcijo Obreži (Crop from). Tako boste odrezali »bele dele« slike. To velja za vse ilustrativne prikaze podatkov.

3. Kako naj opišem strukturni krog oz. pito?

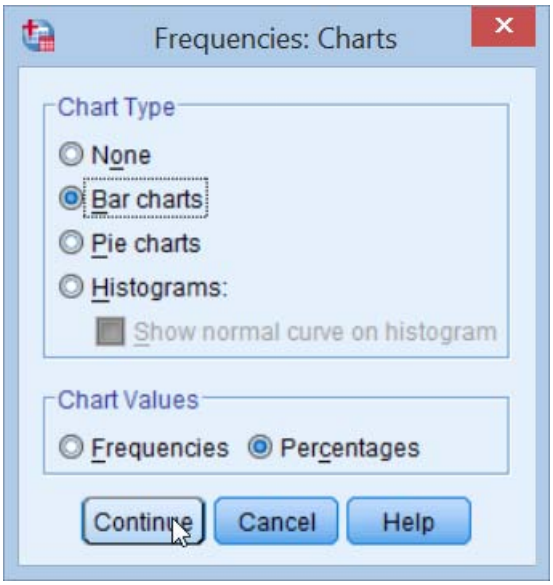
V tem primeru bi lahko naredili takole:
Slika 1 prikazuje strukturni krog za naš vzorec za spremenljivko kraj bivanja. Iz slike vidimo, da predstavlja delež študentk, ki živijo na vasi, 36%, delež študentk, ki živijo v manjšem kraju oz. mestu, 26% ter delež študentk, ki živijo v večjem mestu, 38%.

4. Kako najdem stolpčni grafikon v SPSS-u?

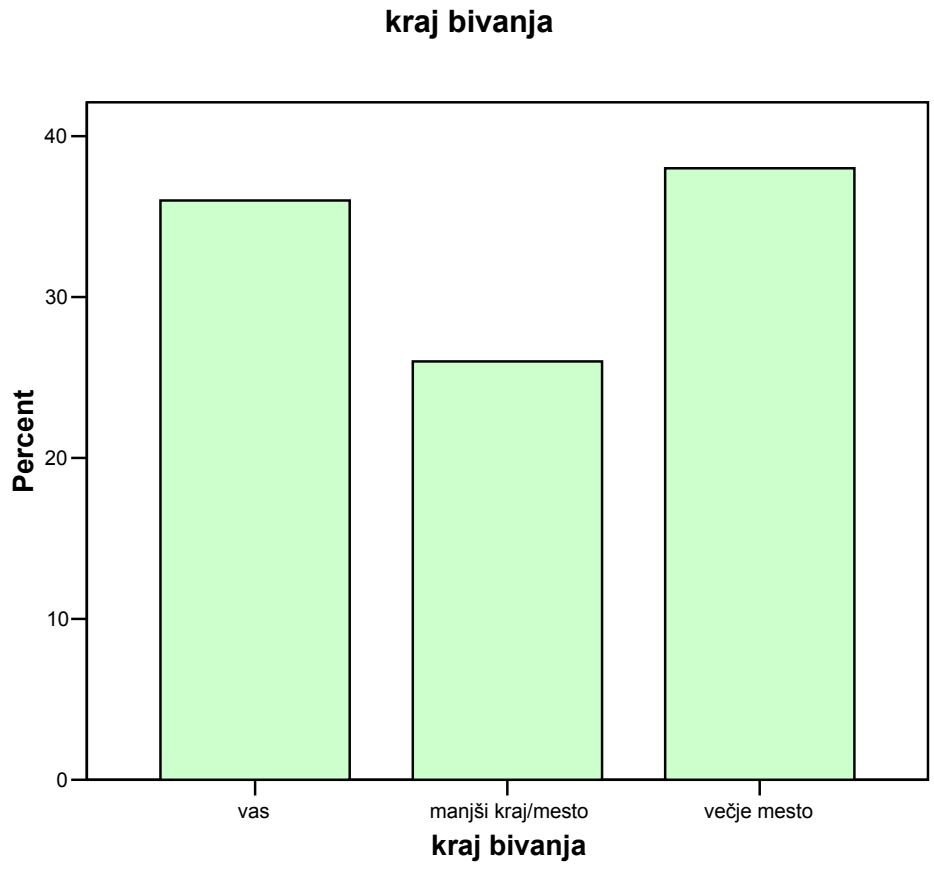
4.1 Stolpčni grafikon za prikaz frekvenc

Odprite pogovorno okno Analyze → Descriptive Statistics → Frequences. V odprtem oknu Frequences kliknite okence Charts (glej zgoraj pri strukturnem krogu) in ko se vam odpre, kliknite Bar charts (in izberite Percentages – če želite prikazati odstotke in ne frekvenc). Druga možnost je, da odprete pogovorno okno Graphs → Legacy Dialogs → Bar (in potem Simple, nato Define in nato v oknu Define prenesete v Category Axis spremenljivko ter

izberete % of cases). Če želite grafikon še nekoliko urediti, enako kot pri piti odprete pogovorno okno Chart Editor z dvoklikom na grafikon.



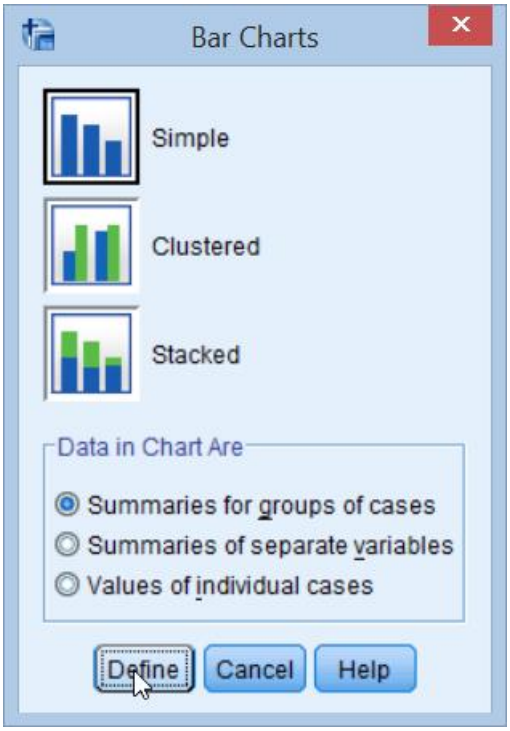
Oglejmo si stolpčni grafikon za enako spremenljivko kot zgoraj (tudi tu sliko enako oblikujete v Wordu).



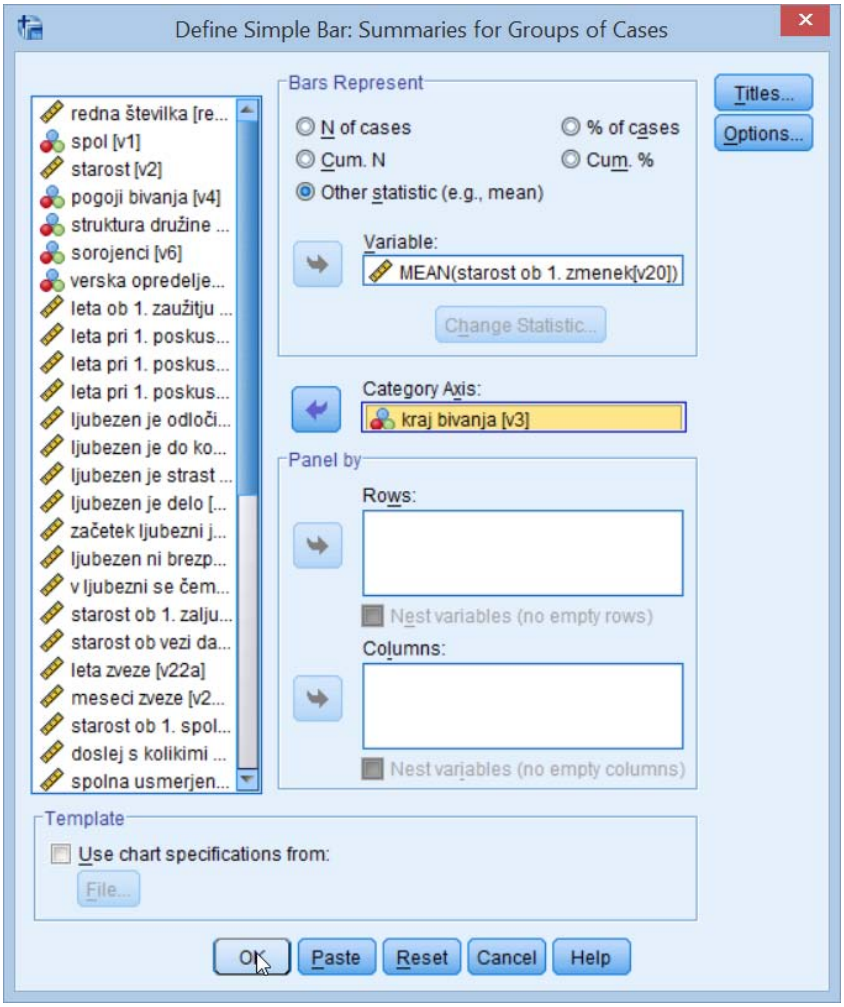
Grafikon opišemo podobno kot pito zgoraj.

4.2 Stolpčni grafikon za prikaz aritmetičnih sredin po skupinah

Odprite pogovorno okno Graphs → Legacy Dialogs → Bar. Izbrana opcija je Simple, kliknite na Define.

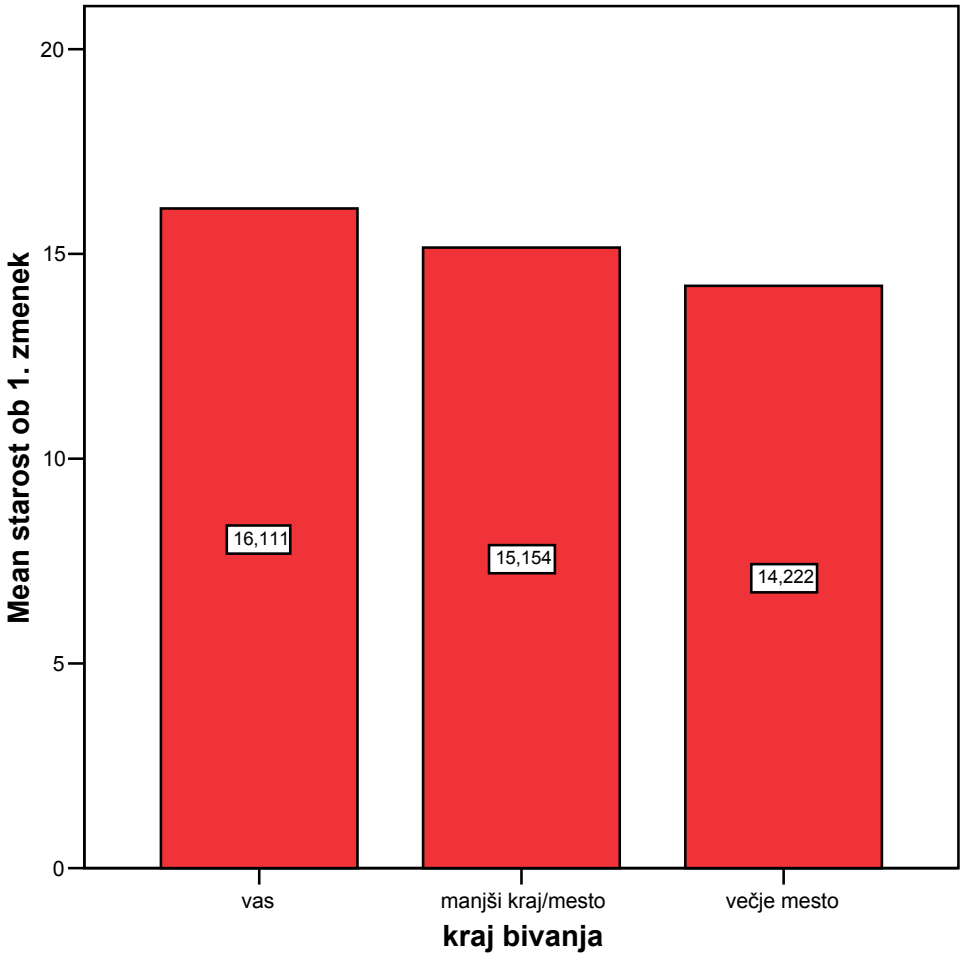


Nato se vam odpre okno Define Simple Bar. V Category Axis prenesete spremenljivko, za katero bi radi prikazali aritmetične sredine po skupinah, ter v funkciji Bar represent izberite funkcijo Other Summary function. Sprosti se vam okence Variable in tja vnesete odvisno spremenljivko. Nato kliknite še ok. Če želite grafikon še nekoliko urediti, enako kot zgoraj odprete pogovorno okno Chart Editor z dvoklikom na grafikon.



Na zgornji sliki so nastavljene opcije za stolpčni grafikon za prikaz aritmetičnih sredin starosti ob prvem zmenku po skupinah glede na kraj bivanja (neodvisna spremenljivka – Category Axis: kraj bivanja, odvisna – Variable: starost ob prvem zmenku) na vzorcu enakem kot zgoraj. Dobljeni stolpčni grafikon zgleda takole.

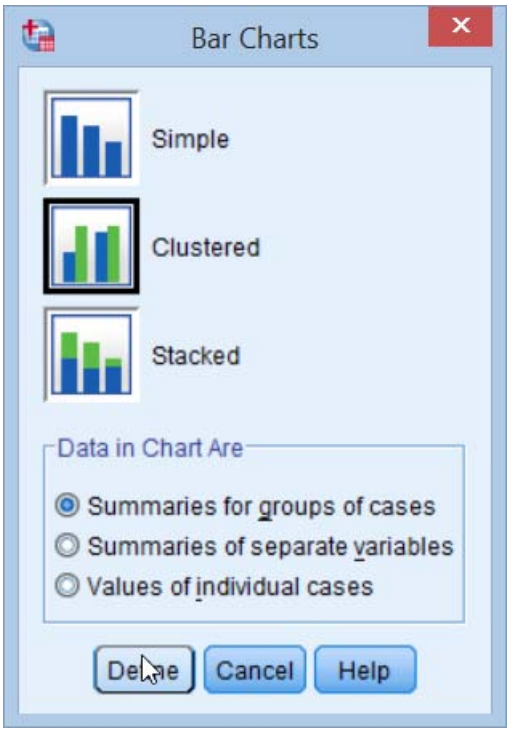
Slika 3: Stolpčni grafikon za spremenljivko starost ob prvem zmenku glede na kraj bivanja



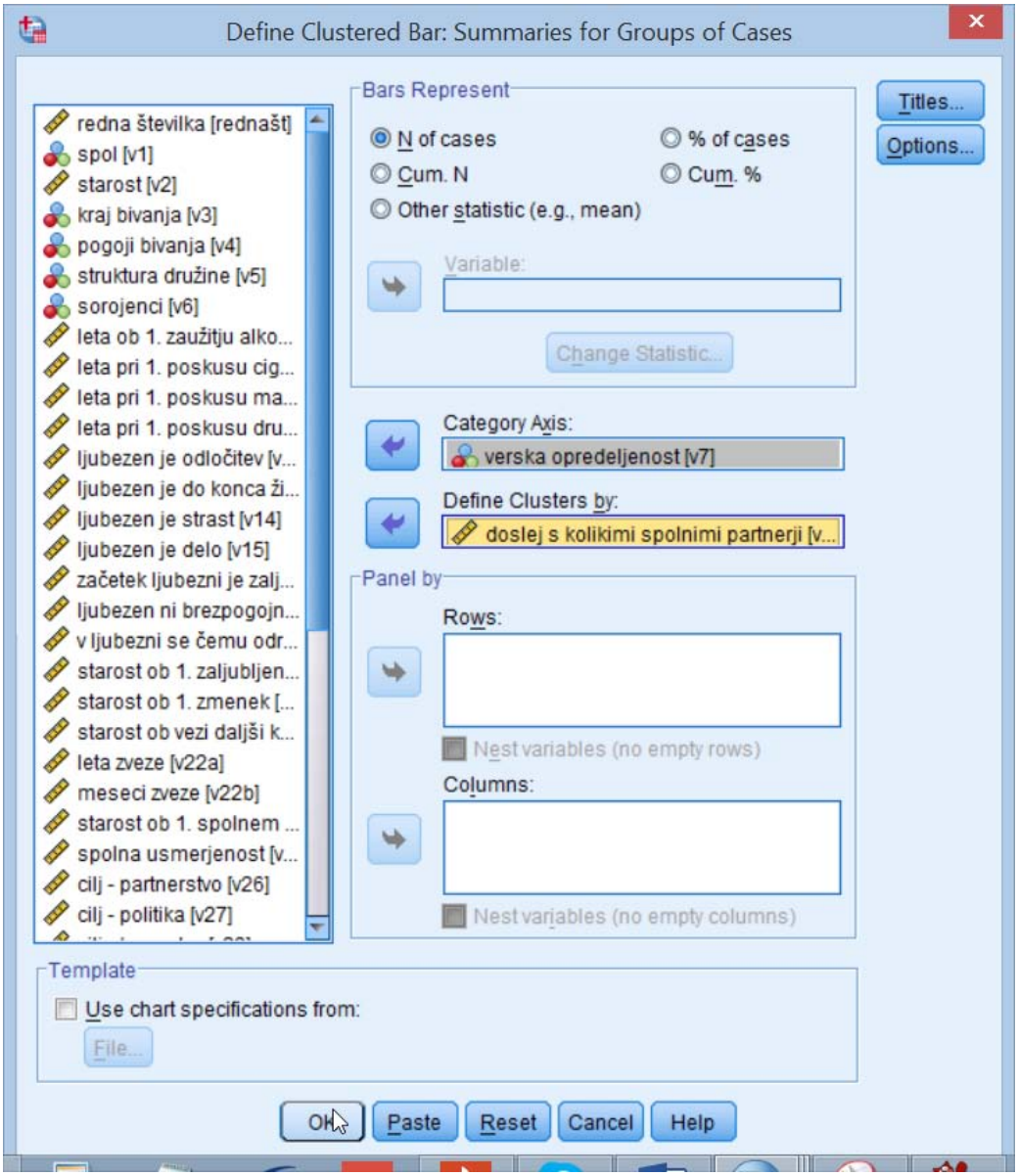
Ta grafikon bi lahko opisali takole:
Na sliki vidimo, da je povprečna starost pri prvem zmenku najvišja pri dekletih, ki živijo na vasi in sicer 16,11 let, nekoliko nižja je povprečna starost pri dekletih, ki živijo v manjšem kraju oz. mestu. Ta znaša 15,15 let. Najnižja pa je pri dekletih iz večjega mesta in sicer 14,22. Za naš vzorec lahko rečemo, da dekleta iz vasi v povprečju najkasneje dobijo izkušnjo zmenka, dekleta iz večjega mesta pa najprej. (Pazi: če bi to želeli preveriti, če to velja za vse študentke socialne pedagogike v Sloveniji, potem bi morali preveriti hipotezo. V tem primeru bi to naredili s testom ANOVA.

4.3 Stolpčni grafikon za prikaz kategorij ene spremenljivke glede na kategorije druge spremenljivke

Odprite pogovorno okno Graphs → Legacy Dialogs → Bar → Clustered ter Define.

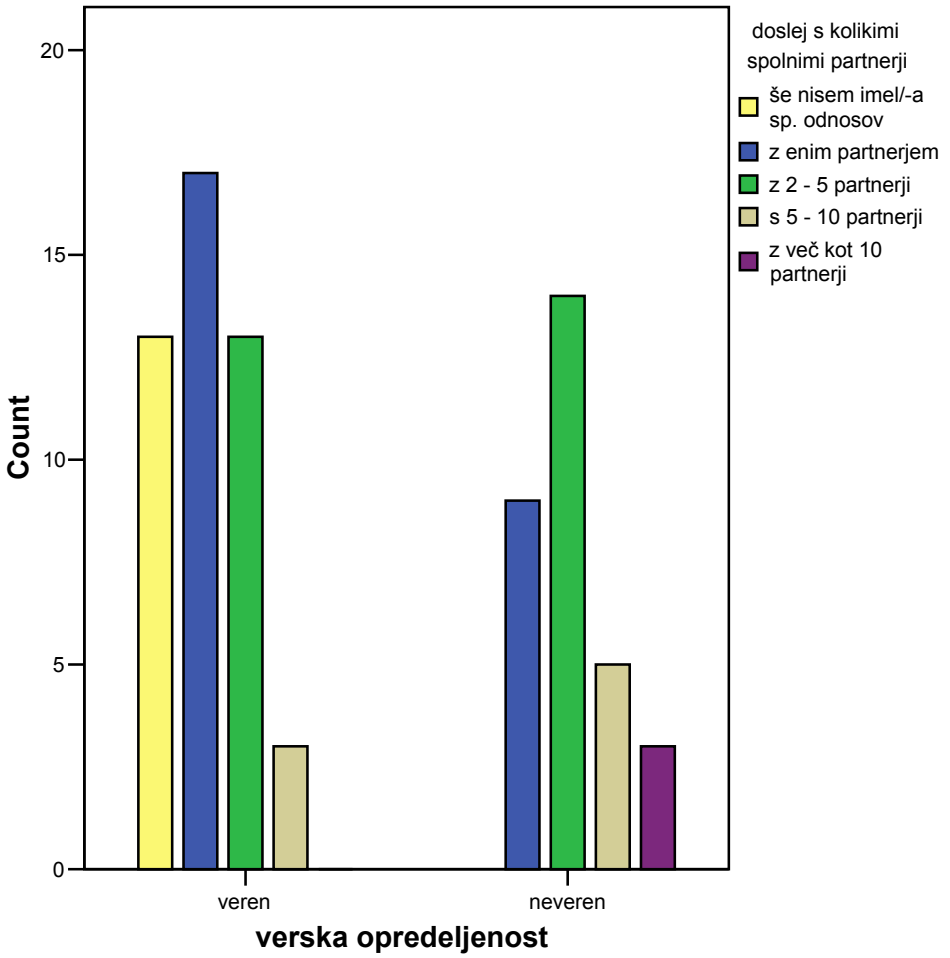


Nato se vam odpre okno Define Clustered Bar. V Category Axis prenesete spremenljivko, po kateri boste opazovali kategorije druge spremenljivke, ki jo vnesete v okence Define clusters by, nato kliknite še ok. Če želite grafikon še nekoliko urediti, enako kot zgoraj odprete pogovorno okno Chart Editor z dvoklikom na grafikon.



Zgornja slika prikazuje nastavitev za stolpčni grafikon za frekvenčni prikaz izkušenj z različnim številom spolnih partnerjev glede na vernost oz. nevernost študentk (neodvisna spremenljivka – Category Axis: veren, neveren, Define clusters by: število spolnih partnerjev doslej) na vzorcu enakem kot zgoraj.
Mimogrede: seveda tudi tu lahko uporabimo Other Summary function in namesto frekvenc naredimo primerjavo po aritmetičnih sredinah neke tretje spremenljivke.
Naš grafikon zgleda takole.

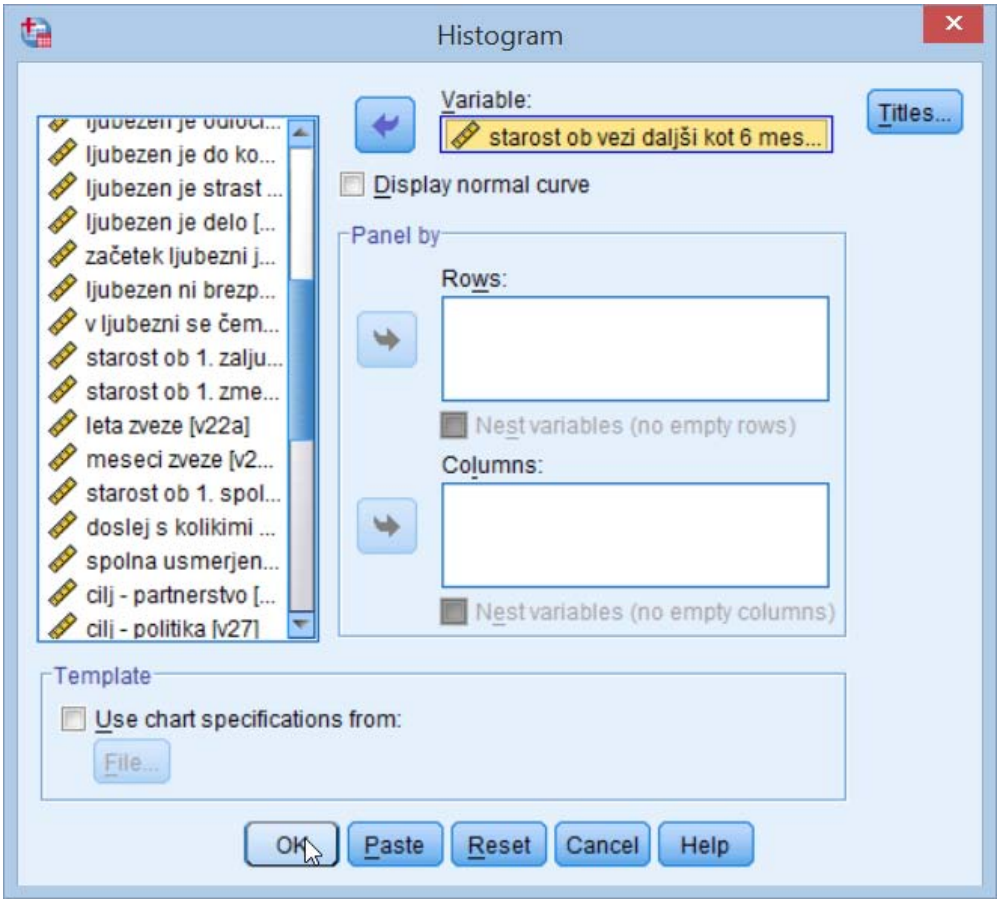
Slika 4: Stolpčni grafikon za frekvenčni prikaz izkušenj z različnim številom spolnih partnerjev glede na versko opredelitev študentk



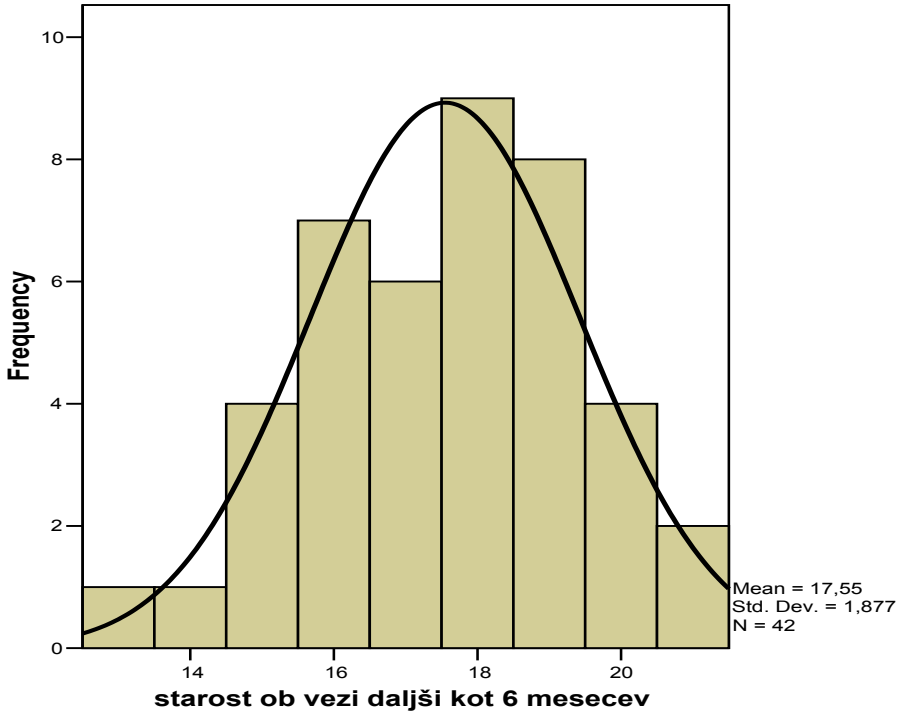
Ta grafikon bi lahko opisali takole:
Na sliki 4 vidimo, da se vse študentke (13), ki še niso imele spolnih odnosov, opredeljujejo kot verne. Več vernih (17) kot nevernih (9) študentk je imelo/ima izkušnje zgolj z enim spolnim partnerjem. Več nevernih (14) kot vernih (13) študentk je imelo/ima spolne izkušnje z 2 do 5 partnerji, prav tako več nevernih (5) kot vernih (3) je imelo/ima izkušnje s 5 do 10 partnerji. Vse študentke (3), ki so imele/imajo izkušnje z več kot 10 partnerji, pa se opredeljujejo kot neverne.

5. Kako najdem histogram v SPSS-u?

Odprite pogovorno okno Graphs → Legacy Dialogs → Histograms. V okence Variable prenesite želeno spremenljivko. Če želite še izrisano normalno krivuljo, odkljukajte Display normal curve in kliknite OK. Če želite grafikon še nekoliko urediti, enako kot zgoraj odprete pogovorno okno Chart Editor z dvoklikom na grafikon (druga možnost je, da odprete pogovorno okno Analyze → Descriptive Statistics → Explore in potem podokence Plots ter odkljukate Histogram). Spodnja slika prikazuje nastavitev za stolpčni grafikon za spremenljivko starosti, ko so imele študentke partnersko vezo daljšo od 6 mesecev.



Slika 5: Stolpčni grafikon za spremenljivko starosti, v kateri so imele študentke partnersko vezo daljšo od 6 mesecev



Na sliki 5 vidimo, da je v povprečju bila starost študentk, ko so pričele partnersko zvezo, ki je trajala dlje od 6 mesecev 17,55 let, najmlajša študentka je bila stara 13 let, najstarejša pa 21 let.

Več o ilustrativnem prikazu podatkov:

Ambrožič, F., Leskošek, B. (1999). Uvod v SPSS (verzija 8.0 za Windows 95/NT). Ljubljana: Fakulteta za šport: Inštitut za kineziologijo. Str. 43 – 47.

Field, A. (2005). Discovering Statistics Using SPSS. London: Sage Publications. Str. 65-76.

Hinton, P. R., Brownlow, C., McMurray, I. & Cozens, B. (2004). SPSS Explained. London& New York: Routledge. Str. 41 – 72.

ZANESLJIVOST VPRAŠALNIKA

1. Kaj pomeni zanesljivost vprašalnika?

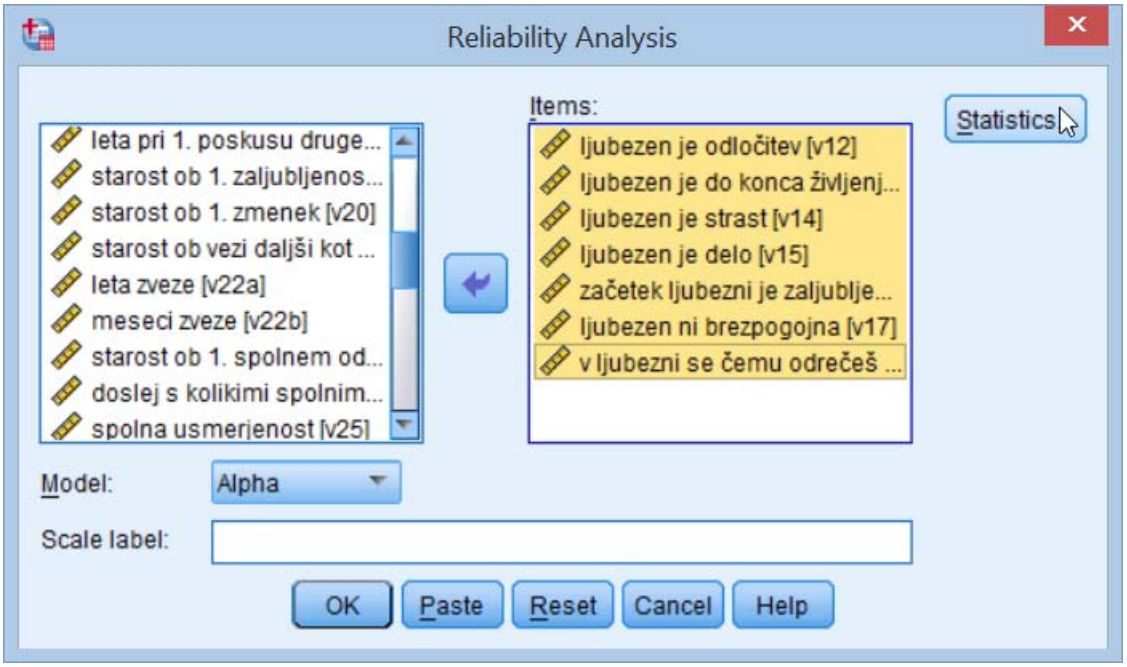
Kot smo rekli že zgoraj je zanesljivost vprašalnika lastnost vprašalnika, da daje pri ponovljenih merjenjih istih lastnosti pri istih osebah enake rezultate. Ali kot razlaga Field (2005), če bi posameznik reševal vprašalnik v različnih časovnih obdobjih, vsi ostali pogoji pa bi bili enaki, bi moral dobiti enake rezultate.

2. Kako vem, če je moj vprašalnik dovolj zanesljiv?

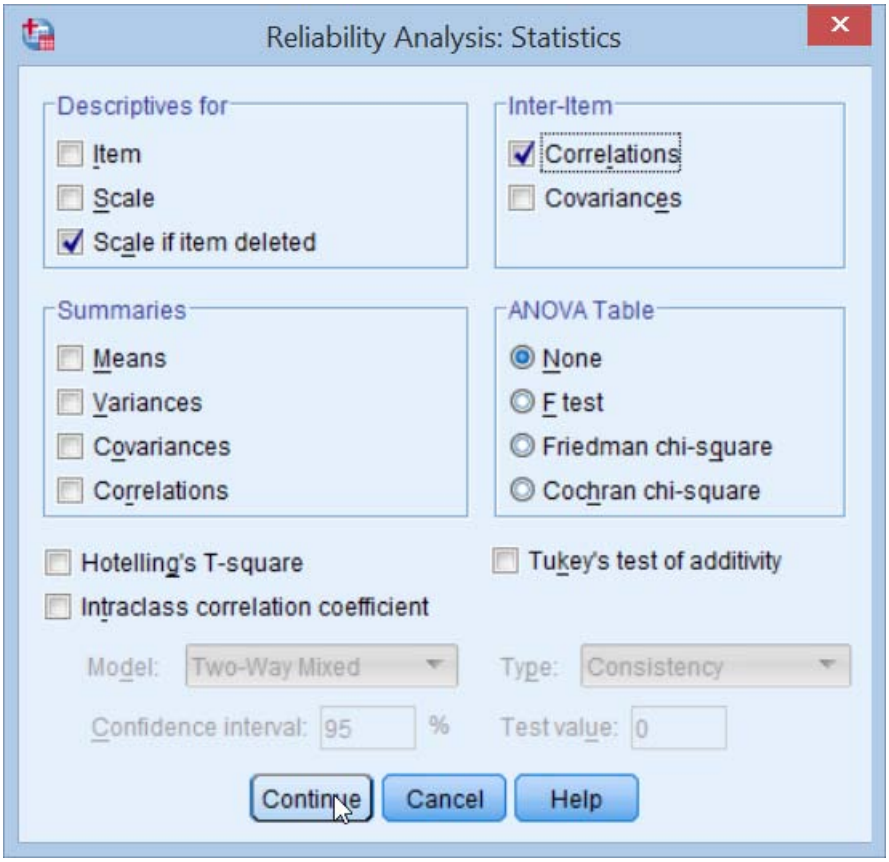
Tega na pamet žal ne moremo trditi. Zanesljivost vprašalnika izračunamo v SPSS-u s testom zanesljivosti oz. Reliability Analysis, ki pokaže vrednost Cronbachovega koeficienta α. Predpogoj za računanje Cronbachovega koeficienta α je intervalna lestvica. Če je vprašalnik sestavljen iz več lestvic, moramo izračunati vrednost Cronbachovega koeficienta α za vsako lestvico posebej. Najbolj strogo merilo za zanesljivost vprašalnika je, da je vrednost Cronbachovega koeficienta α enaka ali večja od 0,8. Za teste sposobnosti naj bi bila lestvica oz. vprašalnik dovolj zanesljiv, če njegova vrednost znaša 0,7.

3. Kako najdem Reliability Analysis v SPSS-u?

Odprite pogovorno okno Analyze → Scale → Reliability Analysis. V okence Items prenesite vse postavke določene lestvice stališč oz. vprašalnika. Nato kliknite na okence Statistics.



Odpre se vam pogovorno okno Reliability Analysis: Statistics. Tu odkljukajte okenci Scale if item deleted ter Correlations ter kliknite Continue ter potem OK.



4. Kako naj razumem output Reliability Analysis?

Oglejmo si output Reliability Analysis za preverbo zanesljivosti lestvice stališč o pravi ljubezni. Štiristopenjska lestvica, kjer je 1 pomenilo »se sploh ne strinjam« ter 4 »se popolnoma strinjam«, je vsebovala 5 izjav. Output Reliability Analysis zgleda takole.[2] Tabela 5 kaže vrednost Cronbachovega koeficienta α, ki v našem primeru znaša .839. Ker je torej njegova vrednost večja od .8, lahko zaključimo, da je naša lestvica dovolj zanesljiva. V tabeli vidimo še, koliko bi znašal ta koeficient na standardiziranih postavkah ter koliko postavk je bilo vključenih v analizo.

[2] Če uporabljate starejšo verzijo SPSS-a , lahko Output zgleda malo drugače, vse tabele so namreč združene v eni.

Tabela 5: Izračun koeficienta zanesljivosti

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,839	,838	5

V tabeli 6 vidimo korelacijske koeficiente med postavkami. Field (2005) priporoča, da v tej tabeli preverimo, kako korelirajo postavke med seboj. Na Cronbachov koeficient α namreč vpliva število postavk lestvice oz. vprašalnika; več ko je postavk, večja je lahko zanesljivost. Zato se priporoča, da imajo lestvice raje več vprašanj, a zaradi načela ekonomičnosti je priporočljivo, da jih je denimo 5 do 10.

Tabela 6: Korelacijski koeficienti med postavkami vprašalnika

Inter-Item Correlation Matrix

	ljubezen je odločitev	ljubezen je do konca življenja	ljubezen je strast	ljubezen je delo	ljubezen je odrekanje
ljubezen je odločitev	1,000	,682	,506	,273	,697
ljubezen je do konca življenja	,682	1,000	,665	,357	,635
ljubezen je strast	,506	,665	1,000	,543	,472
ljubezen je delo	,273	,357	,543	1,000	,257
ljubezen je odrekanje	,697	,635	,472	,257	1,000

The covariance matrix is calculated and used in the analysis.

V tabeli 7 vidimo v stolpcu Corrected Item-Total Correlation korelacije med vsako postavko ter vsemi ostalimi postavkami vprašalnika (Field, 2005). Če je lestvica zanesljiva, morajo postavke dobro korelirati s celoto. Če katera postavka slabše korelira s celoto, t.j. če je vrednost njenega korelacijskega koeficienta v tem stolpcu manjša kot .3, potem je priporočljivo, da jo izločimo iz vprašalnika. Vidimo, da v našem primeru najnižja korelacija nastopi pri postavki ljubezen je delo in sicer .410, kar pa ni pod .3 in zato ni kandidatka za izločitev. V zadnjem stolpcu te tabele Cronbach's Alpha if Item Deleted vidimo vrednosti Cronbachovega koeficienta α , če posamezna postavka vprašalnika ne bi bila vključena v vprašalnik. Vrednosti koeficienta v tem stolpcu ne smejo biti večje od vrednosti Cronbachovega koeficienta α v prvi tabeli, ki v našem primeru znaša .839. Vidimo, da nobeden od koeficientov ne presega te vrednosti. Če pa bi kateri od koeficientov presegel to vrednost, potem moramo izločiti to postavko iz vprašalnika, saj bo brez nje Cronbachov koeficient α višji, to pa pomeni, da bo lestvica bolj zanesljiva.

Tabela 7: Korelacijski koeficienti med posamezno postavko ter vsemi postavkami vprašalnika

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
ljubezen je odločitev	8,93	7,927	,714	,584	,787
ljubezen je do konca življenja	9,40	8,767	,773	,627	,770
ljubezen je strast	9,80	9,541	,672	,553	,800
ljubezen je delo	10,16	11,478	,410	,294	,838
ljubezen je odrekanje	8,16	8,332	,679	,534	,797

5. Zakaj je vrednost mojega Cronbachovega koeficienta α negativna?

To se zgodi v primeru, če imamo nekaj postavk vprašalnika zapisanih v trdilni, nekaj pa v nikalni obliki, kar je priporočljivo za vsak vprašalnik. Ta naj bi imel denimo pol vprašanj v trdilni in pol v nikalni obliki. Npr. da bi želeli ugotoviti, kakšna so menja vzgojiteljic v vrtcu glede vpliva igranja računalniških igric na razvoj agresivnega vedenja pri otroku. Na lestvici stališč imamo lahko več postavk, med katerimi sta dve formulirani takole: a) otroci, ki igrajo več računalniških igric, se bolj nasilno vedejo kot otroci, ki jih igrajo manj ali pa sploh ne, b) računalniške igrice pozitivno vplivajo na celosten razvoj otroka. Seveda taki dve postavki ne smemo zapisati v vprašalniku eno za drugo, temveč ju razporedimo med ostale trditve. Vendar pri štiristopenjski lestvici, kjer je 1 – se sploh ne strinjam in 4 – se popolnoma strinjam, pri prvi postavki 1 pomeni, da računalniške igrice ne vplivajo negativno na otrokov razvoj, pri drugi postavki pa 1 pomeni, da negativno vplivajo.

V takem primeru moramo rekodirati dobljene vrednosti za trdilno postavljene trditve ali nikalno postavljene trditve. Za rekodiranje uporabimo funkcijo Recode (glej poglavje Prilagajanje podatkov in oblikovanje novih spremenljivk) in za naš primer spremenimo vrednosti:

- $1 \rightarrow 4$,
- $2 \rightarrow 3$,
- $3 \rightarrow 2$
- $4 \rightarrow 1$.

Mimogrede: Field (2005) pravi, da bi lahko zgornje podatke lahko spremenili tudi s funkcijo Compute. Več o tem v Field, A. (2005). Discovering Statistics Using SPSS. London: Sage Publications, str. 669.

6. Kako navedem vrednost Cronbachovega koeficienta α?

Pravzaprav vrednosti tega koeficienta ne navajamo eksplicitno, saj je razviden iz tabele. Lahko pa zapišemo npr. tako kot smo zgoraj: vrednost Cronbachovega koeficienta α znaša .839. Ker je njegova vrednost večja od .8, lahko zaključimo, da je lestvica dovolj zanesljiva.

POSTAVLJANJE IN INTERPRETACIJA HIPOTEZ

1. Zakaj potrebujemo hipoteze?

Karkoli raziskujemo, navadno raziskujemo na vzorcu in ne na celotni populaciji (pač ne moremo izprašati vseh ljudi na svetu). Pri preučevanju določenega pojava pa nas ne zanimajo rezultati, ki smo jih dobili na vzorcu (npr. ali obstajajo razlike med 130 vprašanimi moškimi in ženskami glede doživljanja nasilja), temveč nas zanimajo splošna spoznanja o ljudeh in svetu (ali na splošno obstajajo razlike glede doživljanja nasilja med moškimi in ženskami). Ali povedano drugače, v kvantitativnem raziskovanju nas zanimajo ugotovitve, ki so posplošljive na vse člane neke populacije. Ali lahko posplošimo rezultate iz vzorca na populacijo nam pove postopek preverjanja hipotez.

2. Kaj je preverjanje hipotez?

Je način, na osnovi katerega lahko potrdimo ali zavrnemo, da rezultati, ki smo jih dobili na vzorcu, veljajo za celotno populacijo. V hipotezi zapišemo predpostavko o pričakovanih razlikah med ljudmi oz. povezavah med določenimi pojavi itd. Temu pravimo raziskovalne hipoteze. Npr. »Med otroki staršev delavskega razreda ter otroki staršev višjega razreda obstajajo razlike glede obiskovanja interesnih dejavnosti. Otroci staršev višjega razreda obiskujejo več interesnih dejavnosti.«

Če bi rekli, da »med otroki staršev delavskega razreda in otroki staršev višjega razreda ni razlik glede obiskovanja interesnih dejavnosti«, govorimo o ničelni hipotezi. Te nikoli ne pišemo, jo pa imamo v mislih pri ugotavljanju statistično pomembnih razlik med dvema skupinama (v tem primeru otrok). Torej v postavljanju hipotez vedno pišemo raziskovalne hipoteze, saj iščemo razlike oz. povezave med skupinami glede določenih pojavov.

3. Kako postavljamo hipoteze?

Hipoteze praviloma postavljamo vnaprej, torej preden izvedemo raziskavo oz. ko imamo razjasnjeno teoretično ozadje pojava, ki ga raziskujemo. Povedano drugače: hipoteze postavimo tako, da izhajajo iz neke teorije. Če hipoteze postavljamo tako rekoč sproti oz. za nazaj, postavimo take hipoteze, ki seveda potrdijo rezultate. To po drugi strani pomeni, da manipuliramo s številkami, jih malo prirejamo, da dobimo želene rezultate. Torej ne raziskujemo, temveč navidezno ugotavljamo in potrjujemo tisto, za kar smo se že vnaprej odločili, da drži. To pa nima nobene zveze z raziskovanjem.

4. Kako potrdimo oz. zavrnemo hipotezo?

Da lahko hipotezo potrdimo in ovržemo, uporabimo različne teste, ki pokažejo, ali so rezultati statistično pomembni (povedano po domače; ali jih lahko posplošimo na vso populacijo). Če ti testi pokažejo, da so razlike med parametri (npr. razlike med aritmetičnimi sredinami moških in žensk glede starosti ob prvem spolnem odnosu) ali povezave med parametri (npr. korelacija med starostjo ob prvem spolnem odnosu ter starostjo prve nosečnosti) statistično pomembne, potem sprejmemo hipotezo. To z drugimi besedami pomeni, da to, kar smo izračunali na reprezentančnem vzorcu, lahko trdimo, da velja za vso populacijo z določenim tveganjem (npr. 5%). Rečemo, da na ravni $p \leq 0{,}05$, sprejmemo hipotezo. To pomeni, da sprejmemo hipotezo ob 5% tveganju, da smo se zmotili oz. povedano drugače, obstaja 5% možnosti, da se je ta rezultat pojavil naključno. Vedno obstaja določena stopnja tveganja, da smo se zmotili in sicer 5% ali 1% ali 0,1% (odvisno od ravni tveganja), vendar to dejstvo je pač treba sprejeti. Pri še tako dobrih teorijah, ki so bile postavljene z manj kot 0,1% tveganja, to tveganje pač obstaja. Je sicer zanemarljivo majhno, a vendarle je.

Če pa razlike oz. povezave med parametri niso statistično pomembne, hipotezo zavrnemo. Torej rezultatov pridobljenih na vzorcu ne moremo posplošiti na vso populacijo. Vendar pozor! Če razlike niso statistično pomembne, ne rečemo, da ni razlik med skupinami (npr. med moškimi in ženskami glede starosti ob prvem spolnem odnosu), temveč da nismo uspeli dokazati razlik!!!! Razlike morda v populaciji res obstajajo, a je naš vzorec morda premajhen, morda premalo reprezentativen itd., da na njem teh razlik nismo uspeli dokazati.

Ko hipotezo sprejmemo ali zavrnemo, poskušamo interpretirati ta rezultat. Na osnovi spoznanj iz teoretičnega uvoda poskušamo razložiti zakaj je temu tako, navajamo rezultate podobnih raziskav in jih primerjamo s svojimi, povezujemo s prakso... Skratka tu se začne razlaga dobljenih rezultatov in postavljanje teorije.

5. Kakšna je za teorijo dobra hipoteza?

Če nekoliko poenostavljeno rečemo, je dobra hipoteza taka, ki je v povezavi s teorijo oz. je usmerjena na ugotavljanje določenih zakonitosti med pojavi, ki izhajajo iz konceptov neke teorije (npr. iskanje povezav med pojavi, vzročno posledičnih zvez itd.) in ni usmerjena zgolj na ugotavljanje stanja in obsega pojavov. Kakšna hipoteza je usmerjena na ugotavljanje stanja in obsega pojavov? To je hipoteza, ki denimo ocenjuje odstotek nekega pojava v družbi (npr. 25% vseh srednješolcev kadi travo). Sicer je res zanimivo vedeti, kakšno je stanje v družbi (torej koliko srednješolcev kadi travo), vendar take hipoteze temeljijo zgolj na opisovanju tega stanja in zato niso pojasnjevalne. Vsekakor pa so zanimive za ugotavljanje obsega pojava.

Kako bi se glasila hipoteza, ki odkriva zakonitosti med pojavi? Npr. »V populaciji srednješolcev obstaja korelacija med kajenjem trave ter šolskim uspehom. Srednješolci, ki kadijo travo, imajo v povprečju slabši šolski uspeh«. Ta hipoteza odkriva zakonitost relacije med pojavom kajenja trave ter šolskim uspehom.

5. Kako navedem hipoteze v raziskovalni nalogi?

Najbolj pregledno jih navedemo takole:

H1: Verne študentke socialne pedagogike se bolj kot neverne strinjajo s trditvijo, da se v partnerski ljubezni čemu odrečeš.

H2: Obstajajo razlike med študentkami socialne pedagogike, ki so že imele in tistimi, ki še niso imele spolnih odnosov, glede strinjanja s stališčem, da je ljubezen odločitev.

H3: Študentke socialne pedagogike, ki prihajajo iz vasi ali manjšega mesta, so bolj verne kot študentke, ki prihajajo iz večjega mesta.

H4: Pri študentkah socialne pedagogike se ocena doživljanja sreče v študijski dobi povezuje z napovedjo ocene sreče v odrasli dobi

Več o tem: Field, A. (2005). Discovering Statistics Using SPSS. London: Sage Publications. Str. 22 – 33.

DILEME V ZVEZI Z IZBOROM PRAVEGA TESTA ZA PREVERBO HIPOTEZ

1. Ali je od vrste spremenljivk, ki jih meri vprašalnik, odvisen izbor testa, s katerim preverjam hipoteze?

Seveda. Od tega, katere spremenljivke meri vprašalnik, je odvisno, za kateri test se odločite. Vsak test je primeren zgolj za določene vrste spremenljivk ter je pravilno uporabljen le, če se upošteva vse predpostavke tega testa. Zato so v nadaljevanju pri opisu posameznih testov vedno napisane vrste spremenljivk, za katere se določen test uporablja ter predpostavke zanj.

Za ponovitev si na kratko oglejmo vrste spremenljivk:

ATRIBUTIVNE	NOMINALNE	So kakovostne spremenljivke, ki merijo značilnosti, na osnovi katerih so si vprašani po njej enaki ali različni, ne moremo pa jih po tej značilnosti razvrščati na boljše ali slabše, večje, manjše oz. jih primerjati. Npr. spol, zakonski stan, verska opredeljenost…
	ORDINALNE	So spremenljivke, ki merijo značilnosti, za katere lahko ugotovimo, katera je večja in katera manjša. Lahko jih torej razvrstimo po velikosti, vendar pa ordinalno merjenje NE določa velikosti razlik med vrednostmi spremenljivke. Npr. če bi samo po velikosti (izgledu) postavili otroke v vrsto (a ne bi merili velikosti v cm!)
NUMERIČNE	INTERVALNE	So spremenljivke, ki merijo značilnosti, za katere lahko ugotovimo tudi, koliko je katera vrednost večja/manjša od druge. Določeni so intervali med vrednostmi spremenljivke. Nimajo pa absolutne ničelne točke oz. vrednosti; ta je arbitrarno določena. Npr. lestvice stališč, npr. Likartova (ponavadi imajo skalo od se zelo strinjam, se strinjam, se ne strinjam, se sploh ne strinjam), testi osebnostnih lastnosti, testi znanja.
	PROPORCIONALNE oz. RAZMERNOSTNE	So spremenljivke, katerih možne vrednosti imajo absolutno ničlo in zato imajo tudi lastnost razmernosti. Npr. starost, število otrok, enote tedensko zaužitega alkohola, višina osebnega dohodka…

Več o tem glej Mesec, B. (1997). Metodologija raziskovanja v socialnem delu. Ljubljana: VŠSD. Str. 56 – 70.

2. Za katere spremenljivke so primerni posamezni testi?

Ambrožič in Leskošek (1999) jih delita na teste analize razlik ter analize povezanosti. Omenjamo samo nekatere izbrane teste.

<table>
<tr><th colspan="2">ANALIZA RAZLIK</th></tr>
<tr><th>ZA NUMERIČNE SPREMENLJIVKE</th><th>ZA ATRIBUTIVNE SPREMENLJIVKE</th></tr>
<tr><td>t-test: primerjava aritmetičnih sredin dveh skupin med seboj.</td><td rowspan="2">χ^2 – test: primerjava empiričnih ter teoretičnih frekvenc dveh spremenljivk.</td></tr>
<tr><td>ANOVA: primerjava aritmetičnih sredin več skupin med seboj.</td></tr>
</table>

<table>
<tr><th colspan="3">ANALIZA POVEZANOSTI</th></tr>
<tr><th rowspan="2">ZA NUMERIČNE SPREMENLJIVKE</th><th colspan="2">ZA ATRIBUTIVNE SPREMENLJIVKE</th></tr>
<tr><th>NOMINALNE</th><th>ORDINALNE</th></tr>
<tr><td>Pearsonov korelacijski koeficient (r): ugotavljanje povezanosti dveh numeričnih spremenljivk.</td><td>χ^2 – test: ugotavljanje povezave med dvema spremenljivkama.</td><td>Spearmanov korelacijski koeficient (r_s): ugotavljanje povezanosti dveh ordinalnih spremenljivk.</td></tr>
</table>

V primeru, da imamo eno nominalno ter eno intervalno spremenljivko ali pa eno ordinalno in eno intervalno, pa uporabimo koeficient korelacijskega razmerja $\mathcal{E}$.

Več o izboru pravilnega testa v preglednici v:
Cencič, M. (2002). Pisanje in predstavljanje rezultatov raziskovalnega dela: kako se napiše in predstavi diplomsko delo (nalogo) in druge vrste raziskovalnih poročil. Ljubljana: Pedagoška fakulteta, str.68. V preglednici so vključeni tudi testi, ki jih tu zaradi vsebinske omejenosti ne omenjamo.

T-TEST ZA NEODVISNE VZORCE

1. Kdaj uporabim t-test?

T-test za neodvisne vzorce uporabimo, ko želimo med seboj primerjati dve neodvisni skupini. S t-testom ugotavljamo, ali se aritmetični sredini dveh skupin med seboj statistično pomembno razlikujeta.

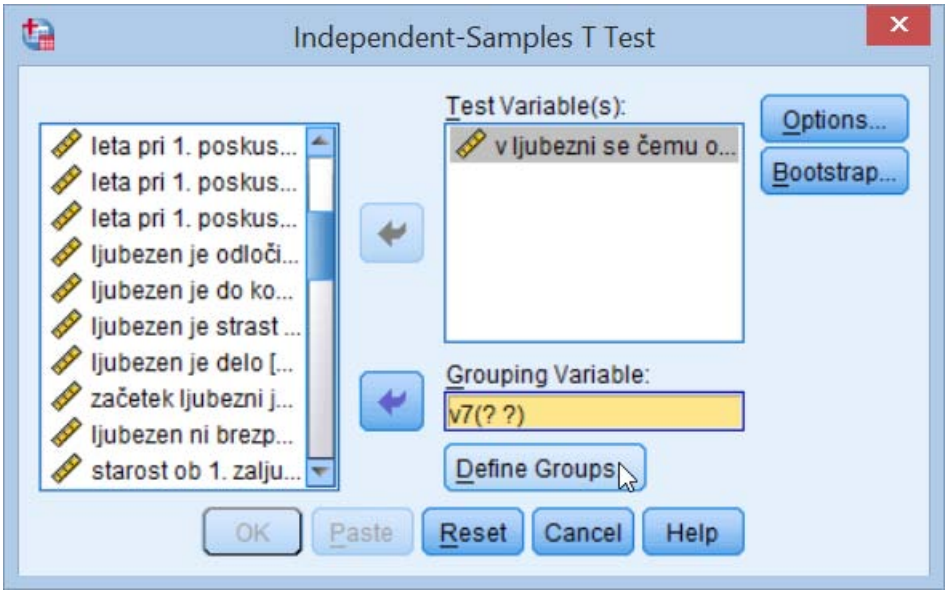
Npr. ali se študentje prava in študentje socialne pedagogike razlikujejo glede povprečnih mesečnih izdatkov za žur. To bi ugotovili s t-testom, ki bi meril aritmetično sredino povprečnih mesečnih izdatkov za žur pri dveh skupinah; študentih prava in študentih socialne pedagogike.

Predpostavke za t-test (Field, 2005: 287):

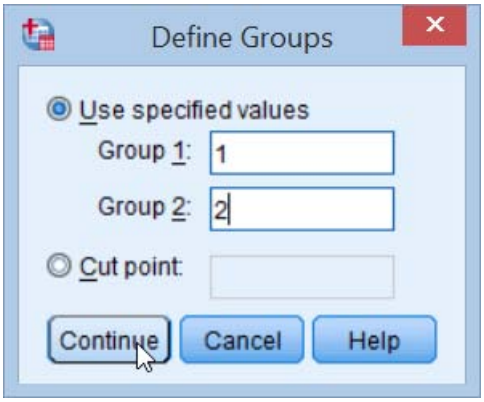
- podatki so vzeti iz normalno distribuirane populacije
- podatki so numerični (tj. intervalni ali razmernostni oz. proporcionalni)
- variance v obeh skupinah populacij so približno enake (homogenost varianc)
- podatki so med seboj neodvisni, saj so pridobljeni na različnih posameznikih.

2. Kako najdem t –test v SPSS-u?

Odprite pogovorno okno Analyze → Compare Means → Independent-Samples T test. V okence Test Variable(s) prenesite odvisno/-e spremenljivko/e, v okence Grouping Variable pa neodvisno spremenljivko. Nato kliknite na okence Define Groups.



Odpre se vam okno Define groups. V okence Group 1 vnesite vrednost, ki označuje prvo skupino (npr. skupini vernih študentk pripada vrednost 1, nevernih pa 2), v okence Group 2 pa vrednost, ki označuje drugo skupino. Nato kliknite Continue in OK.



3. Kako naj razumem output t-testa?

Output t-testa vsebuje:

- opisne mere po skupinah (Group Statistics)
- Levene-ov preizkus razlik med variancama
- t- preizkus razlik med a.s. (preizkus skupin) in meje intervala zaupanja za razliko a.s. pri predpostavki enakosti oz. neenakosti varianc (Rovan, Turk, 1999).

Oglejmo si output t – testa za preverbo hipoteze: »Verne študentke socialne pedagogike se bolj strinjajo s trditvijo, da se v partnerski ljubezni čemu odrečeš.« Vzorec predstavljajo študentke socialne pedagogike v študijskih letih 2003/04 ter 2004/05. Lestvica, ki je vsebovala izjavo »V pravi ljubezni se moraš kdaj čemu odreči v dobro svojega partnerja,« je bila štiristopenjska, kjer je 1 – se sploh ne strinjam ter 4 – se popolnoma strinjam. Output t- testa zgleda takole.
Tabela 8 kaže opisne mere po skupinah oz. deskriptivno statistiko. Vidimo, da je vseh vernih študentk, ki so odgovorile na to vprašanje 30, vseh nevernih pa 17. Aritmetična sredina strinjanja z izjavo je pri vernih študentkah znašala 3.33, pri nevernih pa 2.94, kar je oboje zelo blizu strinjanja s stopnjo 3 – se precej strinjam. Tudi standardna odklona sta si blizu za obe skupine študentk, čeprav je standardni odklon pri nevernih študentkah večji glede na aritmetično sredino kot pri vernih. Zadnji stolpec prikazuje standardno napako aritmetične sredine (to je standardni odklon vzorčne distribucije).

Pazi!!! V SPSS-u ni zapisa decimalnih števil z 0, temveč se začne s piko. Torej .130 pomeni 0.130.

Tabela 8: Opisne mere po skupinah

T-Test

Group Statistics

	verska opredeljenost	N	Mean	Std. Deviation	Std. Error Mean
v ljubezni se čemu odrečeš	veren	30	3,33	,711	,130
	neveren	17	2,94	,899	,218

V tabeli 9 z naslovom Independent Samples Test najprej pogledamo prvi stolpec z naslovom Leven's Test for Equality of Variances ter predvsem kakšna je vrednost Sig. (statistična pomembnost F-testa homogenosti varianc, označimo jo tudi s p) v vrstici Equal variances assumed. Ta vrstica temelji na predpostavki, da sta varianci homogeni.
Če je vrednost Sig. oz. p:

$p \leq 0{,}05 \Rightarrow$ razlike med variancama obeh skupin so statistično pomembne. (H_0, ki pravi: ni razlik med variancama obeh skupin zavrnemo.) Ker so razlike, varianci nista homogeni. Torej moramo ovreči predpostavko o homogenosti varianc. → Glej vrstico Equal variances not assumed.

$p > 0{,}05 \Rightarrow$ razlike med variancama obeh skupin niso statistično pomembne. H_0 obdržimo, varianci sta približno enaki. Torej obdržimo predpostavko o homogenosti varianc. → Glej vrstico Equal variances assumed.

Tabela 9: T-test za preverjanje razlik o mnenju o odrekanju v ljubezni glede na vernost študentk

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
v ljubezni se čemu odrečeš	Equal variances assumed	,141	,709	1,649	45	,106	,392	,238	-,087	,871
	Equal variances not assumed			1,545	27,446	,134	,392	,254	-,128	,913

Ker je **p > 0,05**, obdržimo predpostavko o homogenosti varianc. Sedaj šele gledamo dalje rezultat t-testa v tej vrstici in spodnjo vrstico Equal variances not assumed popolnoma ignoriramo.

Stolpec Sig. (2-tailed) nam pove, ali je razlika med aritmetičnima sredinama obeh skupin statistično pomembna ali ne. Ker je **p > 0,05,** razlike med aritmetičnima sredinama obeh skupin niso statistično pomembne. Zato hipotezo : »Verne študentke socialne pedagogike se bolj strinjajo s trditvijo, da se v partnerski ljubezni čemu odrečeš« zavrnemo. Nismo uspeli dokazati razlik med vernimi in nevernimi študentkami socialne pedagogike glede mnenja o odrekanju v partnerski ljubezni.

$p \leq 0,05$	$\Rightarrow$	razlike med aritmetičnima sredinama obeh skupin so statistično pomembne. (H_0, ki pravi, da ni razlik med njima, zavrnemo.) Torej lahko posplošimo razlike med vzorcema tudi na celotno populacijo s 5 % tveganjem.
$p > 0,05$	$\Rightarrow$	razlike med aritmetičnima sredinama obeh skupin niso statistično pomembne. (H_0 obdržimo.) Torej nismo uspeli dokazati, da bi razlike med aritmetičnima sredinama našega vzorca veljale za celotno populacijo.

Poglejmo si še en primer. Vzorec je isti kot pri zgornjem, prav tako je lestvica stališč štiristopenjska kot zgoraj. Hipoteza se glasi: »Obstajajo razlike med študentkami socialne pedagogike, ki so že imele in tistimi, ki še niso imele spolnih odnosov, glede strinjanja s stališčem, da je ljubezen odločitev.« Output je sledeči:

Tabela 10: Opisne mere po skupinah

T-Test

Group Statistics

	so/niso imele spolnega odnosa	N	Mean	Std. Deviation	Std. Error Mean
ljubezen je odločitev	so imele spolni odnos	39	2,41	,938	,150
	niso imele spolnega odnosa	11	3,09	,831	,251

Iz tabele 10 preberemo naslednje parametre: da je na to vprašanje odgovorilo 39 študentk, ki so že imele spolne odnose ter 11 študentk, ki jih še ni imelo. Aritmetična sredina za prvo skupino znaša 2.41 (so nekje vmes med »se deloma strinjam« ter »se precej strinjam«, za drugo pa 3.09 (v povprečju so se torej odločile za »se precej strinjam«). Standardni odklon za prvo skupino znaša .938, za drugo pa .831, medtem ko standardna napaka aritmetične sredine znaša .150, za drugo pa .251.

Tabela 11: T-test za preverjanje razlik o mnenju o ljubezni kot odločitvi glede na spolno prakso študentk

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
ljubezen je odločitev	Equal variances assumed	2,456	,124	-2,175	48	,035	-,681	,313	-1,310	-,051
	Equal variances not assumed			-2,330	17,868	,032	-,681	,292	-1,295	-,066

Ker je **p > 0,05**, obdržimo predpostavko o homogenosti varianc. Sedaj šele gledamo dalje rezultat t-testa v tej vrstici in spodnjo vrstico Equal variances not assumed popolnoma ignoriramo.

Stolpec Sig. (2-tailed) nam pove, ali je razlika med aritmetičnima sredinama obeh skupin statistično pomembna ali ne. Ker je **p < 0,05,** so razlike med aritmetičnima sredinama obeh skupin statistično pomembne. Študentke, ki niso imele spolnih odnosov, se bolj strinjajo s stališčem, da je ljubezen odločitev, kot pa študentke, ki so spolne odnose že imele.

Za izvedbo t – testa po korakih glej:
Filed, A. (2005).). Discovering Statistics Using SPSS. London: Sage Publications. Str. 299 - 303.
Rovan, Turk (1999). Analiza podatkov s SPSS za Windows. Ljubljana: Ekonomska fakulteta.

4. Kako navedem rezultate t-testa?

Field (2005) pravi, da je po standardih APA-e (American Psychological Association) pomembno, da navedemo naslednje parametre: M (aritmetično sredino), SE (standardno napako ocene aritmetične sredine), t (vrednost t-testa), df (stopnje prostosti) ter p (statistično pomembnost t-testa). Primer navedbe rezultatov za naš prvi primer vernih in nevernih študentk.

»V povprečju se verne študentke bolj strinjajo s trditvijo, da se je v partnerski ljubezni potrebno čemu odreči (M = 3.33, SE = .130) kot pa neverne študentke (M = 2.94, SE = .218). Ta razlika ni statistično pomembna $t(45) = 1.649$, $p > .05$.

Temu dodajam še: Hipotezo tako zavrnemo.

Sedaj pa še navedba rezultatov za drug primer študentk, ki so oz. niso imele spolnih odnosov: »V povprečju se študentke, ki še niso imele spolnih odnosov pomembno bolj strinjajo s trditvijo, da je ljubezen odločitev (M = 2.41, SE = .150) kot pa študentke, ki so že imele spolne odnose (M = 3.09, SE = .251, $t(48) = -2.175$, $p < .05$.

Lahko dodamo še: hipotezo tako sprejmemo.

Od tu dalje pa sledi interpretacija dobljenih rezultatov!

χ2 – TEST

1. Kdaj uporabim χ^2- test?

Pearsonov χ^2 – test je neparametrični preizkus, ki ga uporabimo, ko želimo ugotoviti, ali obstaja zveza med dvema atributivnima spremenljivkama (za ponovitev; to sta lahko nominalni ali ordinalni), ki imata 2 ali več kategorij. S χ^2 – testom ugotavljamo, kako dejanske = empirične frekvence posameznih kategorij odstopajo od teoretičnih = pričakovanih frekvenc (ki bi jih pri teh kategorijah dobili naključno). Te frekvence vidimo v kontingenčni tabeli (Crosstabulation).

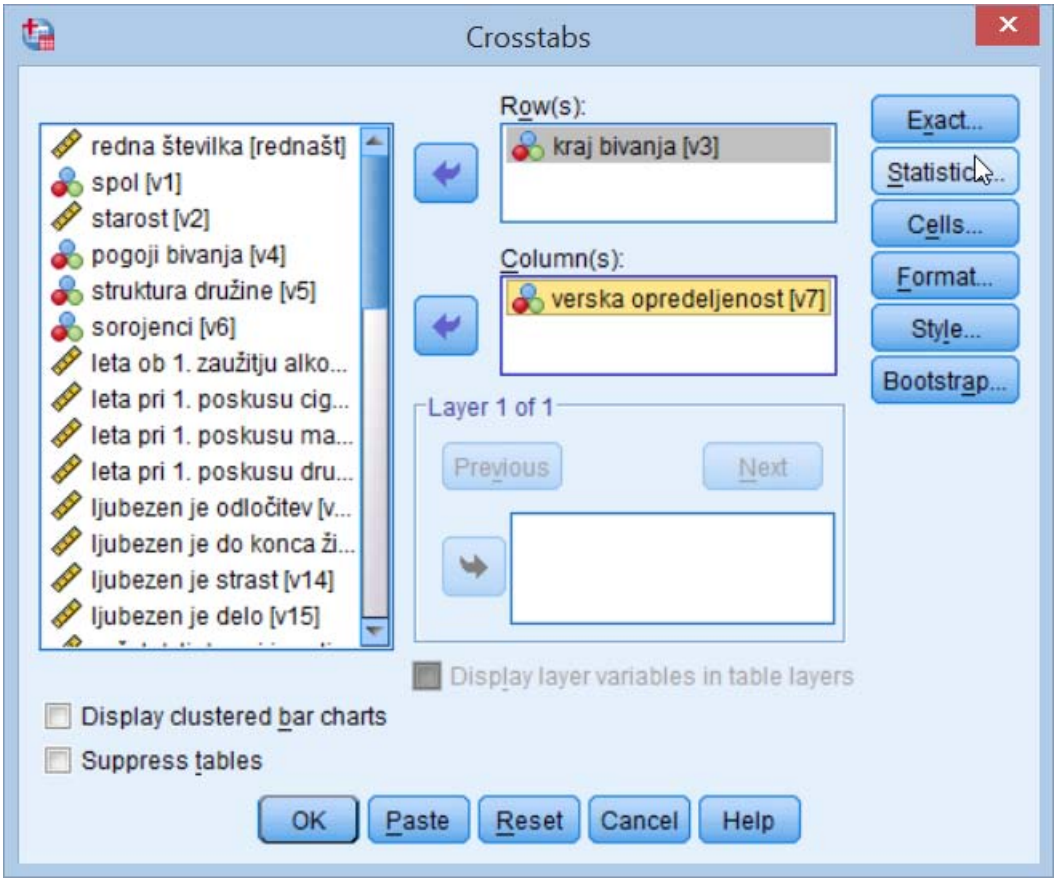
Npr. ali se ženske in moški razlikujejo glede poznavanja nekega računalniškega programa. Obe spremenljivki sta nominalni; spol (m, ž) ter poznavanje programa (da, ne). Ali ta razlika obstaja, bi ugotovili s χ^2 – testom, ki bi meril frekvence poznavanja tega računalniškega programa pri ženskah in pri moških ter jih primerjal s pričakovanimi oz. teoretičnimi frekvencami. Pazi: lahko rečemo, da s χ^2 – testom ugotavljamo povezanost med dvema spremenljivkama oz. razlike med dvema skupinama. Lahko rečemo, da nas zanima povezava med spolom in poznavanjem računalniškega programa ali pa, da nas zanimajo razlike med moškimi in ženskami glede poznavanja računalniškega programa.

Predpostavke za χ^2 – test (Field, 2005: 686):

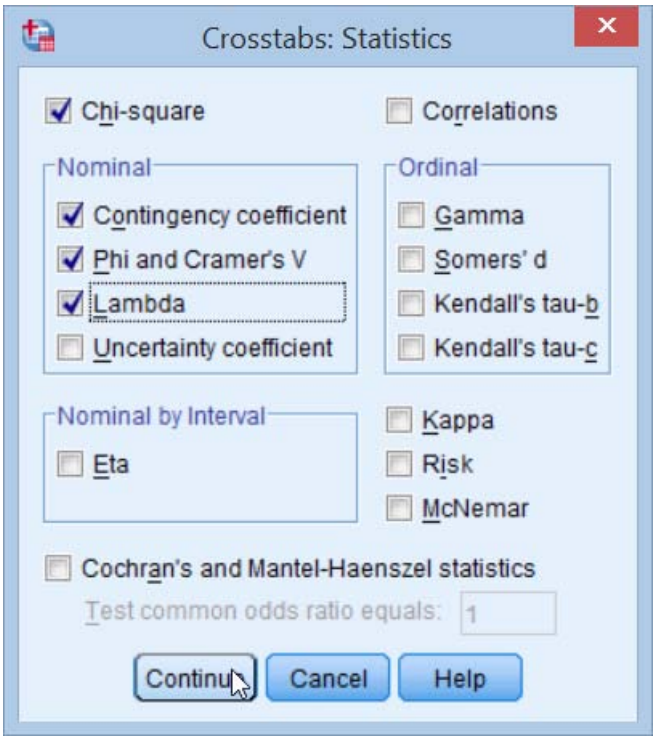
- podatki so med seboj neodvisni, saj so pridobljeni na različnih posameznikih;
- pričakovane frekvence morajo biti večje od 5; za večje kontingenčne tabele je ta pogoj milejši – največ 20% vseh pričakovanih frekvenc je lahko manjši od 5;
- nobena pričakovana frekvenca ne sme biti manjša od 1.

2. Kako najdem χ^2– test v SPSS-u?

Odprite pogovorno okno Analyze → Descriptive Statistics → Crosstabs. Prenesete spremenljivke kot je opisano v poglavju Deskriptivna statistika – funkcija Crosstabs. Kliknite na okence Cells in označite kot je opisano v omenjenem poglavju. Nato kliknite še okence Statistics.



V okencu Statistics odkljukajte vrstice Chi-Square, Contingency coefficient, Phi and Cramer's V ter Lambda. Nato kliknite Continue in potem OK.



3. Kako naj razumem output χ^2- testa?

V Outputu dobimo štiri tabele:

- povzetek značilnosti statistične analize (Case Processing Summary)
- kontingenčno tabelo (Crosstabulation)
- χ^2 – test (Chi-Square Tests)
- moč povezave med spremenljivkama (Symmetric Measures).

Oglejmo si output χ^2 – testa za preverbo hipoteze: »Študentke socialne pedagogike, ki prihajajo iz vasi ali manjšega mesta, so bolj verne kot študentke, ki prihajajo iz večjega mesta.« Pazi, ničelna hipoteza pa – v nasprotju z raziskovalno - pravi, da ni razlik med skupinama. Vzorec predstavljajo študentke socialne pedagogike v študijskih letih 2003/04 ter 2004/05. Podatke smo analizirali iz dveh vprašanj; Kraj bivanja: vas ali manjše mesto ter Moja verska opredeljenost: sem verna, nisem verna. Output χ^2 – testa zgleda takole.

Tabela 12 z naslovom Case Processing Summary kaže numerus ter frekvenčne odstotke za obe kategoriji skupaj.

Tabela 12: Povzetek značilnosti statistične analize

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
kraj bivanja * verska opredeljenost	47	94,0%	3	6,0%	50	100,0%

Tabela 13 z naslovom kraj bivanja * verska opredeljenost Crosstabulation je kontingenčna tabela (njeno razlago ter interpretacijo gled pod poglavjem Deskriptivna statistika).

Tabela 13: : Kontingenčna tabela za kraj bivanja anketiranih glede na njihovo versko opredeljenost

kraj bivanja * verska opredeljenost Crosstabulation

			verska opredeljenost		Total
			veren	neveren	
kraj bivanja	vas ali manjše mesto	Count	18	11	29
		Expected Count	18,5	10,5	29,0
		% within kraj bivanja	62,1%	37,9%	100,0%
		% within verska opredeljenost	60,0%	64,7%	61,7%
		% of Total	38,3%	23,4%	61,7%
	večje mesto	Count	12	6	18
		Expected Count	11,5	6,5	18,0
		% within kraj bivanja	66,7%	33,3%	100,0%
		% within verska opredeljenost	40,0%	35,3%	38,3%
		% of Total	25,5%	12,8%	38,3%
Total		Count	30	17	47
		Expected Count	30,0	17,0	47,0
		% within kraj bivanja	63,8%	36,2%	100,0%
		% within verska opredeljenost	100,0%	100,0%	100,0%
		% of Total	63,8%	36,2%	100,0%

Preden nadaljujemo s χ^2 – testom, moramo še preveriti, če smo zadovoljili kriterij predpostavk, tj. da so vse pričakovane frekvence večje od 5 (ker je tabela 2x2 morajo biti vse pričakovane frekvence večje od 5 in ne velja možnost, da je največ 20 % pričakovanih vrednosti lahko manjših od 5 – to velja le za večje tabele) in da nobena ni manjša od 1. V kontingenčni tabeli preberemo vrednosti podvrstic Expected Count v vseh treh vrsticah: vas ali manjše mesto, večje mesto ter Total. Vidimo, da je najnižja vrednost pričakovanih frekvenc 6.5 (vrstica večje mesto, stolpec neveren), torej smo zadovoljili zahteve predpostavk (Field, 2005).

Če pa kateri od pogojev ne bi bil izpolnjen, potem lahko združimo kategorije – funkcija Recode v pogovornem oknu Transform ali pa zberemo večje število podatkov.

Sedaj poglejmo tretjo tabelo Outputa (Tabelo 14) z naslovom Chi-Square Tests. Za naše potrebe gledamo zgolj vrstico Pearson Chi-Square. Vidimo, da vrednost tega testa znaša .102 pri stopnji prostosti 1. Njegova statistična pomembnost pa .750. Ker je Ker je **p > 0,05**, zavrnemo hipotezo. Nismo uspeli dokazati povezave med krajem bivanja študentk socialne pedagogike ter njihovo vernostjo. Mimogrede, pod tabelo sta zapisani dve opombi in sicer, da je bil χ^2– test izračunan za 2x2 tabelo ter da 0 celic nima pričakovanih frekvenc manj od 5. Najmanjša pričakovana frekvenca znaša 6,51, kakor smo ugotovili že zgoraj. Te opombe nam pomagajo, da nam ni treba v kontingenčni tabeli iskati vrednosti pričakovanih frekvenc, vendar kar v opombah preverimo ali so predpostavke za χ^2– test zadovoljene.

Tabela 14: χ^2 – test za povezavo med krajem bivanja študentk in njihovo versko opredelitvijo

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,102(b)	1	,750		
Continuity Correction(a)	,000	1	,995		
Likelihood Ratio	,102	1	,749		
Fisher's Exact Test				1,000	,500
Linear-by-Linear Association	,100	1	,752		
N of Valid Cases	47				

a Computed only for a 2x2 table
b 0 cells (,0%) have expected count less than 5. The minimum expected count is 6,51.

Pearsonov χ^2 test pove, ali sta dve spremenljivki neodvisni druga od druge:

$p \leq 0{,}05 \Rightarrow$ spremenljivki sta povezani med seboj, sta odvisni
(H obdržimo oz. H_0 zavrnemo)

$p > 0{,}05 \Rightarrow$ nismo uspeli dokazati, da bi bili spremenljivki povezani med seboj
(H zavrnemo oz. H_0 obdržimo)

V tabeli 15 vidimo koeficiente, ki nam povedo, kakšna je povezava med spremenljivkama. Torej logika razumevanja teh koeficientov je enaka razumevanju ostalih korelacijskih koeficientov. Njihove vrednosti se gibljejo med 0 in 1. Field (2005) koeficiente, ki jih vidimo v spodnji tabeli, razlaga takole:

- Phi: ga upoštevamo, ko imamo 2x2 tabelo. Za večje tabele gledamo Contingency coefficient.
- Cramer's V: če gre za 2x2 tabelo, je njegova vrednost enaka koeficientu Phi. Če pa imamo večje tabele, je ta koeficient edini, ki zares lahko doseže maksimalno vrednost 1 in zato najbolj uporaben.
- Contingency coefficient: sicer zagotavlja vrednosti med 0 in 1, a maksimalno vrednost 1 le stežka doseže, zato je Cramer postavil nov koeficient.

Če torej pogledamo vrednost Cramerjevega koeficienta, vidimo, da je .047, kar implicira izredno nizko povezanost med spremenljivkama, vendar pa ta vrednost ni statistično pomembna, saj je raven njegove statistične pomembnosti enaka .750, kar smo videli že v zgornji tabeli.

V tabeli 15: Koeficienti povezanosti med spremenljivkama

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	-,047	,750
	Cramer's V	,047	,750
	Contingency Coefficient	,046	,750
N of Valid Cases		47	

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

Če bi želeli natančneje ugotoviti moč povezave med spremenljivkama Field (2005: 693) predlaga izračun kvocient različnosti, ki je še posebej v 2x2 tabeli hitro izračunljiv in lahko interpretabilen. Računamo ga po naslednji formuli:

$$\text{Različnostverni} = \frac{\text{število vernih, ki živijo na vasi ali v manjšem mestu}}{\text{število vernih, ki živijo v večjem mestu}} = \frac{18}{12} = 1{,}5$$

$$\text{Različnostneverni} = \frac{\text{število nevernih, ki živijo na vasi ali v manjšem mestu}}{\text{število nevernih, ki živijo v večjem mestu}} = \frac{11}{6} = 1{,}83$$

$$\text{Kvocient različnosti} = \frac{\text{odds}_{\text{neverni}}}{\text{odds}_{\text{verni}}} = \frac{1{,}83}{1{,}5} = 1{,}22$$

Ta rezultat nam pove, da so neverne študentke 1,22 – krat bolj verjetno doma na vasi oz. v manjšem mestu. Ta verjetnost je izredno majhna. In kot smo pokazali zgoraj ni statistično pomembna. Si pa vsekakor ta postopek velja zapomniti za primer, ko dobimo statistično pomembne rezultate $\chi 2$ – testa in želimo oceniti moč povezave med spremenljivkama.

4. Kako navedem rezultate χ^2-testa?

Nismo uspeli dokazati pomembne povezave med krajem bivanja študentk ter vernostjo študentk $\chi^2(1) = .102$, $p > 0{,}05$. Hipotezo tako zavrnemo. Nismo torej uspeli pokazati, da bi študentke socialne pedagogike, ki prihajajo iz vasi ali manjšega mesta, bile bolj verne kot študentke, ki prihajajo iz večjega mesta.

PEARSONOV KOEFICIENT KORELACIJE IN KORELACIJSKA ANALIZA

1. Kdaj uporabim Pearsonov koeficient korelacije (r) in test njegove statistične pomembnosti?

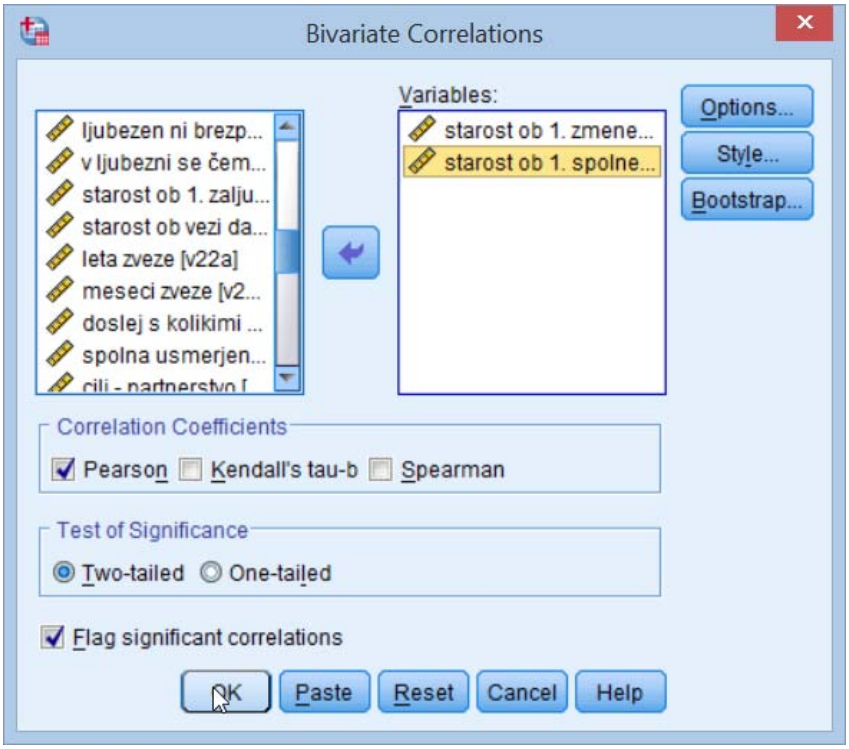
Pearsonov koeficient korelacije uporabimo, ko želimo na reprezentativnem vzorcu ugotoviti ali sta dve numerični spremenljivki medsebojno povezani ter kakšna je smer te povezave. Če je r statistično pomemben (test njegove statistične pomembnosti), potem lahko (ob določenem tveganju, seveda) trdimo, da tudi v populaciji obstaja taka povezava kot smo jo ugotovili na vzorcu.

Predpostavki za preverbo statistične pomembnosti r:

- Podatki obeh spremenljivk se porazdeljujejo normalno (izjema je v primeru, ko imamo eno dihotomno atributivno spremenljivko, npr. spol, drugo pa numerično. V tem primeru govorimo o point-biserialnem koeficientu, ki pa ga izračunamo enako kot Pearsonovega!)
- Korelacija med spremenljivkama je linearna.

2. Kako najdem Pearsonov korelacijski koeficient v SPSS-u?

Odprite pogovorno okno Analyze → Correlate → Bivariate. V okence Variables prenesete spremenljivke, med katerimi želite izračunati korelacijo in kliknete OK.



3. Kako naj razumem output Pearsonovega korelacijskega koeficienta?

Oglejmo si output statistične pomembnosti Pearsonovega korelacijskega koeficienta za preverbo hipoteze: »Pri študentkah socialne pedagogike se starost pri prvem zmenku pomembno povezuje s starostjo pri prvem spolnem odnosu.« Pazi, ničelna hipoteza pa – v nasprotju z raziskovalno - pravi, da ni povezave med tema dvema spremenljivkama. Vzorec predstavljajo študentke socialne pedagogike v študijskih letih 2003/04 ter 2004/05. Podatke smo analizirali iz dveh vprašanj; starost pri mojem prvem zmenku ter prvič sem imela spolne odnose pri starosti. Output zgleda takole.

Tabela 16: Pearsonov korelacijski koeficient med spremenljivkama starosti ob prvem zmenku in prvem spolnem odnosu

Correlations

Correlations

		starost ob 1. zmenek	starost ob 1. spolnem odnosu
starost ob 1. zmenku	Pearson Correlation	1	,428(**)
	Sig. (2-tailed)	.	,007
	N	49	38
starost ob 1. spolnem odnosu	Pearson Correlation	,428(**)	1
	Sig. (2-tailed)	,007	.
	N	38	39

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

Iz tabele 16 odčitamo vrednost Pearsonovega korelacijskega koeficienta, ki v našem primeru znaša r = .428. Vidimo, da je ta povezava statistično pomembna, saj je **p < 0,01** (sprejeli bi ga že na ravni p < 0,05, vendar vzamemo najnižjo stopnjo tveganja – ta je zapisana v opombi pod tabelo). V opombi prav tako preberemo, da je bil test dvostranski, saj smo predvideli povezavo, nismo pa povedali nič o njeni smeri. Če bi npr. postavili hipotezo, da obstaja pozitivna povezanost med starostjo pri prvem zmenku ter starostjo pri prvem spolnem odnosu, potem bi izvedli enostranski test.

Na osnovi vrednosti r (vemo, da se gibljejo od -1 pa do +1), lahko ocenimo, kolikšna naj bi bila moč te povezave. Nekako velja, da (Field, 2005: 32):

- r = .10 nizka povezanost (povezanost pojasni zgolj 1% skupne variance)
- r = .30 srednja povezanost (povezanost pojasni zgolj 9% skupne variance)
- r = .50 visoka povezanost (povezanost pojasni zgolj 25% skupne variance)

V našem primeru gre torej za srednjo povezanost omenjenih dveh spremenljivk.

Statistična pomembnost Pearsonovega korelacijskega koeficienta pove, ali sta dve spremenljivki medsebojno povezani:

$p \leq 0,05 \Rightarrow$ spremenljivki sta povezani med seboj, sta odvisni
(H obdržimo oz. H_0 zavrnemo)

$p > 0,05 \Rightarrow$ nismo uspeli dokazati, da bi bili spremenljivki povezani med seboj
(H zavrnemo oz. H_0 obdržimo)

4. Kako navedem rezultate pomembnosti Pearsonovega korelacijskega koeficienta?

Možnost 1:
Pri študentkah socialne pedagogike obstaja pomembna povezava med starostjo pri prvem zmenku ter starostjo pri prvem spolnem odnosu, r = .428, p (dvostranski) < 0,01. Študentke, ki so mlajše pri prvem zmenku, so tudi mlajše pri prvem spolnem odnosu.

Možnost 2:
Pri študentkah socialne pedagogike se starost pri prvem zmenku pomembno povezuje s starostjo pri prvem spolnem odnosu, r = .428, p (dvostranski) < 0,01. Študentke, ki so mlajše pri prvem zmenku, so tudi mlajše pri prvem spolnem odnosu.

SPEARMANOV KOEFICIENT KORELACIJE IN KORELACIJSKA ANALIZA

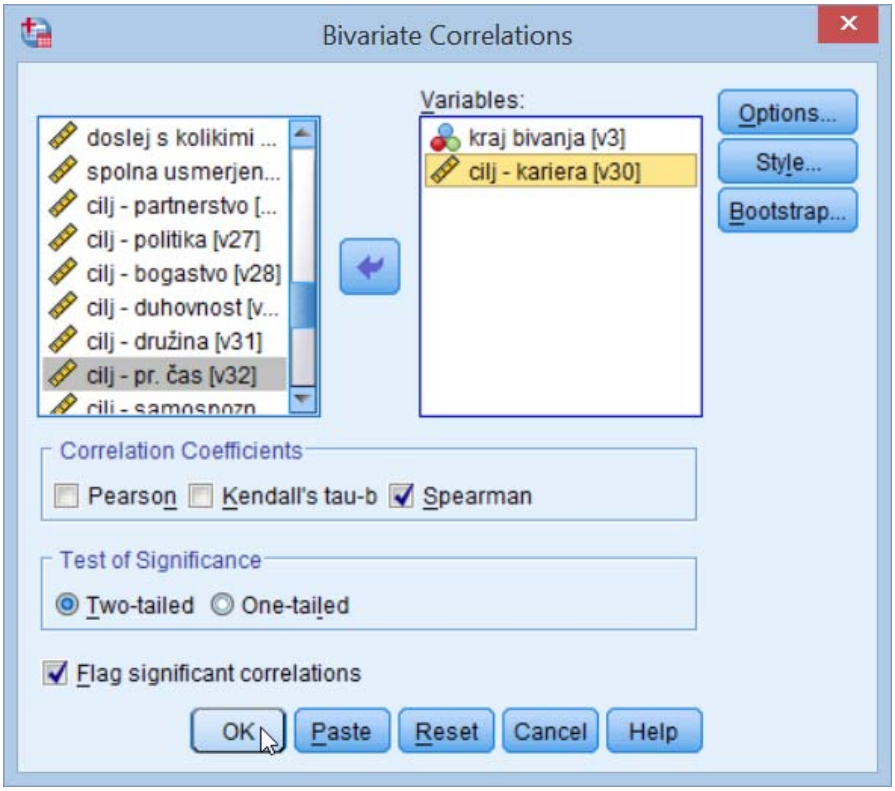
1. Kdaj uporabim Spearmanov koeficient korelacije (r_s ali ρ)in test njegove statistične pomembnosti?

Spearmanov koeficient korelacije uporabimo, ko želimo ugotoviti ali sta dve ordinalni spremenljivki (rangi) medsebojno povezani ter kakšna je smer te povezave. Če je r_s statistično pomemben (test njegove statistične pomembnosti), potem lahko (ob določenem tveganju, seveda) trdimo, da tudi v populaciji obstaja taka povezava kot smo jo ugotovili na vzorcu.

Ker je Spearmanov koeficient korelacije neparametrična statistika, ni potrebne predpostavke normalne porazdelitve podatkov za preverbo statistične pomembnosti r_s.

2. Kako najdem Spearmanov korelacijski koeficient v SPSS-u?

Odprite pogovorno okno Analyze → Correlate → Bivariate... V bistvu računamo Spearmanov korelacijski koeficient enako kot Pearsonovega, le da v pogovornem oknu Bivariate… odkljukamo Spearman.



3. Kako naj razumem output Spearmanovega korelacijskega koeficienta?

Oglejmo si output statistične pomembnosti Spearmanovega korelacijskega koeficienta za preverbo hipoteze: »Pri študentkah socialne pedagogike se pomembnost cilja doseganja uspeha v karieri pomembno negativno povezuje z velikostjo njihovega kraja bivanja.« Pazi, ničelna hipoteza pa – v nasprotju z raziskovalno - pravi, da ni povezave med tema dvema spremenljivkama. Vzorec predstavljajo študentke socialne pedagogike v študijskih letih 2003/04 ter 2004/05. Podatke smo analizirali iz dveh vprašanj: rang pomembnosti cilja oz. vrednote uspešnosti pri strokovnem delu in poklicni karieri ter velikost kraja bivanja.
Output zgleda takole.

Tabela 17: Spearmanov korelacijski koeficient med spremenljivkama kraja bivanja ter ocene pomembnosti cilja kariere v lastnem življenju

Nonparametric Correlations

Correlations

			kraj bivanja	cilj - kariera
Spearman's rho	kraj bivanja	Correlation Coefficient	1,000	-,047
		Sig. (2-tailed)	.	,690
		N	83	76
	cilj - kariera	Correlation Coefficient	-,047	1,000
		Sig. (2-tailed)	,690	.
		N	76	76

4. Kako navedem rezultate pomembnosti Spearmanovega korelacijskega koeficienta?

Pri študentkah socialne pedagogike nismo uspeli dokazati pomembne negativne povezave med oceno pomembnosti cilja oz. vrednote uspešnosti v poklicni karieri ter velikostjo njihovega kraja bivanja, $r = -.047$, p (dvostranski) $> 0,05$.

LITERATURA:

Ambrožič, F., Leskošek, B. (1999). *Uvod v SPSS (verzija 8.0 za Windows 95/NT).* Ljubljana: Fakulteta za šport: Inštitut za kineziologijo.

Cencič, M. (2002). *Pisanje in predstavljanje rezultatov raziskovalnega dela : kako se napiše in predstavi diplomsko delo (nalogo) in druge vrste raziskovalnih poročil.* Ljubljana : Pedagoška fakulteta.

Field, A. (2005). *Discovering Statistics Using SPSS*. London: Sage Publications.

Hinton, P. R., Brownlow, C., McMurray, I. & Cozens, B. (2004). *SPSS Explained*. London & New York: Routledge.

Mesec, B. (1997). *Metodologija raziskovanja v socialnem delu*. Ljubljana: VŠSD

Literatura, ki v tem delu ni bila uporabljena, jo pa prav tako kot zgornje priporočam:

Kožuh, Boris (2000). *Statistične obdelave v pedagoških raziskavah*. Ljubljana: Filozofska fakulteta, Oddelek za pedagogiko in andragogiko.

Rovan, J., Turk, T. (2001). *Analiza podatkov s SPSS za Windows*. Ljubljana: Ekonomska fakulteta.