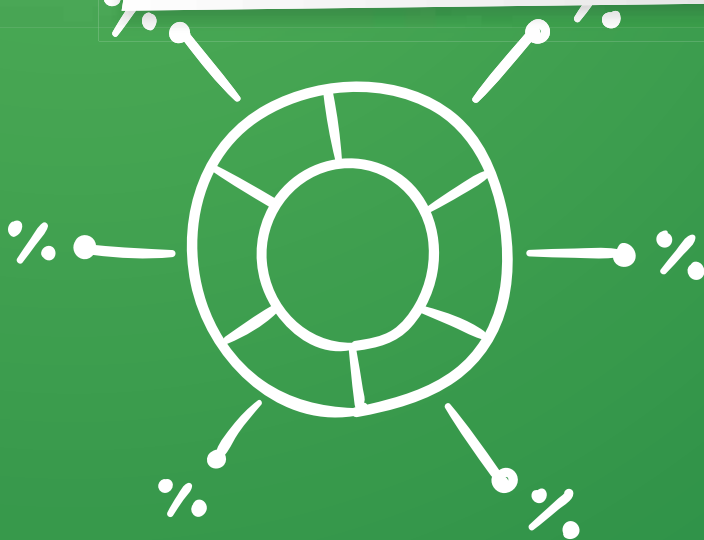




# Analiza podataka u SPSS-u

## Priručnik iz statistike

Slaven Živković





**Impresum****Autor**

Slaven Živković

**Naziv publikacije**

Analiza podataka u SPSS-u  
Priručnik iz statistike

**Izdavač**

De Facto Consultancy  
[www.defacto.me](http://www.defacto.me)

**Dizajn i priprema za štampu**

Bojan Tešić

**Tiraž**

500

Izradu ove publikacije je podržala Britanska ambasada Podgorica. Stavovi izrečeni u njoj nužno ne odražavaju stavove ambasade.



# Analiza podataka u SPSS-u

Priručnik iz statistike

---

Slaven Živković

**DEFACTO**  
CONSULTANCY

*Podgorica, 2015. godine*



# Par riječi na početku

Ova publikacija nastala je u okviru projekta “*Strengthening capacities of Parliament’s administration to prepare and evaluate public policies based on data analysis*”. Prikupljanje i obrada podataka neophodan je preduslov za pripremu kvalitetne osnove za donošenje odgovarajućih i djelotvornih javnih politika. Prvi korak ka tome je da podaci budu dostupni, a drugi da na kvalitetan način budu obrađivani.

Ovaj priručnik je nastao iz potrebe da se istraživačima pri sektoru za istraživanja Skupštine Crne Gore obezbijedi jednostavan i praktičan vodič za korišćenje jednog od najpopularnijih softverskih paketa za analizu i obradu podataka SPSS (zvanični naziv IBM SPSS statistics). Vjerujemo da će biti koristan i drugim istraživačima iz javnog i nevladinog sektora, studentima i svima onima koji se bave empirijskim analizama društvenih procesa. Na žalost, nije bilo moguće obraditi sve mogućnosti ovog programa, ali smatramo da je priručnik dobra osnova za početnike, i uvod u brojne mogućnosti za analizu podataka, koje program nudi.

Zahvalnost dugujemo britanskoj ambasadi u Podgorici koja je prepoznala potrebu za jačanjem istraživanja u službi javnosti. Takođe, zahvalni smo IBM-u koji je za potrebe pripreme priručnika odobrio korišćenje slika i prikaza programa. Na kraju, nadamo se da će ovaj priručnik doprinijeti popularizaciji upotrebe empirijom potvrđenih nalaza u svim sferama javnog života, te podići svijest o značaju pravilnog prikupljanja i korišćenja podataka.

*De Facto*

# SADRŽAJ

9	O programu
11	Formiranje baze
15	Kodiranje varijabli
19	Unošenje podataka
21	Funkcije za obradu podataka
23	Osnovne frekvencije (Frequencies)
31	Deskripcija (Descriptives)
35	Krostabulacije (Crosstabs)
43	Poređenje aritmetičkih sredina (Compare Means)
47	Transformisanje varijabli
49	Rekodiranje varijabli (Recode Variables)
55	Visual Binning
61	Izbor slučajeva

# PREGLED SLIKA

Slika 1:	<i>SPSS spreadsheet</i>	9
Slika 2:	<i>Dva SPSS sheet-a</i>	10
Slika 3:	<i>Primjer liste varijabli u Variable view sheet-u</i>	10
Slika 4:	<i>Variable view u SPSS-u</i>	11
Slika 5:	<i>Variable type prozor u SPSS-u</i>	12
Slika 6:	<i>Variable view poslije kreirane četiri varijable</i>	14
Slika 7:	<i>Otvaranje novog prozora za kodiranje varijabli</i>	15
Slika 8:	<i>Value Labels prozor u SPSS-u</i>	15
Slika 9:	<i>Value Labels poslije kodiranja prve varijable "pol"</i>	16
Slika 10:	<i>Kodirana varijabla "starost"</i>	16
Slika 11:	<i>Kodirana varijabla "zadovoljstvo"</i>	17
Slika 12:	<i>Otvaranje novog prozora za definisanje missing-a</i>	17
Slika 13:	<i>SPSS spreadsheet</i>	18
Slika 14:	<i>Kreirana baza u SPSS-u</i>	18
Slika 15:	<i>Baza sa unijetim odgovorima ispitanika</i>	19
Slika 16:	<i>Baza sa unijetim odgovorima, kao labels, ispitanika</i>	19
Slika 17:	<i>Analyze menu u SPSS-u</i>	21
Slika 18:	<i>Pod-meni Descriptive statistics</i>	22
Slika 19:	<i>Frequencies box u SPSS-u</i>	23
Slika 20:	<i>Polje Statistics u Frequencies box-u</i>	23
Slika 21:	<i>Polje Charts u Frequencies box-u</i>	25
Slika 22:	<i>Polje Format u Frequencies box-u</i>	25
Slika 23:	<i>Frequencies box u SPSS-u</i>	26
Slika 24:	<i>Polje Statistics u Frequencies box-u</i>	27
Slika 25:	<i>Polje Charts u Frequencies box-u</i>	27
Slika 26:	<i>Output prozor u SPSS-u</i>	28
Slika 27:	<i>Deskriptivna statistika u Output-u SPSS-u</i>	28
Slika 28:	<i>Frequency table u Output-u SPSS-a</i>	29
Slika 29:	<i>Frequency table string varijable u Output-u SPSS-a</i>	30
Slika 30:	<i>Grafik u Output-u SPSS-a</i>	30
Slika 31:	<i>Descriptives polej u SPSS-u</i>	31
Slika 32:	<i>References u SPSS-u</i>	31



Slika 33: Polje Options u Descriptives box-u	32
Slika 34: Descriptives tabele u Output-u	33
Slika 35: Crosstabs opcija u SPSS-u	35
Slika 36: Crosstabs prozor u SPSS-u	35
Slika 37: Statistics opcija u Crosstabs	36
Slika 38: Cells opcija u Crosstabs-u	37
Slika 39: Crosstabs window	37
Slika 40: Tabela kontigencije/ukrštanja u Output-u SPSS-a	38
Slika 41: Chi-Square Tests tabela u Output-u SPSS-a	38
Slika 42: Bar chart ukrštanja u Output-u SPSS-a	39
Slika 43: Crosstabs prozor u SPSS-u	40
Slika 44: Three-way Crosstabulation u SPSS-u	40
Slika 45: Chi-Square tests za three-way crosstabulation	41
Slika 46: Analyze – Compare Means – Means	43
Slika 47: Means prozor (window) u SPSS-u	44
Slika 48: Options polje u prozoru Means	44
Slika 49: Means window u SPSS-u	45
Slika 50: Compare Means tabele u Output-u SPSS-a	45
Slika 51: Transform u taskbar-u SPSS-a	47
Slika 52: Recode into Different Variables	49
Slika 53: Recode into Different Variables prozor u SPSS-u	50
Slika 54: Recode into Different Variables prozor poslije određivanja karakteristika nove varijable	50
Slika 55: Recode into Different Variables: Old and New Values	51
Slika 56: Old and New Values prozor poslije rekodiranja	52
Slika 57: Variable View poslije rekodiranja varijabli i kreiranja nove varijable	53
Slika 58: Values za rekodiranu varijablu	54
Slika 59: Tabela frekvencija u Output-u za novoformiranu rekodiranu varijablu	54
Slika 60: Visual binning opcija u SPSS-u	55
Slika 61: Visual Binning menu	55
Slika 62: Visual Binning	56
Slika 63: Make Cutpoints polje u Visual Binning	57
Slika 64: Equal Percentiles Based on Scanned Cases box u Visual Binning-u	58
Slika 65: Visual binning prozor poslije definisanja karakteristika nove varijable	59
Slika 66: Variable View prozor poslije kreiranja Binned varijable	59
Slika 67: Tabele frakvencije za novoformiranu varijablu Zadovoljstvo životom (Binned)	60
Slika 68: Data -> Select Cases	61
Slika 69: Select cases prozor u SPSS-u	62
Slika 70: Polje If u Select Cases	62
Slika 71: Select Cases –if sa filter formulom	64
Slika 72: Data View sheet sa uključenim filterom	65
Slika 73: Tabela frekvencije za varijablu Pol ispitanika pri uključenom filteru	65
Slika 74: Filter On polje u SPSS-u	65
Slika 75: Brisanje varijable filter	66
Slika 76: Isključivanje filtera u SPSS-u	67



# O PROGRAMU

SPSS (*Statistical Package for Social Sciences*) – jedan je od danas najčešće korišćenih programa za obradu podataka. I pored toga što nosi ime “statistički paket za društvene nauka”, oslikavajući primarno i originalno tržište kojem je bio namijenjen, danas se koristi i u drugim naukama i oblastima, poput medicine, marketinga i ostalih. Program od 2009.godine proizvodi američka kompanija IBM (*International Business Machines Corporation*) i danas predstavlja samo dio familije softverskih proizvoda ove kompanije namijenjenih prikupljanju, čuvanju i obradi podataka.

SPSS je samo jedan od paketa koji se koristi u svrhe obrade podataka. Pored ovog, koriste se još i STATA, R, Epi info i mnogi drugi, ali smatramo da je upravo SPSS najlakši za korišćenje, iz prostog razloga što koristi tako-zvani point-and-click interfejs: procedure su već u meniju programa, a na korisniku je da ih samo izabere.

U SPSS-u su podaci predstavljani tabelarno u spreadsheet-u (slično kao i u excel-u), a baze podataka su u .sav formatu. Svaka baza se sastoji od varijabli i podataka. Varijable su zapravo pitanja koja su postavljena, dok pregledom podataka možemo vidjeti konkretne vrijednosti po svakoj varijabli za svaku opservaciju. U zavisnosti od tipa istraživanja, opservacije mogu biti pojedinci (ispitanici u istraživanjima javnog mnjenja), opštine, države (pr. agregatna statistika), firme (pr. istraživanja kompanija) i ostalo. U SPSS-u opservacije su predstavljene u redovima, a varijable u kolonama.

	Broj_anquete	Grad	Mjesna_zajec nica	v001	D_2	D_3	D_4	D_5	D_6	D_7
1	1187	Bijelo Polje	Fond Zdrav...	Muski	18	Srednje	12	Islamska	Crnogorac	Nijesam za...
2	1190	Bijelo Polje	OŠ Patrk	Muski	41	Osnovno o...	8	Pravoslav...	Crnogorac	Nezaposle
3	1195	Bijelo Polje	OŠ Patrk	Muski	59	Srednje	11	Pravoslav...	Crnogorac	Nijesam za...
4	1165	Bijelo Polje	OŠ Vrela ...	Muski	65	Osnovno o...	8	Pravoslav...	Crnogorac	Penzioner
5	1168	Bijelo Polje	OŠ Vrela ...	Muski	65	Osnovno o...	8	Pravoslav...	Crnogorac	Penzioner
6	1188	Bijelo Polje	Fond Zdrav...	Ženski	25	Srednje	12	Islamska	Crnogorac	Nezaposle
7	1163	Bijelo Polje	OŠ Vrela ...	Ženski	30	Srednje	13	Pravoslav...	Crnogorac	Zaposlen ...
8	1196	Bijelo Polje	OŠ Patrk	Ženski	35	Osnovno o...	8	Pravoslav...	Crnogorac	Nezaposle
9	1164	Bijelo Polje	OŠ Vrela ...	Ženski	57	Srednje	Bez odgov...	Pravoslav...	Crnogorac	Penzioner ...
10	1158	Bijelo Polje	OŠ Vrela ...	Muski	20	Srednje	12	Pravoslav...	Srbin	Nezaposle
11	1193	Bijelo Polje	OŠ Patrk	Muski	37	Osnovno o...	8	Pravoslav...	Srbin	Nezaposle
12	1137	Bijelo Polje	OŠ u Cokrl...	Muski	49	Srednje	12	Pravoslav...	Srbin	Nezaposle
13	1176	Bijelo Polje	Fond Zdrav...	Muski	43	Srednje	12	Pravoslav...	Srbin	Nezaposle
14	1135	Bijelo Polje	OŠ u Cokrl...	Muski	40	Srednje	12	Pravoslav...	Srbin	Nezaposle
15	1174	Bijelo Polje	Fond Zdrav...	Muski	38	Vise obraz...	15	Pravoslav...	Srbin	Nezaposle
16	1154	Bijelo Polje	OŠ Vrela ...	Muski	53	Srednje	11	Pravoslav...	Srbin	Nesto drug
17	1145	Bijelo Polje	OŠ u Podi...	Muski	46	Vise obraz...	14	Pravoslav...	Srbin	Nesto drug
18	1142	Bijelo Polje	OŠ u Podi...	Muski	75	Osnovno o...	8	Pravoslav...	Srbin	Penzioner
19	1136	Bijelo Polje	OŠ u Cokrl...	Muski	65	Osnovno o...	8	Pravoslav...	Srbin	Penzioner
20	1150	Bijelo Polje	OŠ u Podi...	Muski	56	Srednje	11	Pravoslav...	Srbin	Nezaposle
21	1151	Bijelo Polje	OŠ u Podi...	Ženski	30	Srednje	12	Pravoslav...	Srbin	Nezaposle

Slika 1: SPSS spreadsheet

Brojevi 1, 2, 3...su unosi različitih anketa, koje su obilježene svojim brojem kao što možemo vidjeti u spreadsheet-u. Riječ je o odgovorima ispitanika u tipičnom javnomnjenjskom istraživanju. U kolonama su varijable (Broj ankete; Grad; Mjesna zajednica), to jest pitanja iz upitnika. Redova ima onoliko koliko ima opservacija (ispitanika), dok kolona ima onoliko koliko ima varijabli (pitanja). SPSS sadrži dva sheet-a, koji se kao i u excel-u nalaze u donjem lijevom uglu.

Case Number	Broj ankete	Grad	Mjesna zajednica	Pol	Starost	Obrazovanje	Ukupan broj za
34	1159	Bijelo Polje	OS Vrela ...	Muski	19	Srednje	Bez odgov...
35	1180	Bijelo Polje	Fond Zdrav...	Muski	18	Srednje	12
36	1133	Bijelo Polje	OS u Godij...	Muski	42	Osnovno o...	8
37	1199	Bijelo Polje	OS u Godij...	Muski	40	Osnovno o...	0

Slika 2: Dva SPSS sheet-a

U *Data View sheet*-u možemo vidjeti sve podatke. U redovima su opservacije, u kolonama su varijable. S druge strane *Variable View* sadrži listu svih varijabli koje imamo u bazi. U Variable view prozoru unosimo nove varijable, kada formiramo bazu, i svaku od varijabli, možemo da definišemo po 11 karakteristika, koliko kolona ima ovaj SPSS sheet.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Broj_anketa	Numeric	9	0	Broj ankete	None	None	9	Right	Scale	Input
2	Grad	Numeric	8	0	Grad	{1, Bijelo Po...	None	8	Right	Scale	Input
3	Mjesna_zaj	String	300	0	Mjesna zajednica	None	None	8	Left	Nominal	Input
4	v001	Numeric	1	0	Pol	{1, Muski}...	9	8	Right	Nominal	Input
5	D_2	Numeric	2	0	Starost	{0, Bez odg...	0	8	Right	Scale	Input
6	D_3	Numeric	1	0	Obrazovanje	{1, Bez obra...	9	8	Right	Scale	Input
7	D_4	Numeric	2	0	Ukupan broj za	{99, Bez od...	99	8	Right	Scale	Input
8	D_5	Numeric	1	0	Ukupan broj za	{1, Bez odg...	9	8	Right	Scale	Input

Slika 3: Primjer liste varijabli u Variable view sheet-u

# FORMIRANJE BAZE

U *variable view* prozoru kreiramo varijable. Kreiranje varijabli, i formiranje baze nastojaćemo da objasnimo na jednom primjeru. Pretpostavimo da želimo ispitati demografske razlike među grupom ispitanika. Želimo da provjerimo da li su naši ispitanici mušakrci ili žene; koliko imaju godina, kojoj nacionalnosti pripadaju; i kog su obrazovanja. Za to je potrebno da kreiramo četiri varijable:

- Pol
- Starost
- Zadovoljstvo životom
- Mišljenje o nacionalizmu

Svaki red u *variable view* označava jednu varijablu, dok u u kolonama te varijable definišemo po raznim karakteristikama.

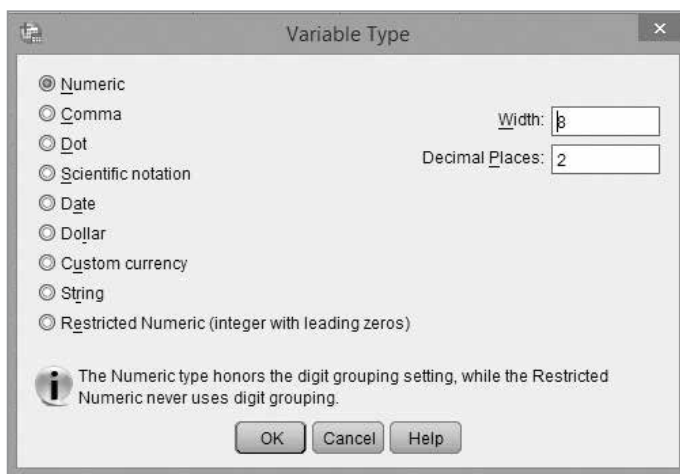
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1											
2											
3											
4											
5											

Slika 4: *Variable view* u SPSS-u

Da bismo kreirali novu varijablu selektujemo prazno polje u *variable view*, u meniju SPSS-a idemo na *edit* i izaberemo opciju *insert variable* (dodaj varijablu).

Varijable u SPSS-u možemo odrediti po 11 karakteristika. Od njih, pet najvažnijih i najčešće korišćenih su:

- Name (ime) – u ovoj koloni možete unijeti ime varijable. Ono što upišete u ovom polju pojavice se na vrhu kolone u *data view sheet*-u. Ime varijable ne smije sadržati nijedan razmak (*space*). Ako bi ime varijable sadržalo više od dozvoljenog broja karaktera ili neki razmak, program bi Vam prijavio grešku kada biste pokušali da pređene na drugu ćeliju.
- Type (tip) – u SPSS-u imate mogućnost da unesete različite tipove podataka. Najčešće je varijabla (a i po *default*-u) označena kao *numeric* (numerička), što znači da je definisana preko brojeva. Druga opcija koja je takođe često u upotrebi je *string* varijabla, što je skraćenica za *string of letters* (niz slova), i ovu opciju koristimo kada je varijabla definisana preko slova (na primjer, ako je varijabla takva da želimo upisati nečije ime, nečiji odgovor koji nije kodiran brojevima). Kada koristimo *string* varijablu u polju *width* definišemo do koliko maksimum karaktera može imati ono što upisujemo u toj varijabli.



Slika 5: Variable type prozor u SPSS-u

- Label (karakteristika) – SPSS ograničava ime varijable na određeni broj karaktera (u starijim verzijama programa to je 8). Međutim u polju *label* varijabli možemo dati detaljniji opis. Riječ je o izuzetno pogodnoj opciji jer se pitanja najčešće imenuju skraćenicama (na primjer var1, var2), a u polju *label* određujemo koje je to pitanje (na primjer “u kojoj mjeri se slažete sa stavom da će ulaskom u Evropsku uniju privreda Crne Gore profitirati”), i kasnije u opcijama za obradu podataka varijable možemo pretraživati po *label*-ima a ne po, najčešće, šturim imenima varijabli, što izuzetno olakšava rad u program, i zato je uvijek poželjno odrediti i *label* varijable.
- Values (vrijednosti) – u ovoj koloni dodjeljujemo brojeve, za određene grupe ljudi, ili odgovora. Ponuđene odgovore iz upitnika kodiramo kako bismo lakše unosili podatke i radili obradu podataka. O načinu kodiranja, i određivanju values-a detaljnije ćemo u sljedećoj podsekciji.
- Missing (nedostajuće vrijednosti) – u ovoj koloni određujemo nedostajuće vrijednosti, to jest one *values* koje iz nekog razloga u datom trenutku ne želimo da uključimo u obradu. I o ovoj opciji će biti više riječi u narednom dijelu, kao i docnije kada budemo predstavljali opcije za obradu podataka.

Kada kreirate varijablu SPSS automatski određuje da njene vrijednosti imaju dvije decimale. U polju *decimals* možemo smanjivati i povećavati broj *decimal* po potrebi. Nije od suštinske važnosti, pravilno definisanje broja decimala će uticati na bolje formatiranje podataka prilikom obrade; (npr. Ukoliko za godine starosti definišemo da imaju dvije decimale, iako nam one nisu potrebne, prilikom računanja prosječne starosti, dobićemo aritmetičku sredinu sa 4 decimale, što je bespotrebno i zbunjujuće). Dodatno u polju *measure* možemo odrediti da li je varijabla *nominalna*, *ordinalna* ili *intervalna*.

- *Nominalne varijable*, odnosno preciznije nominalne skale, predstavljaju one skale na kojima su određenim kvalitetima, odnosno svojstvima date numeričke vrijed-

nosti. Skala je specifična po tome što između numeričkih vrijednosti, odnosno brojeva, ne možemo uspostaviti neki redoslijed, to jest utvrditi kakav je kvantitativan odnos. Na primjer, kada na pitanju o polu ispitanika, u nekom istraživanju, ženski pol označimo sa 1, a muški sa 2, mi smo ovim atributima dali samo numeričke vrijednosti, radi lakše obrade podataka, a one nemaju pored toga nikakva kvantitativna svojstva.

- *Ordinalne varijable*, odnosno ordinalne skale su one skale, gdje vrijednosti imaju numeričke oznake koje stoje u određenom redoslijedu (rangu), to jest posložene sup o određenoj hijerarhiji. Međutim kod ordinalnih skala ne znamo kolika je razlika između vrijednosti na skali. Na primjer, kada bismo varijablu o godinama ispitanika u nekom istraživanju definisali preko ordinalne skale, gdje bi vrijednost 1 bili mladi (od 18 do 34 godine, na primjer), vrijednost 2 srednjovječni (od 35 do 54 godine), a vrijednost 3 stariji (55+ godina), vidimo da među numeričkim vrijednostima postoji jasan redoslijed, to jest hijerarhija. Međutim između vrijednosti 1 i 2, razlika može biti dvije godine (ako jedan ispitanik ima 33, a drugi 35 godina, oni pripadaju različitim kategorijama, a razlika između njih je dvije godine), ili 32 godine (ako jedan ispitanik ima 18 godina, a drugi 50). I u jednom i u drugom slučaju prvi ispitanik pripada kategoriji 1, a drugi kategoriji 2, i iako je razlika između dva data primjera velika, sa aspekta ordinalne skale, u oba slučaja prvi je u kategoriji 1, a drugi ispitanik u kategoriji 2.
- *Intervalne varijable*, odnosno intervalne skale, su numeričke, i vrijednosti na intervalnoj skali, ne samo da su poređane po određenom redoslijedu (hijerarhijski), već, za razliku od ordinalne, razlike između vrijednosti na skali su jednake. Ako godište ispitanika definišemo preko intervalne skale, razlike između njih će uvijek biti jasne, to jest, razlika između jednog ispitanika sa 50 godina, i drugog sa 35, ista je kao i razlika između jednog ispitanika sa 40 godina, i drugog sa 25.

Za potrebe našeg primjera kreiraćemo četiri varijable. Prvu varijablu "pol" ćemo nazvati (name) upravo tako – "pol". To će biti numerička varijabla, pa u polju *type* zadržavamo opciju *numeric*, dok u polju *label* unosimo pitanje onako kako je definisano upitnikom na primjer "Pol ispitanika". Određivanje *values* i *missing* ćemo objasniti u sljedećoj sekciji. Dodatno za varijablu *pol*, dodaćemo još, u polju *measure*, da je to nominalna varijabla (*nominal*). Sličan postupak ponovićemo i za varijable "starost" i "zadovoljstvo". I te dvije varijable su numeričke. Varijablu "starost" ćemo napraviti kao ordinalnu, dok je varijabla "zadovoljstvo" intervalna (*scale*), s obzirom na to da ćemo pretpostaviti da su ispitanici ocjenjivali zadovoljstvo životom na desetostepenoj skali od 1 do 10, gdje 1 znači "u potpunosti nezadovoljan", a 10 znači "u potpunosti zadovoljan".

Za četvrtu varijablu, definisanu upitnikom kao "otvoreno" pitanje (pitanje bez ponuđenih odgovora) unosimo jednu malu promjenu u odnosu na prve tri. Kada odredimo ime varijable (na primjer – mišljenje), i njen *label* (na primjer – mišljenje o nacionalizmu) u polju *type* ovu varijablu ćemo označiti kao *string*, da bismo mogli da unesemo cijeli odgovor ispitanika, važno je da je "povećamo", to jest da unese-

mo željeni maksimalni broj karaktera koje možemo unijeti na ovoj varijabli. U polju *width*, umjesto broja 8 (koji je po *default*-u) unosimo na primjer 200. To je broj karaktera koji možemo unijeti po jednoj opservaciji na ovoj varijabli.

Poslije ovoga, Vaš *variable view sheet* u SPSS-u bi trebalo da izgleda ovako.

	Ime	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Pol	Numeric	8	0	Pol ispitanika	None	None	8	Right	Nominal	Input
2	Starost	Numeric	8	0	Starost ispitanika	None	None	8	Right	Ordinal	Input
3	Zadoljstvo	Numeric	8	0	Zadoljstvo životom	None	None	8	Right	Scale	Input
4	Mišljenje	String	200	0	Mišljenje ispitanika o nacionalizmu	None	None	8	Left	Nominal	Input
5											
6											

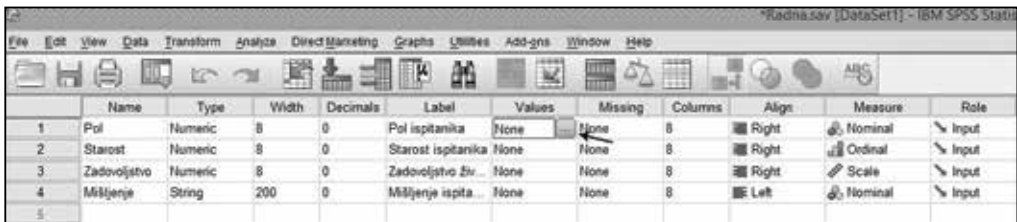
Slika 6: *Variable view* poslije kreirane četiri varijable



# KODIRANJE VARIJABLI

Kodiranje varijabli je proces u kojem ponuđenim odgovorima u upitniku, grupama ljudi, kvalitativnim odgovorima, dajemo numeričke kodove, kako bismo ih grupisali zajedno i kasnije lakše obrađivali. Varijable kodiramo i da bismo olakšali unos podataka u SPSS-u.

Kodiranje varijabli i definisanje nedostajućih vrijednosti (*missing*) objasnićemo kroz naš primjer četiri varijable. Najprije ćemo kodirati varijablu "pol". Na toj varijabli imaćemo dvije opcije: muškarci i žene. Za potrebe kodiranja sve muškarce označićemo kodom 1, a sve žene kodom 2. Dodatno, ukoliko na ovom pitanju nedostaje odgovor (što je rijetkost, ali se dešava), tu opciju označićemo sa 9 – nema odgovora. Da bismo to uradili selektujemo polje *values*, i preko tri tačke u desnom uglu otvorimo novi prozor.



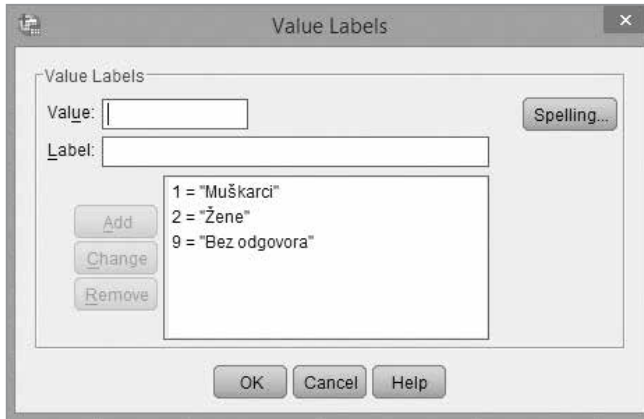
Slika 7: Otvaranje novog prozora za kodiranje varijabli

Pošto izaberemo "tri tačkice" u polju *values* trebalo bi da se otvori sljedeći prozor.



Slika 8: Value Labels prozor u SPSS-u

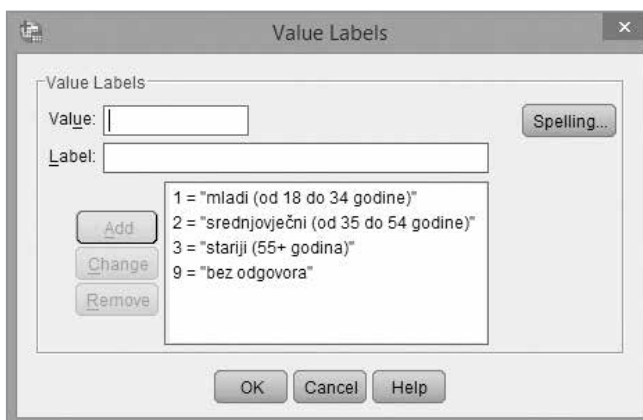
U polju *Value* upisujemo kodove kojima želimo da kodiramo određeni odgovor. Kao što smo naveli kod 1 će biti muškarci. Tako u polju *Value* upisujemo 1, a u polju *Label* upisujemo muškarci, i biramo opciju *add*. Postupak ponovimo i za žene (koje kodiramo sa 2), kao i za opciju bez odgovora (koju kodiramo sa 9). Poslije toga bi *Value Labels* prozor trebalo da izgleda ovako.



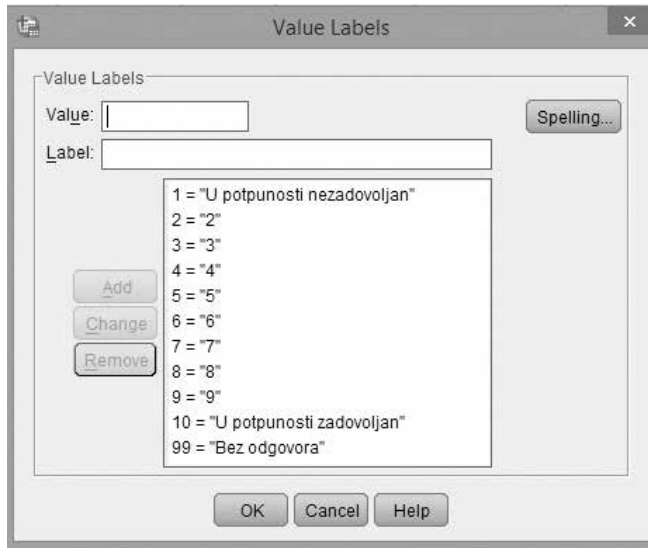
Slika 9: *Value Labels* poslije kodiranja prve varijable "pol"

Promjene na već unijetim kodovima unosimo tako što u istom prozoru izaberemo kod, izvršimo promjenu, i potvrdimo je selektujući opciju *Change*. Da bismo izbrisali neki od dodatih kodova, sve što treba da uradimo je da izaberemo taj kod, i selektujemo opciju *Remove*.

Isti postupak ponovićemo i za drugu i treću varijablu. Na primjer varijablu "starost" smo kodirali tako da je opcija 1 "mladi (od 18 do 34 godine)"; opcija 2 – "srednjovječni (od 35 do 54 godine)"; i opcija 3 – "stariji (55+ godina)"; i dodatno opcija 9 – "bez odgovora". Zadovoljstvo životom ćemo kodirati na sljedeći način: 1 – "U potpunosti nezadovoljan", 2 – "2"; 3 – "3"; 4 – "4"; 5 – "5"; 6 – "6"; 7 – "7"; 8 – "8"; 9 – "9"; 10 – "u potpunosti zadovoljan"; 99 – "bez odgovora". Četvrtu varijablu ne možemo kodirati, s obzirom na to da je riječ o otvorenom pitanju.



Slika 10: Kodirana varijabla "starost"



Slika 11: Kodirana varijabla "zadovoljstvo"

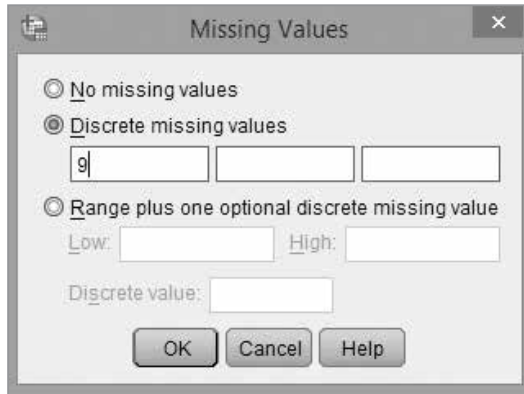
Za potrebe obrade podataka interesovaće nas, za ova konkretna pitanja, samo oni koji su dali neki od ponuđenih odgovora. To jest, željećemo da vidimo distribuciju po polu ukoliko ispitanike posmatramo kao cijelo (kao 100%) i da zbir onih koji su dali odgovor bude 100%. Da bismo to uradili potrebno je definisati *missing* (nedostajuće) vrijednosti. To radimo tako što otvorimo novi prozor selektujući polje pa "tri tačkice" u koloni *missing*.<sup>1</sup>



Slika 12: Otvaranje novog prozora za definisanje *missing*-a

U sljedećem prozoru označimo polje *Discrete missing values* odredimo koje vrijednosti želimo da isključimo. U našem primjeru na varijabli *pol* to će biti opcija 9 – "bez odgovora".

<sup>1</sup> Nemaju sva pitanja i uvijek definisane nedostajuće vrijednosti na varijabli. Nekada u obradi podataka koristimo sve ponuđene opcije kao odgovore, pa i opcije „nema odgovora“.



Slika 13: SPSS spreadsheet

Isto će biti za varijablu “starost”, dok ćemo kod treće varijable, “zadovoljstvo” opciju 99 definisati kao *missing*. Za četvrtu varijablu koja je pitanje otvorenog tipa, bez ponuđenih odgovora, i bez kodiranih opcija, ne definišemo nedostajuće vrijednosti. Naša baza bi sada trebalo da izgleda ovako.

	Ime	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Pol	Numeric	6	0	Pol ispitanika	{1, Muškarci}	8	8	Right	Nominal	Input
2	Starost	Numeric	8	0	Starost ispitanika	{1, mlađi (od 18 do 34 ...}	8	8	Right	Ordinal	Input
3	Zadovoljstvo	Numeric	8	0	Zadovoljstvo životom	{1, U potpunosti nezad. ...}	88	8	Right	Scale	Input
4	Mjerenje	String	200	0	Mjerenje ispitivanja o nacionalizmu	None	None	8	Left	Nominal	Input

Slika 14: Kreirana baza u SPSS-u

Poslije ovih koraka, pošto smo dodali varijabli, odredili njihova imena i karakteristike, kodirali ih, i dodatno definisali nedostajuće vrijednosti (*missing*), baza je spreman za unošenje podataka.

# UNOŠENJE PODATAKA

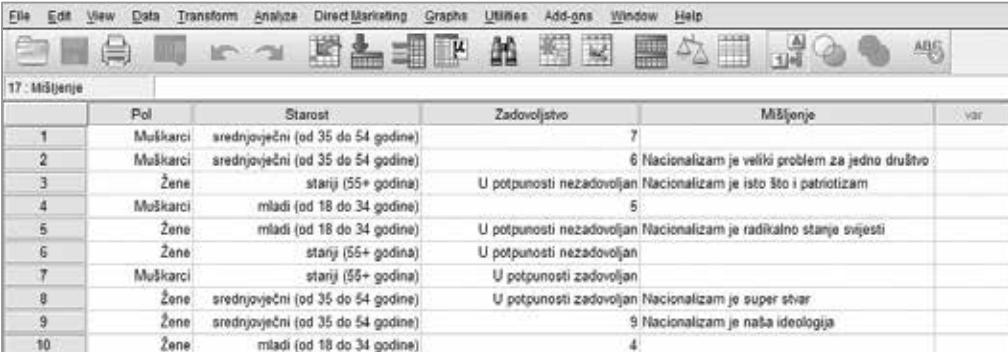
Unošenje podataka je veoma jednostavan proces, koji ponekad zna veoma dugo da traje. Podatke unosimo u *Data view sheet*-u. Ono što je veoma važno za brže unošenje podataka jeste da su varijable u programu pravilno kodirane. Na taj način, ukoliko unosimo podatke sa upitnika, sve što treba da radimo jeste da redom upisujemo brojeve, kojim su na upitniku označeni ponudeni odgovori, identično kao da to radimo u *excel*-u. Kao što smo već napomenuli svaki red je jedna opservacija (u ovom slučaju ispitanik), a svaka kolona je jedna varijabla (u našem primjeru pitanje iz upitnika). Za potrebe daljeg objašnjavanja opcija u SPSS-u unijet ćemo deset nasumičnih ispitanika koji su popunili naša četiri pitanja. Naša baza bi tada trebalo da izgleda ovako.



	Pol	Starost	Zadovoljstvo	Mišljenje	var
1	1	2		7	
2	1	2		6 Nacionalizam je veliki problem za jedno društvo	
3	2	3		1 Nacionalizam je isto što i patriotizam	
4	1	1		5	
5	2	1		1 Nacionalizam je radikalno stanje svijesti	
6	2	3		1	
7	1	3		10	
8	2	2		10 Nacionalizam je super stvar	
9	2	2		9 Nacionalizam je naša ideologija	
10	2	1		4	

Slika 15: Baza sa unijetim odgovorima ispitanika

Kao što možemo vidjeti u polja smo unosili samo one brojeve kako smo kodirali odgovore na definisanim varijablama. Novije verzije SPSS-a pružaju mogućnost da već prilikom unosa vidimo umjesto *values* (kodova) vidimo *values* (vrijednosti koje smo numerički kodirali). Za to je potrebno da selektujemo opciju u *taskbar*-u koja automatski, čak i prilikom unosa, pretvara numeričke kodove u vrijednosti kojima su ti kodovi dodijeljeni. Poslije toga, naša baza, bi trebalo ovako da izgleda:



	Pol	Starost	Zadovoljstvo	Mišljenje	var
1	Muškarci	srednjoječni (od 35 do 54 godine)		7	
2	Muškarci	srednjoječni (od 35 do 54 godine)		6 Nacionalizam je veliki problem za jedno društvo	
3	Žene	stariji (55+ godina)	U potpunosti nezadovoljan	Nacionalizam je isto što i patriotizam	
4	Muškarci	mładi (od 18 do 34 godine)		5	
5	Žene	mładi (od 18 do 34 godine)	U potpunosti nezadovoljan	Nacionalizam je radikalno stanje svijesti	
6	Žene	stariji (55+ godina)	U potpunosti nezadovoljan		
7	Muškarci	stariji (55+ godina)	U potpunosti zadovoljan		
8	Žene	srednjoječni (od 35 do 54 godine)	U potpunosti zadovoljan	Nacionalizam je super stvar	
9	Žene	srednjoječni (od 35 do 54 godine)		9 Nacionalizam je naša ideologija	
10	Žene	mładi (od 18 do 34 godine)		4	

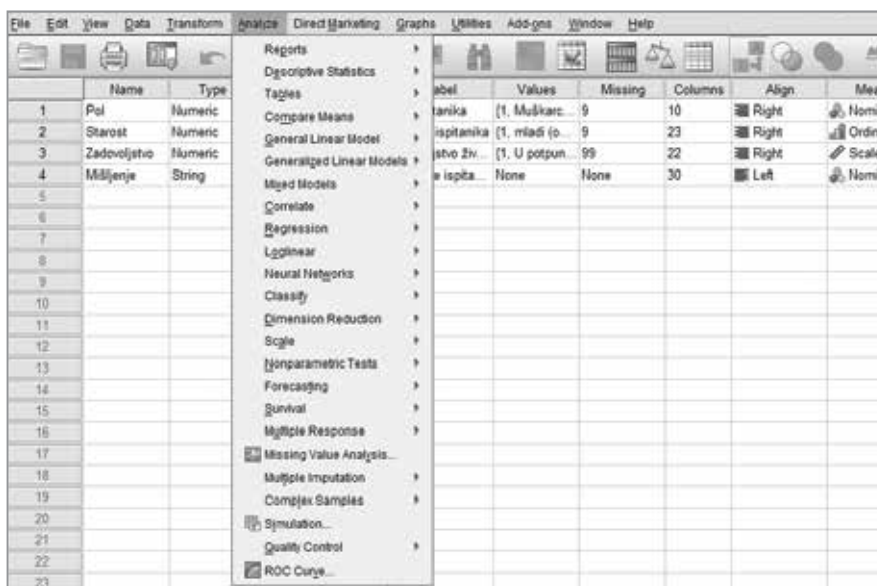
Slika 16: Baza sa unijetim odgovorima, kao labels, ispitanika

SPSS nam neće dozvoliti da u nekoj koloni unesemo broj, to jest kod, kad prije toga nije definisan u polju *Values* u *sheet*-u *Data view*. Čim unesemo neki kod kojeg nema u programu, i pokušamo da pređemo dalje, program će nas upozoriti da je unos nevalidan. Jedino na četvrtoj varijabli možemo da pišemo sve što želimo, s obzirom na to da smo je unaprijed definisali kao string varijablu.

Dodatno u ovom dijelu važno je napomenuti da je SPSS izuzetno *user-friendly* program, ne samo zbog toga što koristi tako-zvani *point-and-click* interfejs, već i zato što osnovne opcije u velikoj mjeri liče, ili su u nekim djelovima čak i potpuno iste kao i osnovne opcije u *microsoft office* porodici programa (*word*, *excel*, *power point*...), koje *windows* korisnici svakodnevno koriste. U osnovnom *taskbar*-u, meni *File* pruža mogućnosti da sačuvamo našu bazu, preko osnovnih opcija *Save* i *Save as*, kao i da otvorimo novu bazu preko opcije *Open – Data*.

# FUNKCIJE ZA OBRADU PODATAKA

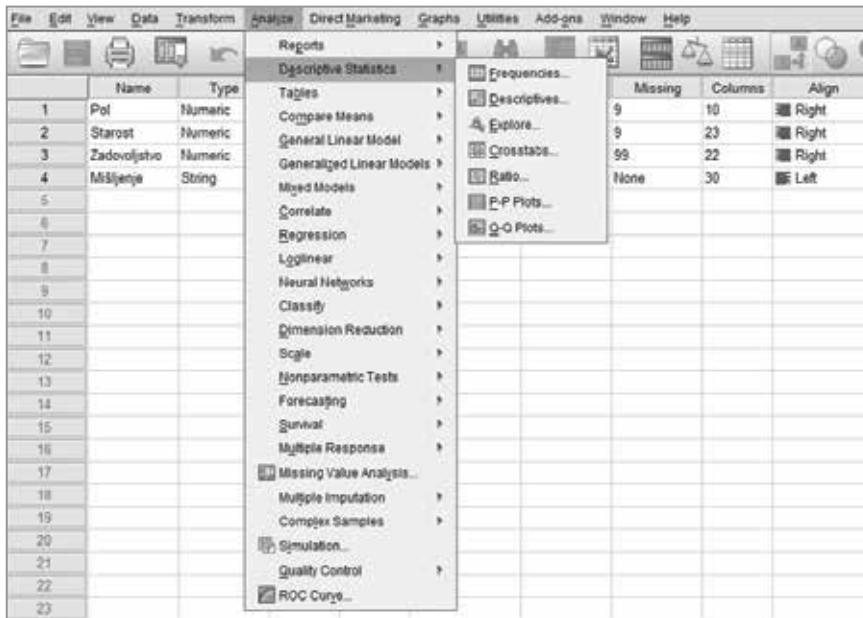
Podatke obrađujemo (analiziramo, ukrštamo...) u polju *Analyze* u *Taskbar*-u.



Slika 17: *Analyze* menu u SPSS-u

Meni *Analyze* pruža veliki broj mogućnosti za najrazličitije statističke postupke koje možemo primjenjivati na podacima. Za potrebe ovog priručnika zadržaćemo se samo na najosnovnijim funkcijama, pomoću kojih možemo raditi deskriptivnu statistiku, odnosno one metode i procedure koje nam omogućavaju sumiranje i prezentovanje podataka. Riječ je obično o prvom koraku u analizi podataka, koja služi opisivanju podataka. Deskripcija može da prethodi statističkom zaključivanju i predviđanju, ali može biti i kraj statističke analize.

Opcije na koje će poseban akcenat biti u ovom priručniku, nalaze se u pod-meniju *Descriptive statistics*.



Slika 18: Pod-meni Descriptive statistics

Za potrebe ovog priručnika redom ćemo analizirati tri opcije iz ovog pod-menija:

- Dobijanje osnovnih frekvencija (*Frequencies*)
- Deskripcija (*Descriptives*)
- Krostabulacije, to jest ukrštanja/kontingencije (*Crosstabs*)



# OSNOVNE FREKVENCIJE (*FREQUENCIES*)

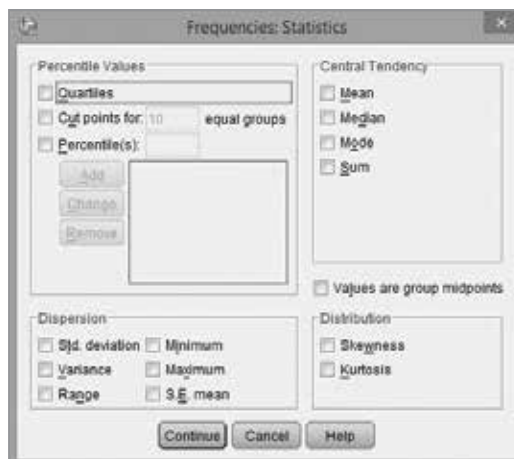
Osnovne frekvencije nam služe da istražimo distribuciju na određenim varijablima (pitanjima) i da „prebrojimo“ grupisane odgovore. Preko polja *Analzye – Descriptive Statistics – Frequencies* stižemo do sljedećeg *box*-a:



Slika 19: *Frequencies box* u SPSS-u

Najjednostavniji postupak je zapravo prebaciti varijable za koje želimo da vidimo sumirane skorove i ukupnu distribuciju odgovora na izabranim varijabalama. Te varijable selektujemo, prebacimo u polje *Variable(s)* na desnoj strani, i selektujemo polje "OK". Prije nego predstavimo rezultate koje dobijamo u *Output*-u analiziraćemo dodatne opcije koje nam *Frequencies box* nudi u desnom dijelu (*Statistics; Charts; Format*).<sup>2</sup>

## Statistics



Slika 20: Polje *Statistics* u *Frequencies box*-u

<sup>2</sup> *Bootstrap* opciju, koja omogućava testiranje analitičkih modela i njihove pouzdanosti, nećemo dodatno obrađivati u ovom priručniku

U ovom polju možemo štrahirati operacije koje želimo da program izračuna i prikaže za izabrane varijable.

U polju *Percentile Values* određujemo percentile koje želimo da vidimo u *Output*-u. Percentili nijesu isto što i procenti. Percentili zapravo dijele distribuciju na dva dijela. To je onaj broj koji na određenoj tački distribucije daje odgovarajući rezultat do te tačke, uključujući i taj rezultat. Na primjer dvadeseti percentil daje sumu svih rezultata do dvadesetog procenta distribucije, uključujući i dvadeseti procenat. Najučestalija je upotreba 25-og, 50-og i 75-og percentila, s obzirom na to da ova četiri percentila dijele distribuciju u četiri grupe. Ove percentile nazivamo kvartilima. Njih ćemo dobiti ako štrahiramo polje *Quartiles*. Možemo odrediti i tačke koje ravnomjerno sijeku distribuciju na određeni broj jednakih grupa, a sami možemo definisati koliko takvih grupa želimo. To radimo štrahirajući polje *Cut points for...* Takođe, kada izaberemo treću opciju *Percentile(s)* možemo sami odrediti tačno koje percentile želimo da vidimo u *Output*-u.

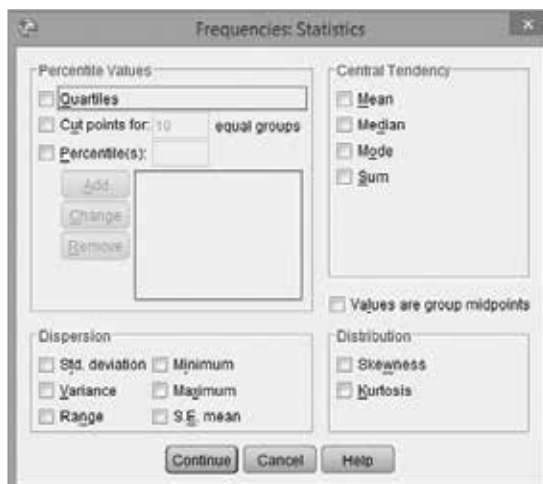
Sa desne strane je polje *Central Tendency* gdje možemo štrahirati da program izračuna i prikaže mjere centralne tendencije distribucije na određenoj varijabli. *Mean* je srednja vrijednost distribucije, *Median* je zapravo 50-ti percentil, odnosno ona tačka na distribuciji koja dijeli distribuciju na dva jednaka dijela. *Mode* je najučestalija vrijednost na distribuciji. A *Sum* je zbir svih vrijednosti na izabranoj distribuciji.

U dijelu *Dispersion* možemo da štrahiramo polja *Std.deviation* (standardna devijacija) koja ukazuje na to koliko vrijednosti na distribuciji variraju, odnosno koliko su udaljene od aritmetičke sredine. Polje *Variance* nam daje varijansu odnosno disperziju, to jest matematičko očekivanje odstupanje slučajne promjenljive od njene srednje vrijednosti. *Range* je zapravo opseg distribucije, to jest razlika minimalne i maksimalne vrijednosti na distribuciji. Ove dvije vrijednosti možemo dobiti štrahirajući polja *Minimum* i *Maximum*. Polje *S.E.mean* (*standard error mean*) predstavlja procijenjenu standardnu grešku uzorka.

U dijelu *Distributions* možemo štrahirati dvije opcije: *Skewness* i *Kurtosis*. *Skewness* je iskrivljenost, to jest nagib distribucije na varijabli koji nam pokazuje da li je distribucija asimetrična ulijevo ili udesno. *Skewness* će biti jednak nuli kada je distribucija simetrična, dodatno ako je vrijednost *skewness*-a negativna (-) tada je distribucija zakrivljena ulijevo, a ako je vrijednost pozitivna (+), onda je zakrivljena udesno. *Kurtosis* predstavlja spljoštenost distribucije, i vrijednost *kurtosis*-a govori u kojoj mjeri su vrijednosti koncentrisane oko aritmetičke sredine. Idealna distribucija ima vrijednost *kurtosis*-a 0, kod tako-zvanih spljoštenih distribucija vrijednost *kurtosis*-a je negativna a kod zašiljenih distribucija je pozitivna.

Polje *Values are group midpoints* štrahiramo samo kada vrijednosti podataka predstavljaju središte nekog niza. Na primjer, ako bismo na varijabli starost, sve one koji imaju između 20 i 29 godina kodirali kao 25.

## Charts

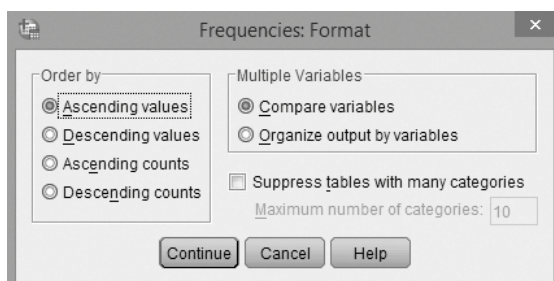


Slika 21: Polje *Charts* u *Frequencies* box-u

SPSS u svom *Output*-u pruža mogućnost da sam kreira najosnovnije grafike za izabrane varijable. Možemo izabrati *Bar charts* to jest histogram koji je grafik koji prikazuje distribuciju tako što se na horizontalnoj osi vrijednosti varijable (ono što smo definisali u polju *values* prilikom formiranja varijable), a na vertikalnoj osi su brojčane vrijednosti (to jest koliko procenata odgovora je grupisano na toj vrijednosti varijable). *Pie charts* će nam prikazati klasičnu „pitu“ to jest okrugli grafik gdje će djelovi pite biti označeni različitim bojama, a biće ih toliko koliko imamo vrijednosti varijable. Možemo izabrati i polje *Histograms* sa štrahiranih dijelom *Show normal curve on histogram* i onda ćemo na standardnom histogramu dobiti liniju koja pokazuje normalnu distribuciju, što olakšava upoređivanje koliko naša distribucija, i u kom dijelu odstupa od normalne.

U dijelu *Chart Values* možemo štrahirati jedno od dva polja: *Frequencies* i *Percentages*. Prvo znači da će vrijednosti prikazane na graficima biti numeričke vrijednosti na toj varijabli (na primjer na varijabli pol 4 i 6, što znači da smo imali četiri muškarca i šest žena). Kada štrahiramo polje *Percentages* na graficima ćemo dobiti procenata koje te numeričke vrijednosti zauzimaju (na istom primjeru, varijabli pol, to će biti 40% muškaraca i 60% žena).

## Format



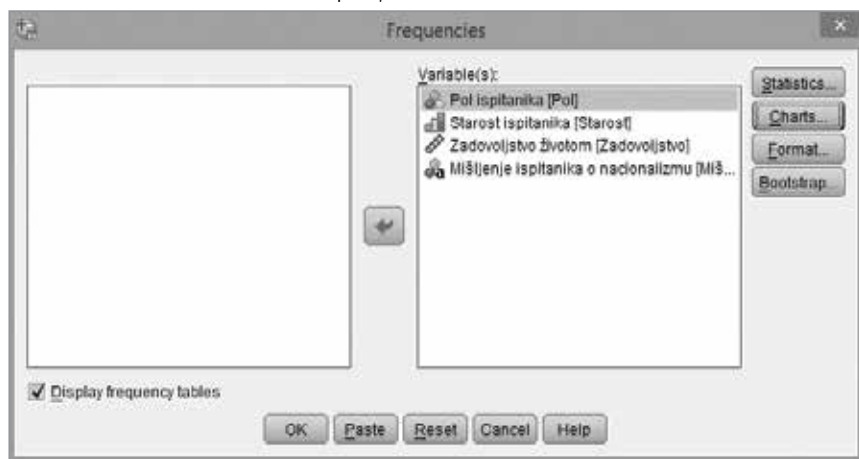
Slika 22: Polje *Format* u *Frequencies* box-u

U prvom dijelu polja *Format* možemo izabrati kako želimo da sortiramo vrijednosti na varijablama. *Values* su i ovdje vrijednosti na varijabli, i SPSS po *default*-u sortira rezultate u *Output*-u onako kako smo redom definisali *Values* prilikom kreiranja varijabli. *Ascending* je sortiranje od najniže ka najvišoj vrijednosti, dok je *Descending* suprotno, sortiranje od najviše ka najnižoj vrijednosti po tome kako su vrijednosti definisane. Dakle, ako smo definisali muški pol kao 1, a ženski kao 2, onda će nam tim redom biti dati rezultati u *Output*-u, kada je štrahirana opcija *Ascending Values*. Pored toga, vrijednosti na varijablama u *Output*-u možemo sortirati po counts, to jest po bročanoj vrijednosti tog ponuđenog odgovora, sortirane redom od one vrijednosti koja ima najmanje procenat ka onoj vrijednosti koja ima najveći (*Ascending counts*), i suprotno od najvećih ka najmanjim (*Descending counts*).

Kada analiziramo rezultate na nekoliko varijabli istovremeno, treba da obratimo pažnju na dio *Multiple Variables*. U principu najčešće u upotrebi je ono što je default opcija u SPSS-u: *Compare variables*, koja je pogodna i za poređenje među varijablama. Ovo polje svu deskriptivnu statistiku za zadate varijable sortira jednu do druge (*side-by-side*). Ako bismo štrahirali polje *Organize output by variables* onda bi program u *Output*-u napravio posebne varijable sa cijelom deskriptivnom statistikom za svaku varijablu posebno.

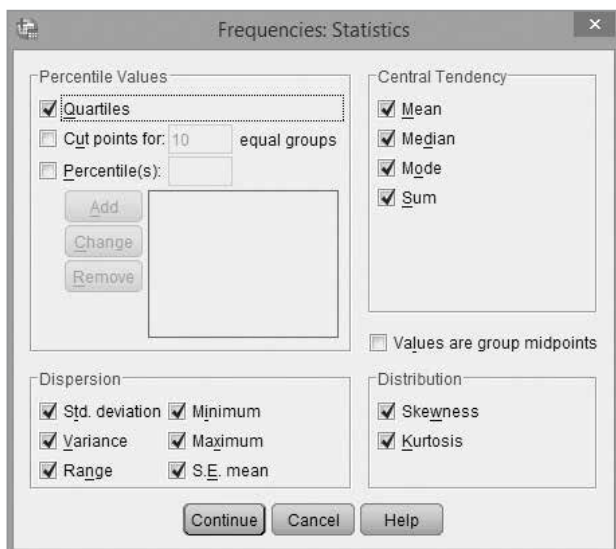
Polje *Suppress tables with many categories* omogućava da sami definišemo ako tabele imaju mnogo kategorija po koliko njih želimo da ih SPSS razdvoji ili spoji.

Za potrebe našeg primjera u *Frequencies box*-u prebacićemo sve četiri varijable prabacićemo sa desne strane u polje *Variable(s)*.



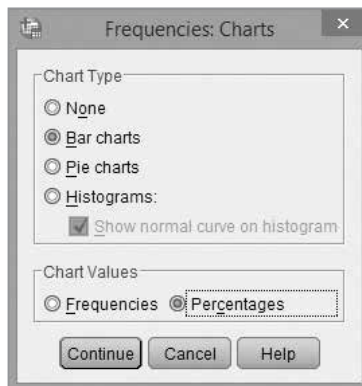
Slika 23: *Frequencies box* u SPSS-u

Dodatno na ovom primjeru želimo da provjerimo u *Output*-u sve opcije koje polje *Frequencies* nudi. Zato ćemo u polju *Statistics* štrahirati gotovo sva polja.



Slika 24: Polje *Statistics* u *Frequencies* box-u

Slično ćemo uraditi i za polje *Charts*.



Slika 25: Polje *Charts* u *Frequencies* box-u

U poljima *Format* i *Bootstrap* nećemo mijenjati ništa u odnosu na default opcije SPSS-a.

Poslije toga u polju *Frequencies* možemo izabrati *OK* i u *Output*-u ćemo vidjeti rezultate.

Opcije u *Output*-u SPSS-a su vrlo jednostavne za korišćenje, i *taskbar* nije puno različit u odnosu na prozor samog programa (*program window*).



Slika 26: Output prozor u SPSS-u

Opcije u *taskbar*-u *Output*-a u SPSS-u ne razlikuju se puno od glavnog programskog prozora. Dodatno, gotovo sve funkcije (na primjer: obrada podataka, transformisanje varijabli...) u vezi sa podacima iz otvorene baze se mogu obaviti i u *Output* prozoru. Radna površina podijeljena je da dva dijela. Na desnoj strani dobijamo informacije o svim koracima koje radimo u bazi podataka (na primjer: transformacija varijabli, ponderisanje podataka, kreiranje novih varijabli...), ali i sve analize podataka koje obavljamo. Tako će nam se u tom dijelu pojaviti tabele sa traženim frekvencijama, deskriptivna statistika, i sve ono što smo tražili od programa da izračuna i prikaže. Sa lijeve strane je *Navigation* meni (koji podsjeća na *Navigation menu* iz *word*-a), i služi nam za lakše snalaženje u *Output*-u. SPSS pošto obavi tražene operacije rezultate koje da u *Output*-u sortira i sa lijeve strane sa naslovima, i prostim klikom na taj naslov odmah prebacuje na datu tabelu u desnom dijelu.

Na početku SPSS *Output*-a vidimo tabelu koja sumira svu traženu deskriptivnu statistiku za četiri izabrane varijable.

		Statistics			
		Pol ispitnika	Starost ispitnika	Zadovoljstvo životom	Misljenje ispitnika o nacionalizmu
N	Valid	10	10	10	10
	Missing	0	0	0	0
Mean		1.00	2.00	5.40	
Std. Error of Mean		.143	.200	1.147	
Median		2.00	2.00	5.50	
Mode		3	3	1	
Std. Deviation		.516	.816	3.627	
Variance		.267	.667	13.156	
Skewness		-.404	.000	-.037	
Std. Error of Skewness		.687	.687	.687	
Kurtosis		-3.277	-1.293	-1.587	
Std. Error of Kurtosis		1.334	1.334	1.334	
Range		1	2	9	
Minimum		1	1	1	
Maximum		2	3	10	
Sem		1.6	2.0	5.4	
Percentiles	25	1.00	1.00	1.00	
	50	2.00	2.00	5.50	
	75	2.00	3.00	9.25	

Slika 27: Deskriptivna statistika u Output-u SPSS-u

Četiri izabrane varijalbe nalaze se u kolonama, dok cijela deskriptivna statistika. Prvo vidimo broj validnih *case*-ova (slučajeva) i nedostajućih (*missing*), a zatim redom sve ono što smo štrahirali u polju *Statistics* i vrijednosti koje nam je program izračunao. Primjećujemo da tih vrijednosti nema za četvrtu varijablu „Mišljenje ispitanika o nacionalizmu“ s obzirom na to da je varijabla definisana kao *string* (otvoreno pitanje) i za nju ne možemo računati mjere deskriptivne statistike. Naravno, pažljivom čitaocu neće promaći da su neke od ovih varijabli nominalne (pol ispitanika) ili definisane kao takve (starost ispitanika), te da neke mjere deskriptivne statistike nema smisla računati za takav tip varijable. Tako smo na primjer dobili podataka da je srednja vrijednost (mean) na varijabli pol ispitanika jednaka 1.60 što nema nikakvog smisla, s obzirom na to da je 1 – muškarci, a 2 – žene. Slično je i za neke druge mjere centralne tendencije, kao što je srednja vrijednost (*Median*).

Sa lijeve strane možemo vidjeti navigacioni meni koji nam dozvoljava da se krećemo brže i jednostavnije kroz *Output*.

Poslije tabele sa deskriptivnom statistikom SPSS nam daje pregled tabele frekvencije (*Frequency Table*).

**Pol ispitanika**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Muškarci	4	40.0	40.0	40.0
	Žene	6	60.0	60.0	100.0
Total		10	100.0	100.0	

Slika 28: *Frequency table* u *Output*-u SPSS-a

Sa lijeve strane, u redovima *Frequency Table*-a vidimo sve ono što smo definisali kao values prilikom kreiranja varijabli. U konkretnom slučaju, na primjeru varijable “pol ispitanika” to su dvije vrijednosti (dva *values*-a). U prvoj koloni, *Frequency*, vidimo broj opservacija, to jest frekvencija na datoj vrijednosti te varijable. Konkretno u našem primjeru 4 ispitanika su muškarci, a 6 ispitanica su žene. Nema *missing*, odnosno nedostajućih vrijednosti na ovom pitanju. Sljedeće tri kolone operišu procentima, i iako u našem primjeru među njima nema razlike, to nije uvijek tako. Kolona *Percent* predstavlja procenat svih slučajeva, uzimajući u obzir prilikom računanja i *missing cases* odnosno nedostajuće slučajeve. U koloni *Valid Percent* se računaju procenti uzimajući u obzir samo *valid cases* odnosno validne slučajeve, dakle bez nedostajućih (*missing*). *Cumulative Percent* kolona nam služi samo kada radimo sa ordinalnim ili intervalnim varijablama, to jest skalama, na kojima postoji više od dvije vrijednosti (*values*) s obzirom na to da ova kolona računa kumulativni procenat do tog određenog ponuđenog odgovora, odnosno uzima u obzir sve procenat prije toga. Na taj način, ova kolona nam može biti korisna kod ordinalnih i intervalnih varijabli jer ćemo odmah znati da li je neka opservacija “ispod ili iznad” određenog mjesta na distribuciji. S obzirom na to da na ovom pitanju imamo nominalnu varijablu, koja ima dva ponuđena odgovora, to jest dvije vrijednosti na varijabli, i dodatno nema nevažjećih slučajeva (*missing cases*), vrijednosti u kolonama su identične.

SPSS nam slične tabele daje i za one varijable koje smo definisali kao *string*, to jest za otvorena pitanja.

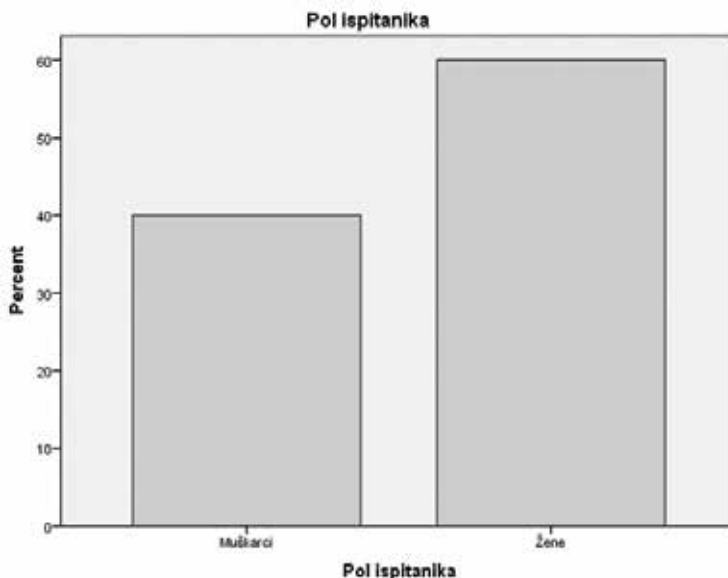
**Mišljenje ispitanika o nacionalizmu**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	5	50.0	50.0	50.0
Nacionalizam je isto što i patriotizam	1	10.0	10.0	60.0
Nacionalizam je naša ideologija	1	10.0	10.0	70.0
Nacionalizam je radikalno stanje svijesti	1	10.0	10.0	80.0
Nacionalizam je super stvar	1	10.0	10.0	90.0
Nacionalizam je veliki problem za jedno društvo	1	10.0	10.0	100.0
Total	10	100.0	100.0	

Slika 29: *Frequency table string varijable u Output-u SPSS-a*

Ovdje redom možemo vidjeti šta su sve ispitanici govorili na ovom pitanju. U slučaju da je bilo dva ili više identičnih odgovora program ih ne bi prikazivao odvojeno već kumulativno, pa bi i *Valid Percent* pored te vrijednosti bio veći u odnosu na ostale. Ovako prikazana string varijabla može biti informativna u nekim slučajevima, ali rijetko može pružiti prave podatke za kvantitativnu analizu ovakvog tipa pitanja.

Dodatno, u *Output-u* nalazimo i grafike koje smo štrihirali u polju *Charts*. Na primjeru varijable "pol ispitanika" grafik izgleda ovako:

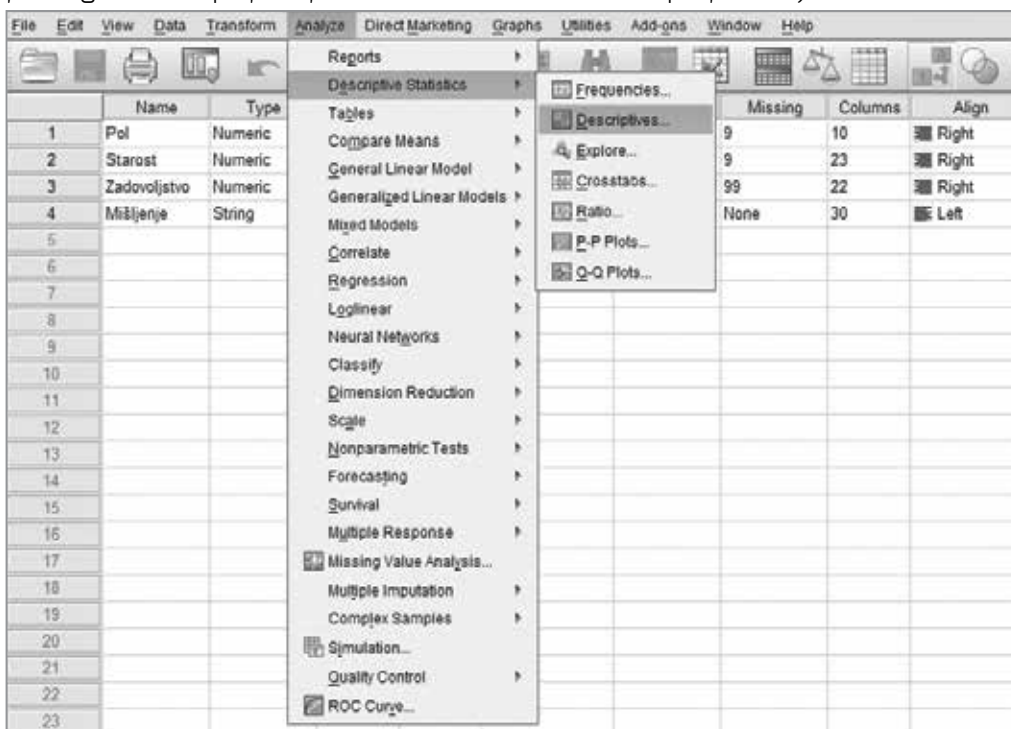


Slika 30: *Grafik u Output-u SPSS-a*



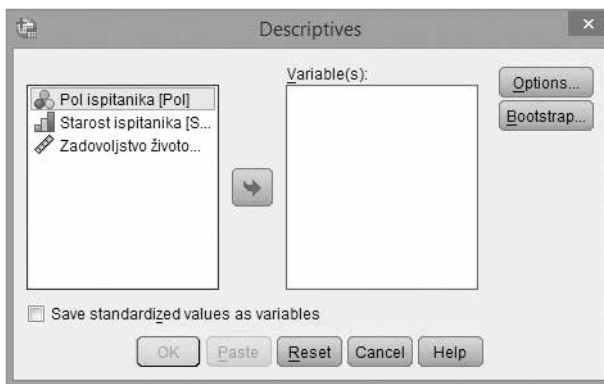
# DESKRIPCIJA (DESCRIPTIVES)

Do polja *Descriptives* dolazimo preko *Analyze – Frequencies – Descriptives*, i to je druga od tri opcije koju ćemo obrađivati u okviru polja *Analyze*.



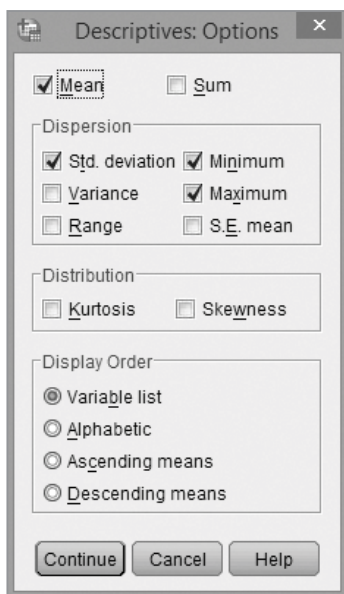
Slika 31: *Descriptives* polej u SPSS-u

Deskriptivnu statistiku (mjere centralne tendencije, disperziju i distribuciju vrijednosti na varijabli, kao i sređivanje (*ordering*) možemo izračunati i preko ove opcije u SPSS-u.



Slika 32: *References* u SPSS-u

Sve to pronalazimo u polju *Options* koje se nalazi desno u *Descriptives box-u*.<sup>3</sup> Pažljivom čitaocu neće promaći da u ovom prozoru nema sve četiri varijable koje smo kreirali u našoj bazi. Nedostaje četvrta “Mišljenje ispitanika o nacionalizmu”. Ove varijable nema zato što smo je definisali kao string varijablu, to jest otvoreno pitanje, gdje nijesmo imali unaprijed predviđene odgovore, i samim tim zbog toga za nju ne možemo raditi deskriptivnu statistiku.



Slika 33: Polje *Options* u *Descriptives box-u*

Sve od ovih opcija pomenuli smo i objasnili šta znače u okviru *Statistics* opcije u *Frequencies* prozoru. Ove opcije i na ovom mjestu imaju istu funkciju, sa tom razlikom što nam polje *Descriptives* pruža još jednu mogućnost. U osnovnom prozoru ove opcije, možemo štrihirati polje *Save standardized values as variables* (sačuvati standardizovane skorove kao varijable), što nam omogućava da varijable transformišemo u z-skorove preko kojih standardizujemo sve varijable, bez obzira na to što one imaju različite skale, i zatim možemo da poredimo rezultate na više varijabli. Program će, ako štrahiramo ovo polje sam sačuvati nove varijable, koje možemo vidjeti u osnovnom prozoru SPSS-a u *sheet-u Variable View* na dnu liste varijabli.

Kako tipično izgledaju tabele u *Output-u* iz opcije *Descriptives* vidimo na primjeru treće varijable “Zadovoljstvo životom” koju ćemo prebaciti u desni prozor *Variable(s)*.

<sup>3</sup> I za *Descriptives box* opciju *Bootstrap* ostavićemo kako je po default-u podešena u SPSS-u

	N	Minimum	Maximum	Mean	Std. Deviation
Zadovoljstvo životom	10	1	10	5.40	3.627
Valid N (listwise)	10				

Slika 34: Descriptives tabele u Output-u

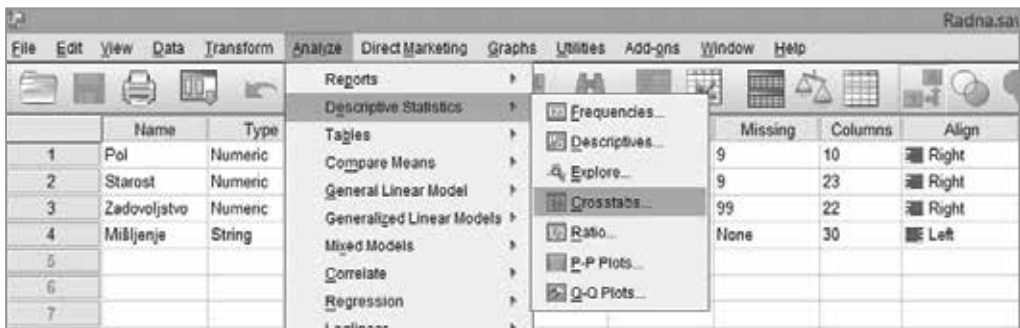
Po *default*-u SPSS nam u redovima daje varijable koje smo izabrali (u ovom slučaju varijabla “zadovoljstvo životom”) daje u redovima, dok u kolonama imamo sljedeće opcije:

- *N* – broj analiziranih slučajeva (*case*-ova)
- *Minimum* – minimalna vrijednost na izabranoj varijabli
- *Maximum* – maksimalna vrijednost na izabranoj varijabli
- *Mean* – srednja vrijednost
- *Std. Deviation* – standardna devijacija



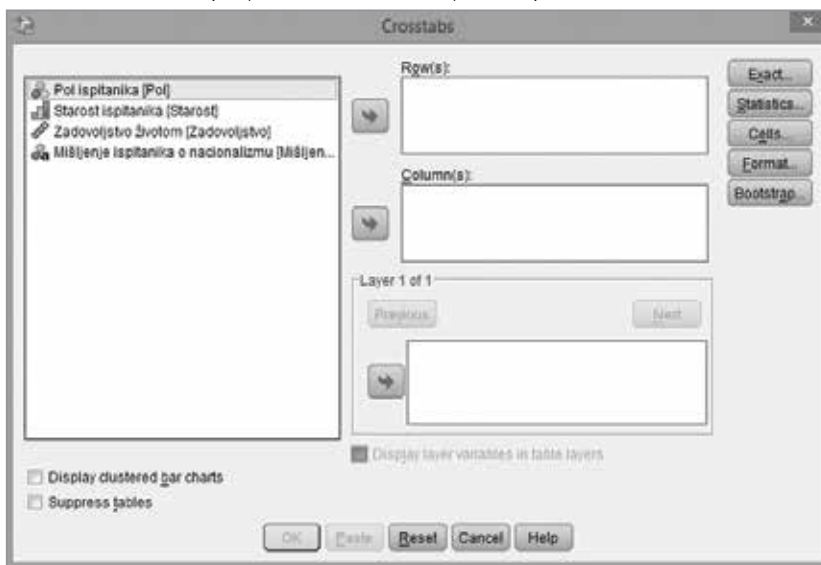
# KROSTABULACIJE (CROSSTABS)

Krostabulacija (tabela ukrštanja/kontingencije) je postupak kojim ispituujemo odnose između varijabli. Odnos između dvije varijable može biti potpuno nezavistan ili u potpunosti povezane varijable. Krostabulacije su samo jedan od čitavog niza statističkih metoda koji se bavi pitanjem povezanosti između varijabli. Najjednostavnija među njima je tabela ukrštanja/kontingencije.



Slika 35: *Crosstabs* opcija u SPSS-u

Kada izaberemo ovu opciju otvoriće se sljedeći prozor.



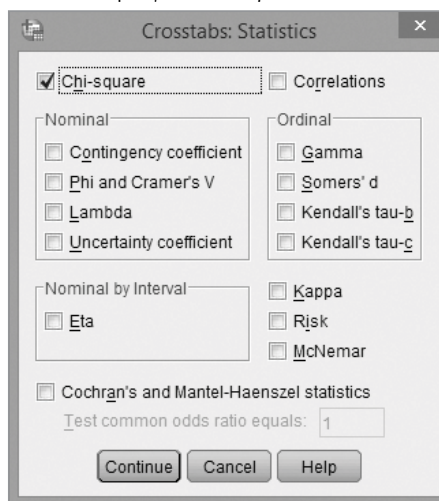
Slika 36: *Crosstabs* prozor u SPSS-u

Za varijable koje ukrštamo važno je da znamo u kakav odnos ih stavljamo. To jest, važno je da znamo odnos zavisna – nezavisna varijabla. Na ovom mjestu samo ćemo se podsjetiti da je nezavisna varijabla ona varijabla koja utiče na

promjene vrijednosti na drugoj varijabli. Ta druga varijabla, čije se vrijednosti mijenjaju, kao ishod promjena na nezavisnoj varijabli, zove se zavisna varijabla.

Za potrebe našeg primjera pretpostavimo da želimo da testiramo odnos između prve dvije varijable, to jest odnos između varijabli “Pol ispitanika” i “Starost ispitanika”. Važno je podsjetiti da to da li je neka varijabla zavisna ili nezavisna zavisi od dizajna istraživanja, te pretpostavki samog istraživača. U našem primjeru pretpostavićemo na primjer da su muškarci u našem uzorku stariji. Iz te pretpostavke slijedi da želimo ispitati kako promjene na varijabli “Pol ispitanika” utiču na promjenu na varijabli “Starost ispitanika”. Samim tim, varijabla “Pol ispitanika” u ovom ukrštanju biće nezavisna varijabla, dok će varijabla “Starost ispitanika” biti zavisna varijabla. Radi preglednosti podataka obično je lakše kada su nezavisne varijable u redovima, a zavisne u kolonama (mada to nije, naravno, pravilo). Tako ćemo, u našem primjeru varijablu “Pol ispitanika” prebaciti u prozor *Row(s)*, dok ćemo zavisnu varijablu “Starost ispitanika” prebaciti u prozor *Column(s)*. Prije nego što vidimo odnos između ove dvije varijable u *Output*-u potrebno je da razmotrimo još neke opcije u *Crosstabs* prozoru koje se nalaze sa desne strane.

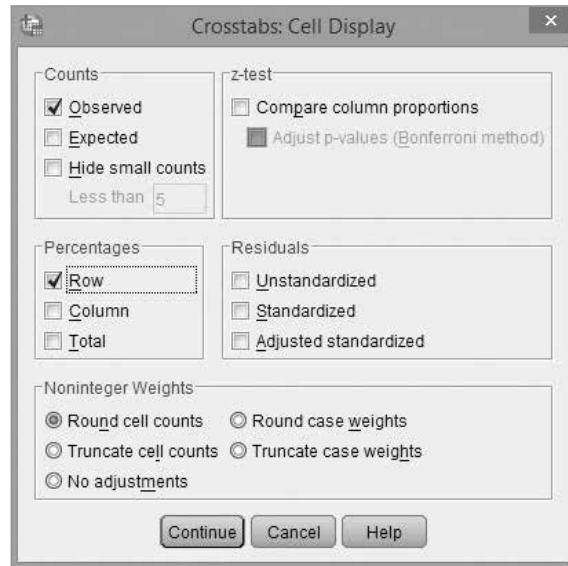
U polju *Statistics* štrahiramo opciju *Chi-Square*.



Slika 37: *Statistics* opcija u *Crosstabs*

Ovim smo od programa tražili da prilikom ukrštanja varijabli izračuna i *Pearson Chi-Square*. Riječ je o testu koji zapravo računa da li je ukrštanje izabranih varijabli statistički značajno, to jest mjeri da li je distribucija vrijednosti po redovima i kolonama nezavisna. Prema konvenciji, kada je vrijednost ovog parametra manja od 0,05 ( $p < 0,05$ ) to ukazuje da je moguće da postoji izvjesna veza između varijabli koje smo izabrali. To jest, naše ukrštanje je izdržalo test statističke značajnosti. Ako je  $p > 0,05$  onda možemo konstatovati da ne postoji veza između varijabli koje smo izabrali, to jest da u našem slučaju ne postoje statistički značajne razlike u starosti između muškog i ženskog pola.

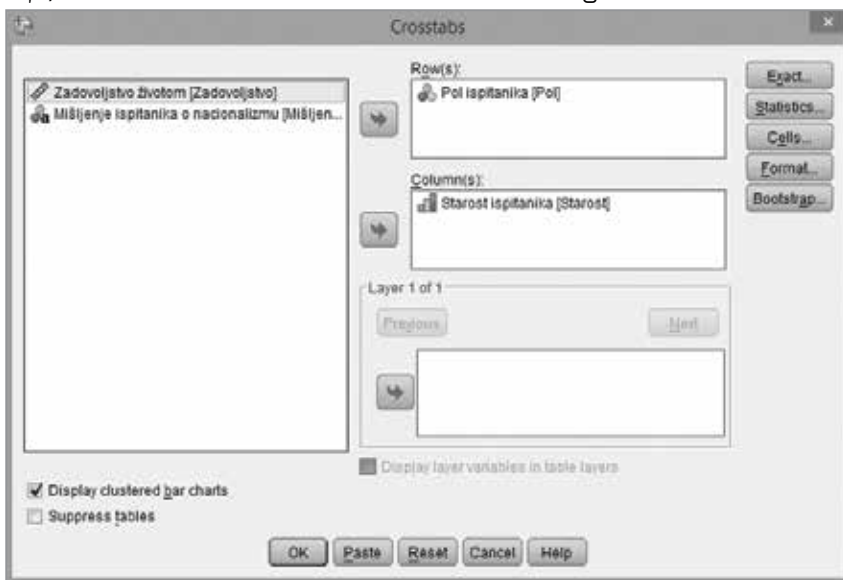
Dodatno u polju *Cells* na desnoj strani, štrahiramo u dijelu *Percentages* polje *Row*. Na ovaj način u *Output*-u procenti će biti predstavljeni po redovima, što će nam dodatno olakšati interpretaciju podataka.



Slika 38: *Cells* opcija u *Crosstabs*-u

U osnovnom *Crosstabs* prozoru izabraćemo još da želimo da u *Output*-u vidimo i grafik (*chart*). To ćemo uraditi tako što štrahiramo polje *Display clustered bar charts* u lijevom donjem uglu.

Poslije toga ispunili smo sva polja koja su potrebna za osnovne krostabulacije (ukrštanja) i *Crosstabs window* bi sada trebalo da izgleda ovako.



Slika 39: *Crosstabs window*

Zatim selektujemo *OK* i naša krostabulacija sa dodatnim testom i grafikom naći će se u *Output*-u programa.

**Pol ispitanika \* Starost ispitanika Crosstabulation**

		Starost ispitanika			Total	
		mladi (od 18 do 34 godine)	srednjovječni (od 35 do 54 godine)	stariji (55+ godina)		
Pol ispitanika	Muškarci	Count	1	2	1	4
		% within Pol ispitanika	25.0%	50.0%	25.0%	100.0%
	Žene	Count	2	2	2	6
		% within Pol ispitanika	33.3%	33.3%	33.3%	100.0%
Total		Count	3	4	3	10
		% within Pol ispitanika	30.0%	40.0%	30.0%	100.0%

Slika 40: Tabela kontingencije/ukrštanja u *Output*-u SPSS-a

Tabele kontingencije/ukrštanja u *Output*-u SPSS-a je prilično jednostavna i lako čitljiva. U redovima je nezavisna varijabla, to jest varijabla "Pol ispitanika" u našem primjeru, dok je u kolonama zavisna varijabla, u našem primjeru varijabla "Starost ispitanika". Tabela nam pokazuje koliko muškaraca, to jest žena ima u raznim kategorijama starosti. Zahvalna je za analizu dodatno i zato što nam prikazuje i frekvenciju (broj slučajeva, *case*-ova), ali i taj broj izražava u procentima. Posljednje kolone, desno i na dnu tabele prikazuju Total, to jest ukupno koliko je bilo slučajevapo svim kategorijama, i na zavisnoj i na nezavisnoj varijabli.

Pored ove tabele, u *Output*-u SPSS-a nalazimo i tabelu sa pomenutim *Pearson Chi-Square* testom.

**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	.278 <sup>a</sup>	2	.870
Likelihood Ratio	.277	2	.871
Linear-by-Linear Association	.000	1	1.000
N of Valid Cases	10		

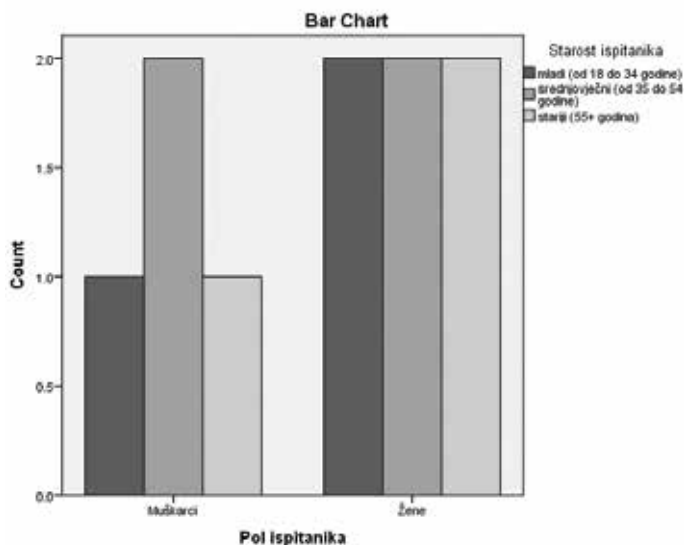
a. 6 cells (100.0%) have expected count less than 5. The minimum expected count is 1.20.

Slika 41: *Chi-Square Tests* tabela u *Output*-u SPSS-a

Podatak koji je nama bitan je zaokružen. Riječ je o vrijednosti *Pearson Chi-Square* testa, koji nam pokazuje statističku značajnost našeg ukrštanja. Kao što smo već istakli, samo kada je vrijednost *Pearson Chi-Square* manja od 0,05 (<0,05) onda možemo reći da je naše ukrštanje varijabli statistički značajno. U ovom slučaju vidimo da to nije slučaj jer je vrijednost testa 0,870. Naravno, u našoj bazi smo mogli i očekivati ovako nešto, s obzirom na to da imamo veoma mali broj slučajeva (*case*-ova), svega 10, pa gotovo izvjesno nijedno ukrštanje neće biti statistički značajno.

Pored ove tabele u *Output*-u nalazimo i grafik koji smo tražili.



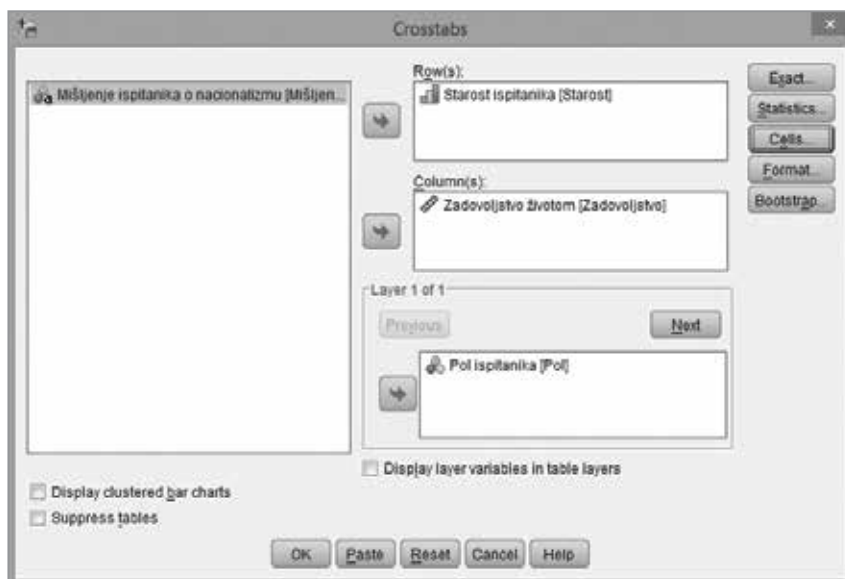


Slika 42: Bar chart ukrštanja u Output-u SPSS-a

Bar chart sa legendom sa desne strane, grafički je prikaz ukrštanja dvije varijable. Grafik je manje informativan u odnosu na tabelu, jer ne daje procenete, već samo frekvencije, to jest broj slučajeva.

Vrlo jednostavno i u okviru istog prozora, SPSS nam pruža mogućnost da napravimo i trostruko ukrštanje varijabli (*three-way crosstabulation*). Pretpostavimo da želimo da ispitamo kako godine utiču na zadovoljstvo životom, pri tom dodatno provjeravajući kako pripadnici različitog pola, različitih godina ocjenjuju zadovoljstvo svojim životom. U tom primjeru naša zavisna varijabla je "Zadovoljstvo životom", naša nezavisna varijabla je "Starost ispitanika", a dodatno vršimo svojevrsnu kontrolu ukrštanja i za varijablu "Pol ispitanika".

U tom slučaju u polje *Row(s)* prebacujemo našu nezavisnu varijablu, "Starost ispitanika", u polje *Column(s)* prebacujemo našu zavisnu varijablu, "Zadovoljstvo životom", a u treće polje *Layer*, prebacićemo našu kontrolnu varijablu, "Pol ispitanika". Opet ćemo štrahirati opciju *Chi-Square* u polju *Statistics*, kao i opciju *Row* u polju *Cells*. Tada bi naš *Crosstabs* prozor trebalo da izgleda ovako.



Slika 43: Crosstabs prozor u SPSS-u

Tabele koje dobijemo u *Output*-u SPSS-a ovom metodom tek malo su složenije za interpretaciju u odnosu na ukrštanje dvije varijable. Najprije treba razumjeti tabelu kontingencije ukrštanja tri varijable.

Starost ispitanika * Zadovoljstvo životom * Pol ispitanika Crosstabulation				Zadovoljstvo životom						Total	
Pol ispitanika				U potpunosti nezadovoljan	4	5	6	7	9		U potpunosti zadovoljan
Muškarci	Starost ispitanika	mlađi (od 18 do 34 godine)	Count			1	0	0		0	1
			% within Starost ispitanika			100.0%	0.0%	0.0%		0.0%	100.0%
	srednjovječni (od 35 do 54 godine)	Count		0	1	1				0	2
		% within Starost ispitanika		0.0%	50.0%	50.0%				0.0%	100.0%
	stariji (55+ godina)	Count		0	0	0				1	1
% within Starost ispitanika			0.0%	0.0%	0.0%				100.0%	100.0%	
Total			Count		1	1	1	1		1	4
			% within Starost ispitanika		25.0%	25.0%	25.0%			25.0%	100.0%
Žene	Starost ispitanika	mlađi (od 18 do 34 godine)	Count	1	1				0	0	2
			% within Starost ispitanika	50.0%	50.0%				0.0%	0.0%	100.0%
	srednjovječni (od 35 do 54 godine)	Count	0	0				1	1	1	2
		% within Starost ispitanika	0.0%	0.0%				50.0%	50.0%	50.0%	100.0%
	stariji (55+ godina)	Count	2	0				0	0	0	2
% within Starost ispitanika		100.0%	0.0%				0.0%	0.0%	0.0%	100.0%	
Total			Count	3	1				1	1	6
			% within Starost ispitanika	50.0%	16.7%				16.7%	16.7%	100.0%
Total	Starost ispitanika	mlađi (od 18 do 34 godine)	Count	1	1	1	0	0	0	0	3
		% within Starost ispitanika		33.3%	33.3%	33.3%	0.0%	0.0%	0.0%	0.0%	100.0%
	srednjovječni (od 35 do 54 godine)	Count	0	0	0	1	1	1	1	1	4
		% within Starost ispitanika		0.0%	0.0%	0.0%	25.0%	25.0%	25.0%	25.0%	100.0%
	stariji (55+ godina)	Count	2	0	0	0	0	0	0	1	3
% within Starost ispitanika		66.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	33.3%	100.0%	
Total			Count	3	1	1	1	1	1	2	10
			% within Starost ispitanika	30.0%	10.0%	10.0%	10.0%	10.0%	10.0%	20.0%	100.0%

Slika 44: Three-way Crosstabulation u SPSS-u

Kao što vidimo u prvom dijelu tabele vidimo ukrštanje svih muškaraca po tri starosne kohorte kako smo ih definisali prilikom kreiranja varijable, sa varijablom "Zadovoljstvo životom". U drugom dijelu tabele vidimo slično ukrštanje, s tim što je u tom dijelu riječ o ukrštanju svih žena po tri starosne kohorte, sa njihovom ocjenom zadovoljstva života. U trećem dijelu vidimo ukupno ukrštanje, to jest isto ono što bismo dobili prostim ukrštanjem dvije varijable "Starosti ispitanika" kao nezavisne, i "Zadovoljstva životom" kao zavisne, da nijesmo uključivali, to jest kontrolisali i za treću varijablu "Pol ispitanika". Neka polja u ovoj tabeli su prazna, zbog nedostajućih slučajeva, s obzirom na to naša baza ima svega 10 unijetih opservacija, to jest ispitanika.

Činjenica da je samo 10 opservacija još jednom će uticati tako da ni ovo ukrštanje neće, ni po jednoj od tri kategorije izdržati test statističke značajnosti, što možemo vidjeti u *Chi-Square Tests* tabeli u *Output*-u.

**Chi-Square Tests**

Pol ispitanika		Value	df	Asymp. Sig. (2-sided)
Muškarci	Pearson Chi-Square	8.000 <sup>b</sup>	6	.238
	Likelihood Ratio	8.318	6	.216
	Linear-by-Linear Association	2.679	1	.102
	N of Valid Cases	4		
Žene	Pearson Chi-Square	8.000 <sup>c</sup>	6	.238
	Likelihood Ratio	9.364	6	.154
	Linear-by-Linear Association	.129	1	.720
	N of Valid Cases	6		
Total	Pearson Chi-Square	12.639 <sup>a</sup>	12	.396
	Likelihood Ratio	15.186	12	.231
	Linear-by-Linear Association	.051	1	.822
	N of Valid Cases	10		

a. 21 cells (100.0%) have expected count less than 5. The minimum expected count is .30.

b. 12 cells (100.0%) have expected count less than 5. The minimum expected count is .25.

c. 12 cells (100.0%) have expected count less than 5. The minimum expected count is .33.

Slika 45: *Chi-Square tests za three-way crosstabulation*

Iz ove tabele vidimo da je program tri puta računao *Pearson Chi-Square* koeficijent, za sve tri pomenute varijacije ukrštanja varijabli. Još jednom, premali broj slučajeva uslovio je da nijedno ukrštanje nije statistički opravdano, to jest značajno (u svakom od tri slučaja *Pearson Chi-Square* je veći od 0,05).

Ukrštanje tri varijable (*three-way crosstabulation*) ponekad je veoma korisna metoda, s obzirom da može da pokaže statističku značajnost i opravdanost

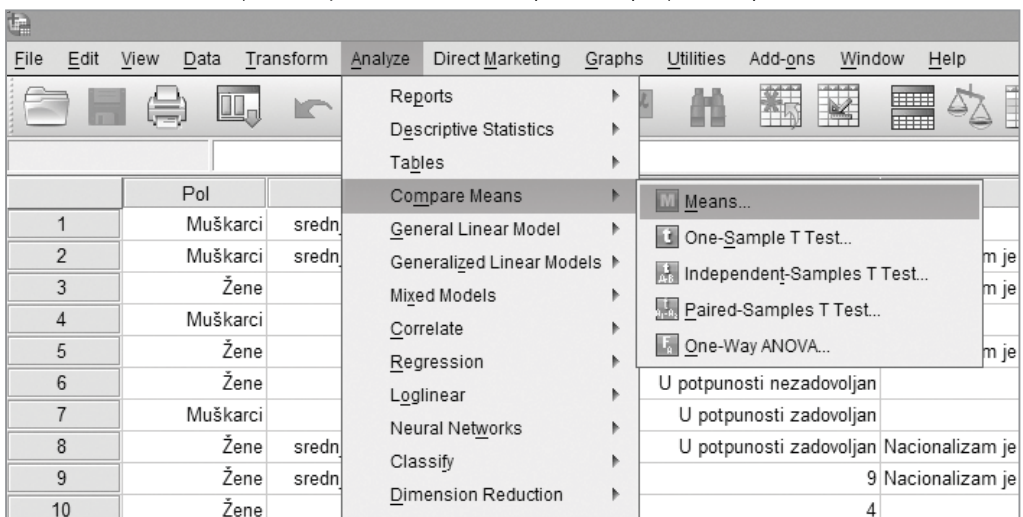
ukrštanja unutar jedne grupe ispitanika, a ne na cijelom uzorku. Na primjer, jedan od mogućih scenarija u našem primjeru je da samo ukrštanje muškaraca u tri starosne kohorte sa zavisno varijablom "Zadovoljstvo životom" bude statistički značajno (*Pearson Chi-Square* < 0,05), a da to nije slučaj kada su u pitanju žene, ili ukupno svi ispitanici. Prostim ukrštanjem dvije varijable "Starost ispitanika" kao nezavisne, te "Zadovoljstvo životom" kao zavisne, ne bismo dobili ovaj podatak, i on bi mogao da nam promakne.

Grafici, to jest *Bar chart*-ovi u *Output*-u SPSS-a, u slučaju ukrštanja tri varijable, zapravo pokazuju tri nezavisna grafika, za tri ukrštanja koja smo analizirali.

# POREĐENJE ARITMETIČKIH SREDINA (COMPARE MEANS)

Intervalne varijable (*scale* u SPSS-u), pogotovo ako su definisane preko petostepenih, ili desetostepenih numeričkih skala, često su nepogodne za tumačenje preko jednostavnih tabela frekvencije (*Frequencies tables*), i podatak koji za ove varijable kao što smo već vidjeli koji je mnogo informativniji je zapravo, aritmetička sredina (*mean*).

U našem primjeru, sa četiri varijable koje imamo, intervalna je naravno varijabla "Zadovoljstvo životom", definisana preko desetostepene numeričke skale. U našem prethodnom primjeru, za ukrštanje tri varijable (*three-way crosstabulation*) koristili smo ovu varijablu sa cijelom skalom, za potrebe primjera. Pažljiv čitalac je sigurno primijetio, čak i da smo radili sa bazom podataka koja ima mnogo više opservacija (slučajeva, *case-ova*), te i da je to naše ukrštanje bilo statistički značajno, da bi dobijeni podaci bili veoma "neinformativni" i nezahvalni za analizu, s obzirom na to da ne znači puno ako, na primjer, kažemo da je na varijabli "Zadovoljstvo životom" najveći broj ispitanika svoje zadovoljstvo ocijenio šesticom, pa sedmicom, pa četvorkom. Mnogo više nam, kao što smo istakli na ovom primjeru govori aritmetička sredina (*mean*). Već smo pokazali, kako se na jednostavan način preko opcija *Frequencies* i *Descriptives* može izračunati aritmetička sredina (*mean*) na varijablama u SPSS-u. Na ovom mjestu, želimo čitaoca upoznati da se i aritmetičke sredine mogu ukštati sa drugim varijablama. To jest, možemo posmatrati i kako se aritmetička sredina na jednoj varijabli mijenja, u odnosu na promjena na drugoj varijabli (nezavisnoj). To radimo, takođe u meniju *Analyze* u *taskbar*-u preko opcije *Compare Means – Means*.



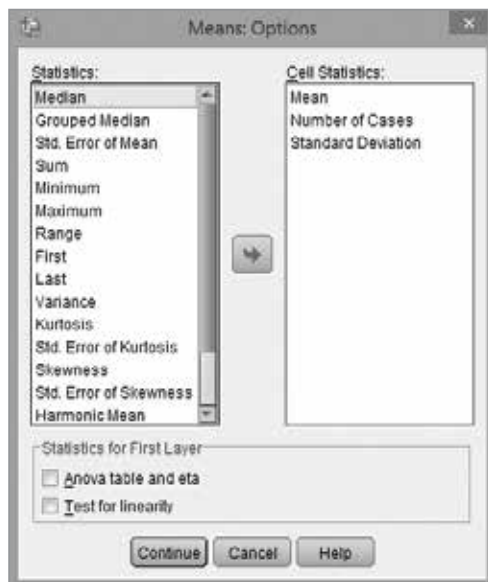
Slika 46: *Analyze – Compare Means – Means*

Onda bi trebalo da se otvori sljedeći prozor.



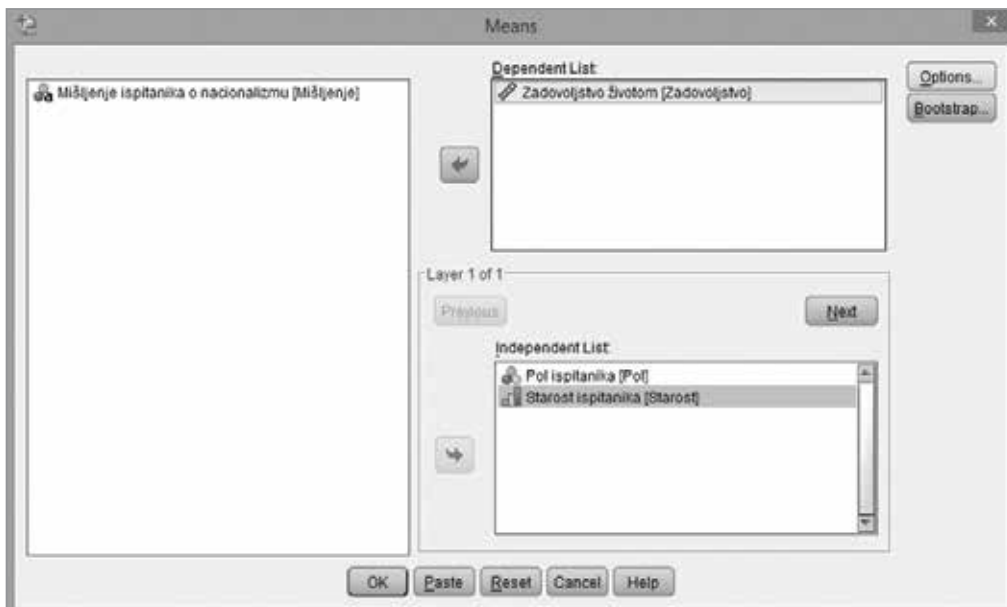
Slika 47: Means prozor (window) u SPSS-u

U polju *Dependent list* ubacujemo zavisne varijable. U našem slučaju, to je varijabla "Zadovoljstvo životom". U polju *Independent list* prebacujemo nezavisne varijable. Za potrebe našeg primjera, nezavisne varijable neka budu "Pol ispitanika" i "Starost ispitanika". Polje *Options* u ovom prozoru pruža listu podataka koje možemo uključiti u našu analizu. Sve što treba da uradimo je da ih izaberemo, i prebacimo iz lijevog polja *Statistics*, u desno polje *Cell Statistics*. Za potrebe našeg primjera, ostavićemo opcije u tom polju onako kako su po default-u podešene u SPSS-u, s obzirom na to da na ovom mjestu samo želimo predstaviti poređenje aritmetičkih sredina unutar grupa dvije varijable.



Slika 48: Options polje u prozoru Means

Pošto sve podesimo kako smo opisali, naš *Means* prozor bi trebalo da izgleda ovako.



Slika 49: *Means* window u SPSS-u

Izabrali smo, i prebacili u odgovarajuća polja zavisnu i nezavisne varijable. Pošto izaberemo polje *OK* rezultate ćemo vidjeti u *Output*-u.

#### Zadovoljstvo životom \* Pol ispitanika

Zadovoljstvo životom

Pol ispitanika	Mean	N	Std. Deviation
Muškarci	7.00	4	2.160
Žene	4.33	6	4.179
Total	5.40	10	3.627

#### Zadovoljstvo životom \* Starost ispitanika

Zadovoljstvo životom

Starost ispitanika	Mean	N	Std. Deviation
mladi (od 18 do 34 godine)	3.33	3	2.082
srednjovječni (od 35 do 54 godine)	8.00	4	1.826
stariji (55+ godina)	4.00	3	5.196
Total	5.40	10	3.627

Slika 50: *Compare Means* tabele u *Output*-u SPSS-a

U *Output*-u se nalaze dvije tabele, za dva poređenja aritmetičkih sredina koje smo izabrali. Vidimo da su i na ovaj način, nezavisne varijable (Pol ispitanika;

Starost ispitanika) prikazane u redovima, a kako se po grupama ispitanika unutar ovih varijabli mijenja aritmetička sredina (*mean*) na zavisnoj varijabli (Zadovoljstvo životom), vidimo u koloni *Mean*. Pogodnost predstavlja i to što je u posljednjem redu *Total*, tako da grupe možemo porediti i sa ukupnim uzorkom. Pored toga u *N* koloni vidimo broj slučajeva (*case-ova*) za svaku grupu. Treća kolona, *Std. Deviation*, prikazuje standardne devijacije za svaku grupu. Opet, zbog malog broja slučajeva u našoj bazi, standardne devijacije su veoma visoke.

Koje varijable i u kakav odnos stavljamo (to jest, koja varijabla će biti zavisna a koja nezavisna, ili više njih), zavisi od mnogo faktora. Prvi od njih je dizajn istraživanja, i šta se tim istraživanjem želi saznati. Dalje to može zavisi i od postavljenih hipoteza, to jest pretpostavki naših zaključaka, a na koncu i od kreativnosti istraživača.

Važno je istaći da na ovaj način mjerimo korelaciju među varijablama, a ne kauzalnost. Podsjetimo se, korelacija je onaj tip povezanosti između varijabli u kojem mi nemamo nikakvu kontrolu nad vrijednostima varijabli. Korelacije koristimo kao jednu od metoda za utvrđivanje moguće kauzalnosti, koju kasnije testiramo drugim statističkim metodama. Kauzalni (uzročno-posljedični) odnosi postoje kada promjene u jednom konceptu, to jest na jednoj varijabli, vode promjenama u drugom ili drugim konceptima (to jest varijabli/ama). Korelacije nijesu isto što i kauzalnost jer nikada ne možemo biti sigurni da su baš dvije posmatrane varijable u uzročno-posljedičnom odnosu, to jest da ne previđamo djelovanje neke treće varijable (ili više njih) koje vode takvom ishodu na zavisnoj varijabli. Ta treća varijabla se naziva intervenišuća varijabla.

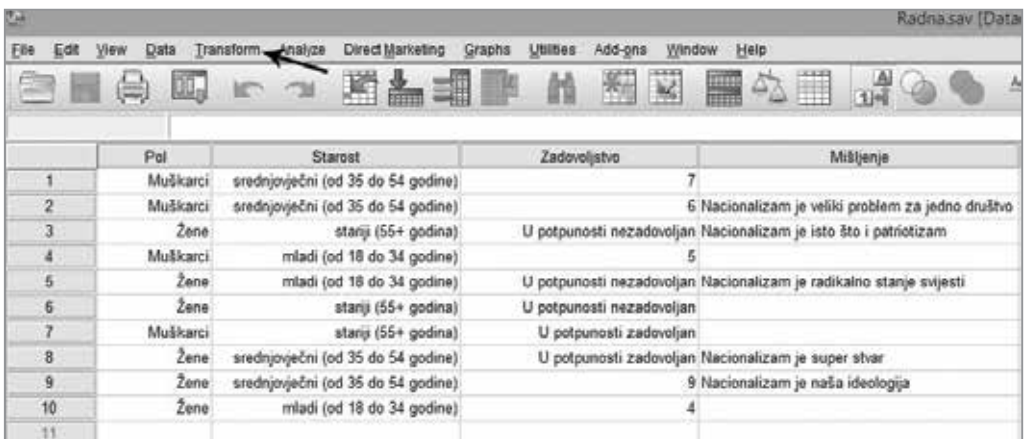


# TRANSFORMISANJE VARIJABLI

U nekim slučajevima odgovori koji su predviđeni upitnikom na nekim varijablama (to jest values u SPSS-u) su ili neinformativni, ili previše dijele opservacije po raznim odgovorima. Dodatno, nekada prosto pitanje sa većom skalom želimo predstaviti dihotomno (to jest sa dva pola/dvije opcije). Takođe, na nekim pitanjima, nekada želimo naše opservacije (ispitanike) podijeliti recimo na nekoliko jednakih djelova, ili na određenim tačkama distribucije, koje sami odredimo kao presjeke (*cutpoints*). To jednostavno možemo uraditi u SPSS-u preko dvije opcije, koje ćemo proučiti u ovom poglavlju:

- Rekodiranje varijabli (*Recode Variables*)
- *Visual Binning*

Obje funkcije obavljamo u *Transform* u *taskbar*-u SPSS-a.



	Pol	Starost	Zadovoljstvo	Mišljenje
1	Muškarci	srednjoječni (od 35 do 54 godine)		7
2	Muškarci	srednjoječni (od 35 do 54 godine)		6 Nacionalizam je veliki problem za jedno društvo
3	Žene	stariji (55+ godina)	U potpunosti nezadovoljan	Nacionalizam je isto što i patriotizam
4	Muškarci	mladi (od 18 do 34 godine)		5
5	Žene	mladi (od 18 do 34 godine)	U potpunosti nezadovoljan	Nacionalizam je radikalno stanje svijesti
6	Žene	stariji (55+ godina)	U potpunosti nezadovoljan	
7	Muškarci	stariji (55+ godina)	U potpunosti zadovoljan	
8	Žene	srednjoječni (od 35 do 54 godine)	U potpunosti zadovoljan	Nacionalizam je super stvar
9	Žene	srednjoječni (od 35 do 54 godine)		9 Nacionalizam je naša ideologija
10	Žene	mladi (od 18 do 34 godine)		4
11				

Slika 51: *Transform* u *taskbar*-u SPSS-a



# REKODIRANJE VARIJABLI (RECODE VARIABLES)

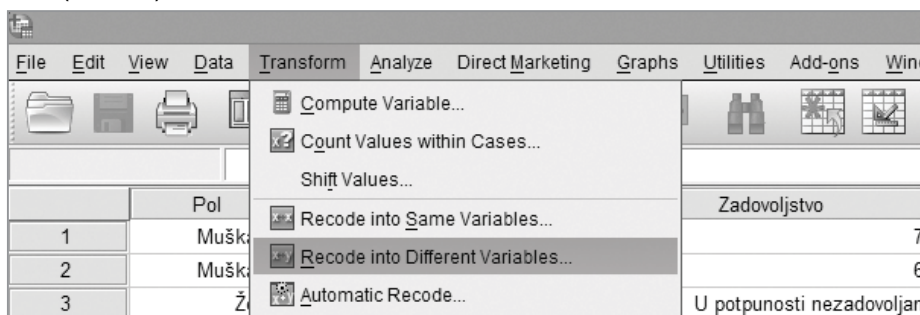
Nekada nam je potrebno da određene vrijednosti (*values*) na varijabli spojimo, ili da im promijenimo raspored. To radimo preko opcije rekodiranje varijabli (*recode variables*). Ova opcija ima dvije varijacije u Transform polju u *taskbar*-u:

- *Recode into Same Variables*
- *Recode into Different Variables*

Ove dvije opcije obavljaju identične funkcije sa jednom malom razlikom. Kada izaberemo prvu opciju (*Recode into Same Variables*) sve promjene koje unese-mo program će izvršiti u toj varijabli, preko (*overwrite*) postojećih podataka. S druge strane, ako izaberemo opciju *Recode into Different Variables*, uradićemo istu stvar, s tim što će program sam kreirati novu varijablu, na kraju liste varijabli (u *Variable View sheet*-u) sa promjenama koje smo unijeli, pritom će sačuvati istu varijablu na početku gdje je ona bila.

Da bismo razjasnili šta se ovom opcijom pokušava postići, osmislićemo jedan primjer. Napominjemo, na ovom mjestu ćemo raditi u opciji *Recode into Different Variables* kako bismo proučili i sve dodatke koje ova opcija nudi u odnosu na prvu. Pošto savlada ovu opciju, čitalac sasvim sigurno neće imati problema da se snađe i sa opcijom *Recode into Same Variables*.

Za potrebe našeg primjera pretpostavićemo da želimo da grupišemo sve ispitanike koji su odgovarali na pitanje o zadovoljstvu životom, u tri kohorte. Reći ćemo da su oni koji su odgovarali sa 10, 9 i 8 – optimisti, oni koji su odgovarali sa 7, 6, 5 i 4 – neutralni, i oni koji su odgovarali sa 3, 2 i 1 pesimisti. Optimiste ćemo označiti sa 1, pesimiste sa 2, a neutralne sa 3. Upravo to će biti i kodovi koje ćemo dodijeliti ovim odgovorima. Dakle, naša nova varijabla, imaće tri vrijednosti (*values*).



Slika 52: *Recode into Different Variables*

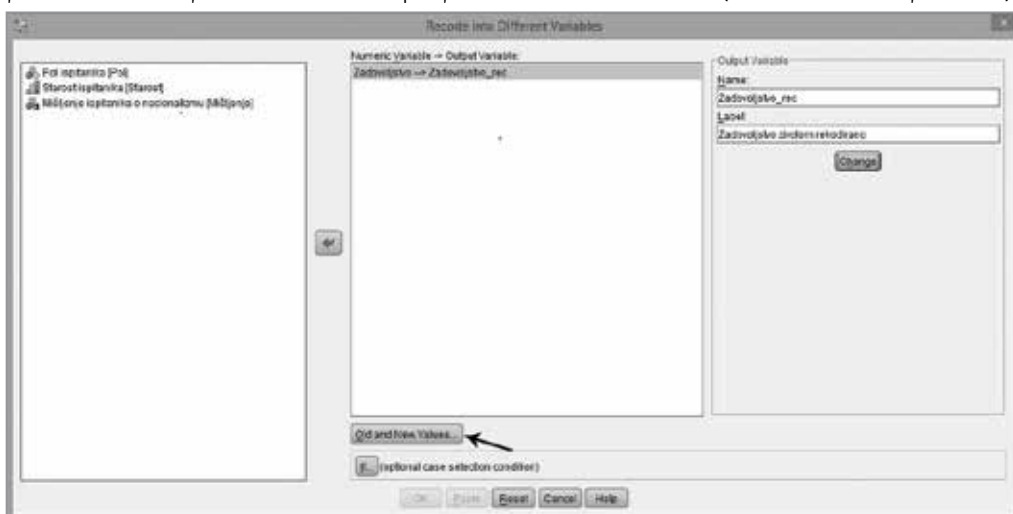
Pošto izaberemo ovu opciju, otvoriće se sljedeći prozor.



Slika 53: *Recode into Different Variables* prozor u SPSS-u

U ovom prozoru izabraćemo željenu varijablu (u našem slučaju „Zadovoljstvo životom“) i prebaciti je u desno polje *Input Variable -> Output Variable*. Prije nego što pređemo na samo rekodiranje, potrebno je da uradimo još jednu stvar. U desnom dijelu ovog prozora, u box-u *Output Variable* upisujemo ime i karakteristiku (*label*) varijable koju će nam program sam kreirati na kraju ovog procesa. U našem primjeru u polju *Name* (ime varijable) upisujemo „Zadovoljstvo\_rec“<sup>4</sup>, dok u polju *Label* (karakteristika) upisujemo „Zadovoljstvo životom rekodirano“. Da bismo izvršili promjenu pritisnemo opciju *Change*.

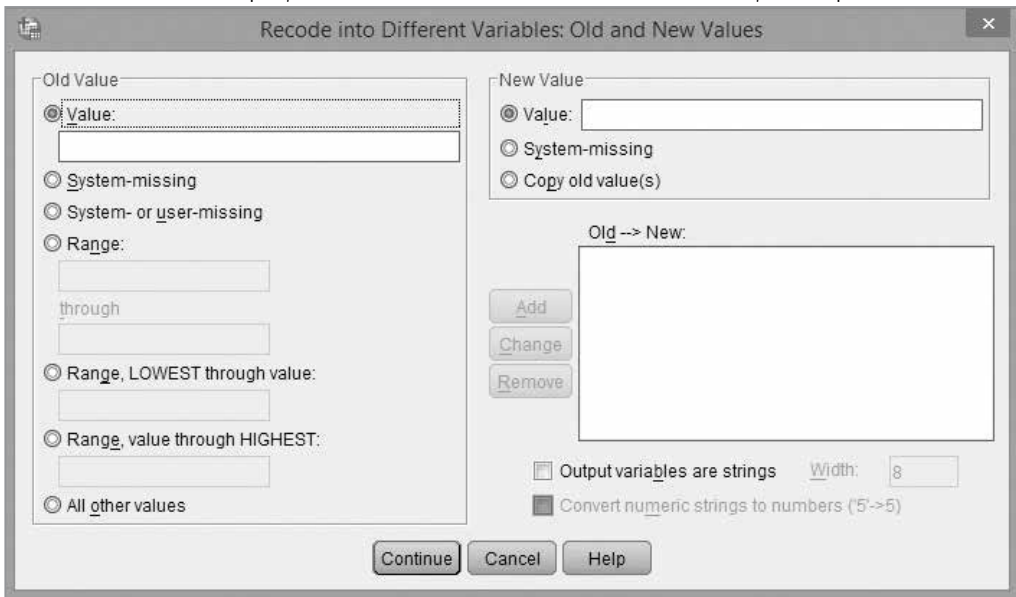
Pošto smo to uradili, možemo pristupiti samom rekodiranju, to jest promjeni vrijednosti na varijabli. To radimo u polju *Old and New Values* (Stare i nove vrijednosti).



Slika 54: *Recode into Different Variables* prozor poslije određivanja karakteristika nove varijable

<sup>4</sup> Pažljiv čitalac će se ovdje sjetiti da u imenu varijable ne možemo da upisujemo bilo šta sa razmacima (*space*-ovima)

Pošto izaberemo opciju *Old and New Values* otvoriće se sljedeći prozor.



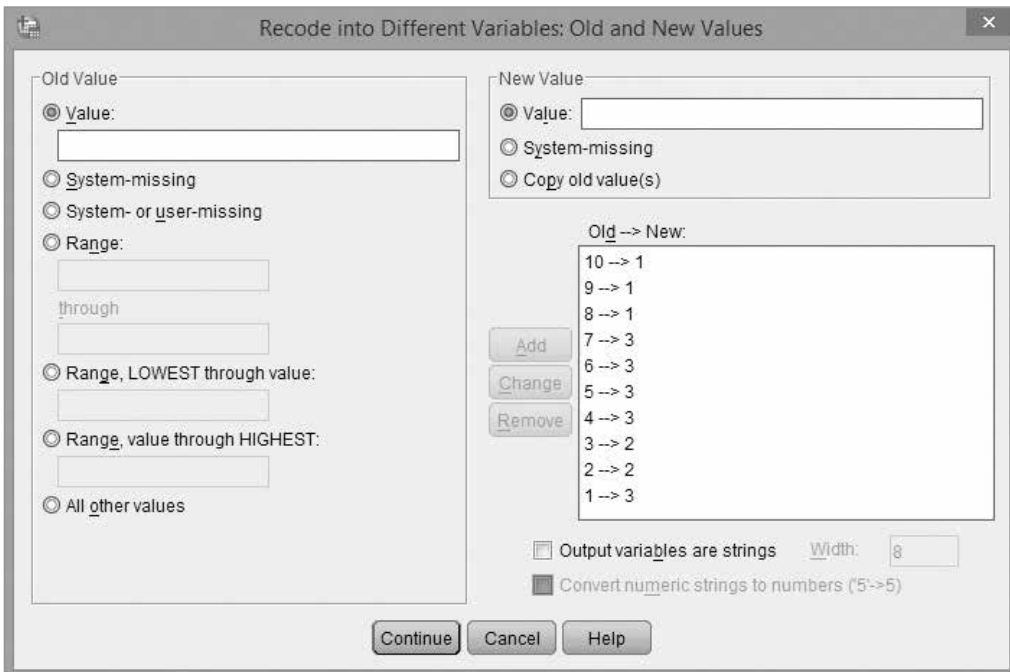
Slika 55: *Recode into Different Variables: Old and New Values*

U lijevom dijelu, gdje je *box Old Value*, u polju *Value* upisujemo vrijednost na varijabli koju mijenjamo (staru vrijednost), dok u desnom dijelu, u *box-u New Value*, u polju *Value* upisujemo novu vrijednost koju želimo da ima nova varijabla. Potom pritisnemo opciju *Add* i ono što smo radili će se pojaviti u polju *Old -> New*.

U našem primjeru, redom ćemo rekodirati jednu po jednu vrijednost na varijabli. U *Old Value* upisujemo 10, u *New Value* 1, i izaberemo *Add*. Zatim u *Old Value* upisujemo 9, a u *New Value* upisujemo 1, i izaberemo *Add*. Taj postupak ćemo redom ponoviti za sve vrijednosti, i to:

- 8 rekodiramo u 1
- 7, 6, 5, 4 rekodiramo u 3
- 3, 2, 1 rekodiramo u 2

Poslije toga naš *Old and New Values* prozor bi trebalo da izgleda ovako:



Slika 56: Old and New Values prozor poslije rekodiranja

Ukoliko smo napravili grešku, lako je možemo ispraviti. Izaberemo u *box-u Old -> New* tu vrijednost koju želimo da promijenimo, unesemo u polje *Value* izmjenu i selektujemo opciju *Change* (promijeni). Ako želimo da izbrišemo ono što smo rekodirali, prosto izaberemo te vrijednosti u istom *box-u* i selektujemo opciju *Remove* (ukloni).

Da bismo potvrdili sve promjene treba da pritisnemo opciju *Continue* na dnu prozora. Međutim prije nego to uradimo, treba obratiti pažnju na nekoliko dodatnih opcija u ovom prozoru, koje mogu biti od koristi i ubrzati rad.

- *System-missing* polje će odmah označiti sve one vrijednosti koje smo ranije u *Variable View sheet-u* označili kao nedostajuće (*missing*). Ova opcija zna u velikoj mjeri skratiti posao, jer su nam često nedostajuće vrijednosti na varijabli i na novoj, rekodiranoj varijabli nedostajuće. Tada je sve što treba da uradimo da i u *box-u Old Values*, i *New Values* izaberemo *System-missing*, pritisnemo *Add* sve nedostajuće vrijednost na ranijoj varijabli, biće nedostajuće i na novoj. Sličnu funkciju obavlja i polje *System-or-user-missing* s tim što ovo polje automatski i sve one vrijednosti na kojima nije unijet nikakav odgovor, tretira kao nedostajuće.
- *Range* (opseg) nam omogućuje da odredimo početnu i krajnju vrijednost na varijabli, i da onda te dvije vrijednosti i sve između njih, rekodiramo automatski, jednim korakom. U našem primjeru, mogli smo proces ubrzati, tako da smo izabrali opciju *Range*, u prvo polje upisali 8 *through* 10, i sa desne strane u *New Value* upisali 1. Program bi odmah sve vrijednosti, i 8 i 9 i 10 rekodirao u vrijednost 1 na novoj varijabli. Sličnu funkciju obavljaju i naredne dvije opcije.

*Range, LOWEST through value* pretpostavlja da se krene od najniže vrijednosti na varijabli, do neke koju sami odredimo, i da skup tih vrijednosti rekodiramo u jednom koraku. Na primjer mogli smo izabrati, ovo polje, upisati broj 3, i rekodirati sa 2. Tada bi sve naše vrijednosti od 1 (početna vrijednost na varijabli) do 3 (uključujući i sve vrijednosti između, to jest u našem primjeru i 2) bile rekodirane u jednom koraku. *Range, value through HIGHEST* traži da upišemo početnu vrijednost od koje želimo da anketiramo, a pretpostavlja da je završna vrijednost upravo ona koja je i završna vrijednost na varijabli.

- *All other values* (sve druge vrijednosti) polje nam olakšava rekodiranje, kada želimo da sve preostale vrijednosti na varijabli rekodiramo u jednu vrijednost. Ova opcija zna biti od velike koristi, kada pravimo tako-zvane *dummy* varijable, to jest one varijable koje imaju samo dvije vrijednosti. Pretpostavimo na primjer da želimo da imamo varijablu gdje su nam svi ispitanici iz prve starosne kohorte (od 18 do 34.godine) označeni kao 1, a sve ostale kohorte kao 0. Tada ćemo rekodirati vrijednost koju smo dodijelili na varijabli „Starost ispitanika“, u našem slučaju 1 u 1, a sve ostale vrijednosti u 0. Prosto ćemo izabrati polje *All other Values*, a u *box-u New Value* upisati 0, i izabrati *Add*.

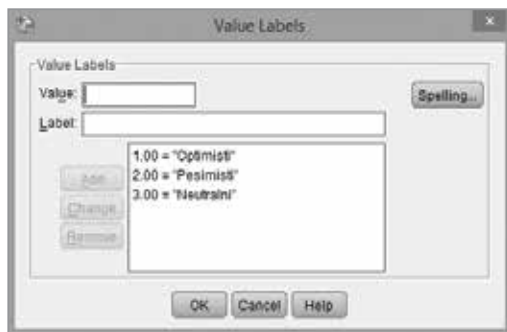
Sve ove opcije ostavljamo čitaocu da dodatno sam istražuje, u zavisnosti od potreba. U našem primjeru već smo obavili rekodiranje, korak po korak, i možemo nastaviti preko opcije *Continue*. Zatim u osnovnom, *Recode into Different Variables*, prozoru izaberemo *OK* i *Output* SPSS-a bi trebalo da pokaže da smo uspješno izvršili rekodiranje varijable „Zadovoljstvo životom“, i da je kreirana nova varijabla “Zadovoljstvo životom rekodirano”. To možemo vidjeti i u *Variable View sheet-u*. Nova varijabla nalaziće se na dnu liste varijabli, i naš prozor bi trebalo sada da izgleda ovako.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns
1	Pol	Numeric	8	0	Pol ispitanika	(1, Muškarci)	9	10
2	Starost	Numeric	8	0	Starost ispitanika	(1, mladi (od 18 do 34 godine))	9	23
3	Zadovoljstvo	Numeric	8	0	Zadovoljstvo životom	(1, U potpunosti nezadovoljan)	99	22
4	Mišljenje	String	200	0	Mišljenje ispitanika o nacionalizmu	None	None	30
5	Zadovoljstvo_rec	Numeric	8	2	Zadovoljstvo životom rekodirano	None	None	18
6								

Slika 57: Variable View poslije rekodiranja varijabli i kreiranja nove varijable

Ono što je program uradio je zapravo sumiranje različitih vrijednosti u jednu, i kreirao novu varijablu na kojoj će te vrijednosti (kojih je na prvoj varijabli nekoliko), prikazivati kao jednu vrijednost (jedan *value*), na novoformiranoj, rekodiranoj, varijabli.

Radi preglednosti podataka u analizi, u polju *values* (vrijednosti), na novoformiranoj varijabli „Zadovoljstvo životom rekodirano“ unijećemo one vrijednosti koje smo predvidjeli. 1 će biti optimisti; 2 će biti pesimisti; 3 će biti neutralni.



Slika 58: Values za rekodiranu varijablu

Tada ćemo u *Output*-u SPSS-a dobiti tabelu frekvencije koja nam je pregledna i mnogo čitljivija.

**Zadovoljstvo životom rekodirano**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Optimisti	3	30.0	42.9	42.9
	Neutralni	4	40.0	57.1	100.0
	Total	7	70.0	100.0	
Missing	System	3	30.0		
Total		10	100.0		

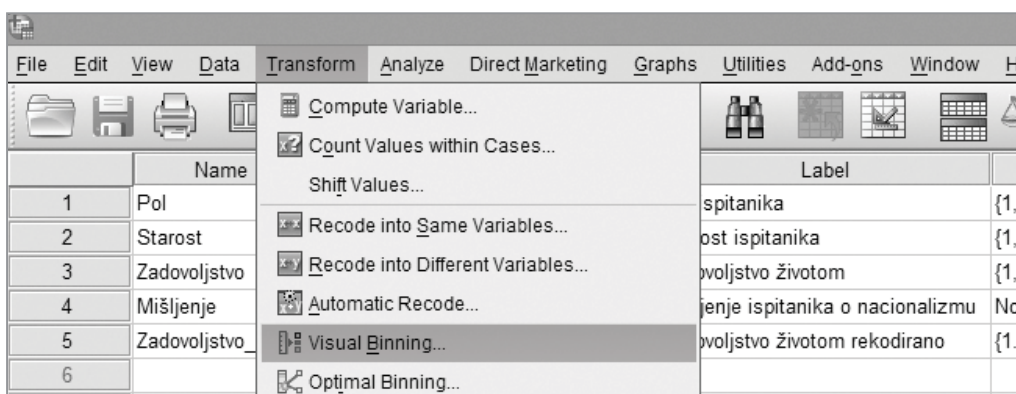
Slika 59: Tabela frekvencija u *Output*-u za novoformiranu rekodiranu varijablu

Ovim smo završili rekodiranje, i dobili smo novu varijablu koja ima vrijednosti koje smo željeli. Podsjećanja radi, ovo isto smo mogli uraditi i preko polja *Recode into Same Variables*. Tada bismo također dobili isti rezultat, s tim što ne bi bila formirana nova varijabla, već bi ova varijabla zamijenila staru varijablu "Zadovoljstvo životom".



# VISUAL BINNING

U nedostatku adekvatnog prevoda ovu opciju ostavljamo onako kako je možete pronaći u SPSS-u, u taskbar meniju *Transform*.



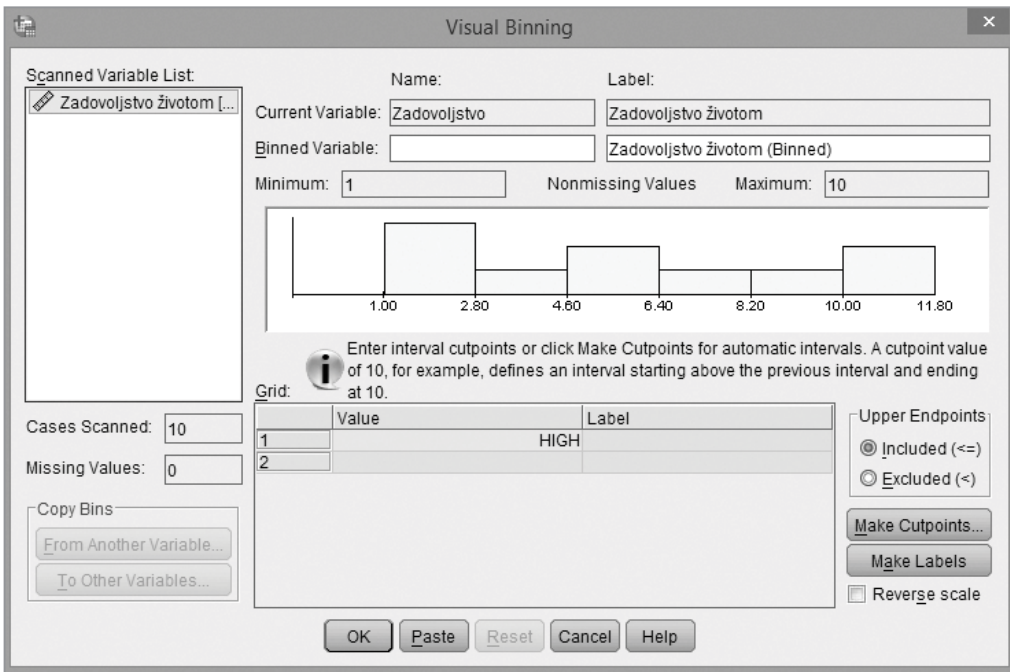
Slika 60: *Visual binning* opcija u SPSS-u

Ova opcija nam omogućuje da grupišemo odgovore na distribuciji na određenoj varijabli (baš kao i rekodiranje varijabli), s tim što u *Visual binning* opciji imamo par dodatnih pogodnosti. U ovoj opciji možemo podijeliti distribuciju na jednake djelove, i dodatno na onoliko djelova koliko želimo. Takođe, možemo podijeliti distribuciji na određeni broj djelova, ali da sami odredimo šta je prvi trenutak od kojeg program počinje da dijeli distribuciju. I ovu opciju ćemo proći korak po korak, kako bismo razjasnili sve što je preko *Visual Binning*-a moguće učiniti. Kada izaberemo to polje, dobićemo sljedeći prozor.



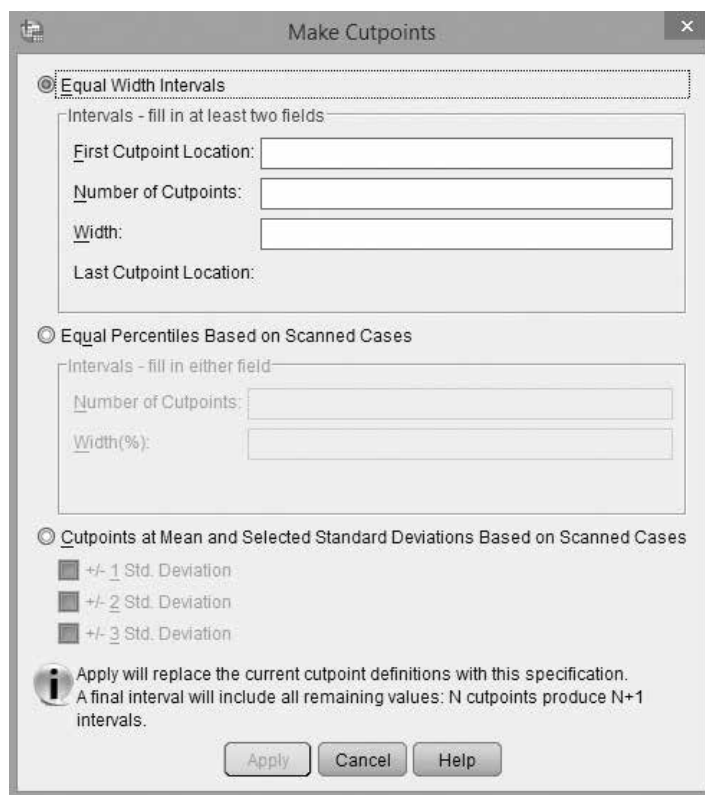
Slika 61: *Visual Binning* menu

I u ovom primjeru koristit ćemo varijablu “Zadovoljstvo životom”, koju ćemo izabrati, prebaciti u desni *box Variables to bin* i selektovati opciju *Continue*. Tada ćemo dobiti sljedeći prozor.



Slika 62: Visual Binning

U ovom prozoru vidimo nekoliko informacija. Najprije treba popuniti polje *Binned Variable* koje se nalazi pri vrhu. U ovom polju upisujemo ime novoformirane varijable koju će program kreirati, i koja će imati grupisane vrijednosti. Za potrebe našeg primjera novoformiranu varijablu ćemo nazvati “Zadovoljstvo\_rec2”. Takođe, u polju *do*, možemo mijenjati i *Label* varijable koja će biti formirana, ali u našem primjeru ostavićemo da to bude, po *default*-u, „Zadovoljstvo životom (*Binned*)”. Mali dio sa grafikom, zna biti veoma koristan jer prikazuje distribuciju vrijednosti na varijabli, a naknadno će prikazati crtama sva mjesta na distribuciji gdje je napravljen presjek (*cutpoint*). Da bismo to i uradili, izabraćemo opciju *Make cutpoints* u donjem desnom uglu. Tada ćemo dobiti sljedeći prozor.

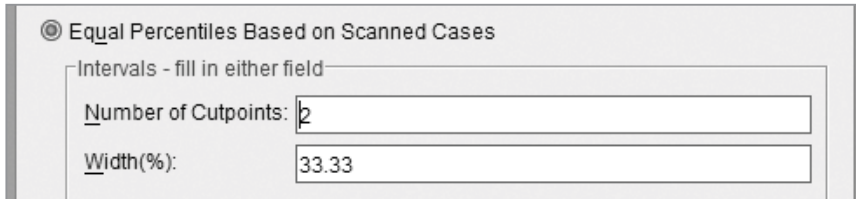


Slika 63: Make Cutpoints polje u Visual Binning

U prvom *box*-u ovog prozora, u kojem možemo unositi izmjene kada nam je štrahirana opcija *Equal Width Intervals* (intervali jednakih širina) možemo sami odrediti šta želimo da nam bude prvo mjesto presjeka, prvog mjesto od kojeg će program početi da dijeli distribuciju. To radimo u polju *First Cutpoint Location*. Dodatno, možemo sami odrediti na koliko djelova želimo da podijelimo sitribuciju na varijabli, što definišemo u polju *Number of Cutpoints*. Nije na odmet na ovom mjestu napomenuti, da ako želimo da podijelimo distribuciju na 5 djelova, na primjer, potrebna su nam četiri presjeka na distribuciji varijable, pa zato u polju *Number of Cutpoints* upisujemo 4. Takođe, u ovoj opciji, unutar *box*-a, program će sam, kada upišemo prvo mjesto presjeka, i broj presjeka koje želimo na distribuciji, izračuti *Width* (širinu jednog presjeka), kao i posljednje mjesto na distribuciji gdje je napravljen presjek. Na primjeru naše varijable „Zadovoljstvo životom“, ako bismo unijeli da je *First Cutpoint Location* 2, a *Number of Cutpoints* 3, program će sam izračunati da je *Width* 2,667, a da je *Last Cutpoint Location*: 7.

Ipak, od ove opcije, mnogo češće se koristi drugi box, unutar ovog prozora, u kojem možemo unositi promjene kada štrahiramo polje *Equal Percentiles Based on Scanned Cases* (jednaki percentili prema posmatranim slučajevima). Ovaj box koristimo kada želimo da podijelimo distribuciju na jednaki broj djelova, pri čemu sami definišemo koliko djelova želimo. Za potrebe našeg primjera, pret-

postavimo da varijablu „Zadovoljstvo životom“ želimo da podijelimo na tri jednaka dijela. Tada u polje *Number of Cutpoints* (broj presjeka), unosimo broj 2, i program automatski računa da će svaki dio na varijabli pokriti naravno 33,33% od distribucije na originalnoj varijabli, što prijavljuje u polju *Width(%)*. Upravo tako ćemo i napraviti presjek za potrebe našeg primjera.



Slika 64: *Equal Percentiles Based on Scanned Cases* box u *Visual Binning-u*

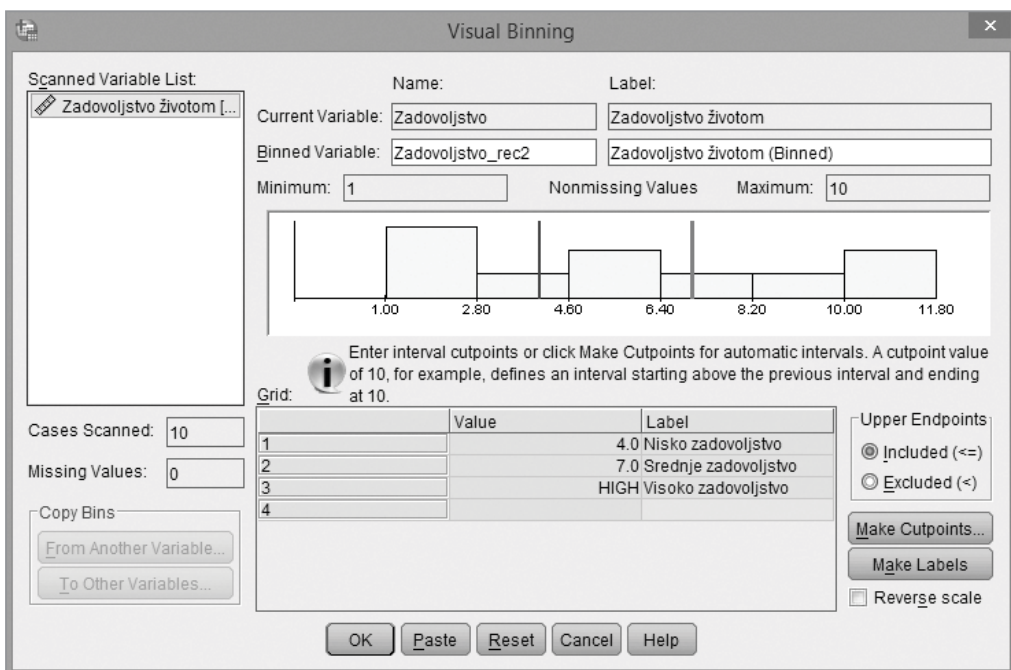
Prije nego izaberemo polje *Apply* da bismo potvrdili unijete promjene, razmotrimo još jednu mogućnost koju nam ovaj prozor pruža u svom trećem *box-u*. Kada štrahiramo opciju *Cutpoints at Mean and Selected Standard Deviations Based on Scanned Cases* mi zapravo želimo da napravimo takve presjeka na distribuciji varijable, gdje će jedan presjek biti na srednjoj vrijednosti, a preostali udaljeni od njega određeni broj standardnih devijacija (jednu, dvije ili tri), što sami definišemo štrahirajući jednu od tri ponudene opcije.

U našem primjeru, ipak, odlučili smo da podijelimo distribuciju na tri jednaka dijela, i pošto unesemo 2, u polju *Number of Cutpoints* u *box-u Equal Percentiles Base on Scanned Cases* možemo da izaberemo *Apply* na dnu našeg prozora.

Tada bi trebalo na našoj distribuciji, u malo dijelu gdje je grafik, i da vidimo dva mjesta na kojima je napravljen presjek. Pošto izaberemo opciju *Make Labels* u donjem desnom uglu ovog prozora, vidjećemo i *label-e* za vrijednosti na novotformiranoj varijabli.<sup>5</sup> Vidimo da je program tako podijelio distribuciju na varijabli da prva trećina, od ukupnog broja ispitanika, obuhvata sve ispitanike koji su na varijabli „Zadovoljstvo životom“ dali odgovor od 1 do 4. Druga trećina su svi oni koji su dali odgovor od 5 do 7, a treća grupa su oni koj su dali odgovor od 8 do završne vrijednosti na varijabli, to jest 10.

*Label-e* vrijednosti (*Value*), koje smo dobili ovim procesom, možemo mijenjati i unutar ovog prozora. Tako u polju *Label* na prvoj vrijednosti, pored broja 4, upisujemo „Nisko zadovoljstvo“, kako ćemo označiti sve one koji su na varijabli „Zadovoljstvo životom“ dali odgovor od 1 do 4. Za drugu vrijednost u polju *Label* upisujemo „Srednje zadovoljstvo“, a za treću vrijednost u polju *Label* upisujemo „Visoko zadovoljstvo“. Tada bi naš *Visual Binning* program trebalo da izgleda ovako.

<sup>5</sup> Polje ispod *Reverse scale* štrahiramo kada želimo okrenuti skalu, to jest da je najveća vrijednost na distribuciji prva, a najmanja posljednja. Takođe, u malom *box-u* sa desne strane *Upper Endpoints* štrahiramo da li želimo da naše vrijednosti uključuju granične vrijednosti na kojima je napravljen presjek – *Included (<=)*; ili da ih isključuju – *Excluded (<)*.



Slika 65: Visual binning prozor poslije definisanja karakteristika nove varijable

Sve karakteristike nove varijable su definisane, i sve što treba da uradimo jeste da izaberemo opciju *OK* na dnu našeg prozora. Program će prijaviti da će *Binning specifications* kreirati jednu novu varijablu, i to potvrdimo opcijom *OK*, i u *Output*-u SPSS-a vidimo da je kreirana nova varijabla, sa zadatim karakteristikama. To možemo vidjeti i na listi svih varijabli, u *Variable View sheet*-u.

	Name	Type	Width	Decimals	Label	Values	Missing
1	Pol	Numeric	8	0	Pol ispitanika	{1, Muškarci}...	9
2	Starost	Numeric	8	0	Starost ispitanika	{1, mladi (od 18 do 34 godine)}...	9
3	Zadovoljstvo	Numeric	8	0	Zadovoljstvo životom	{1, U potpunosti nezadovoljan}	99
4	Mišljenje	String	200	0	Mišljenje ispitanika o nacionalizmu	None	None
5	Zadovoljstvo_rec	Numeric	8	2	Zadovoljstvo životom rekodirano	{1.00, Optimisti}	None
6	Zadovoljstvo_rec2	Numeric	5	0	Zadovoljstvo životom (Binned)	{1, Nisko zadovoljstvo}...	99
7							

Slika 66: Variable View prozor poslije kreiranja Binned varijable

Zbog malog broja slučajeva (*case*-ova), kada bismo htjeli da vidimo tabelu sa osnovnom frekvencijom na varijabli koju smo kreirali preko *Visual Binning*-a, vidimo da distribucija nije baš tri jednaka dijela (po 33.33%).

**Zadovoljstvo životom (Binned)**

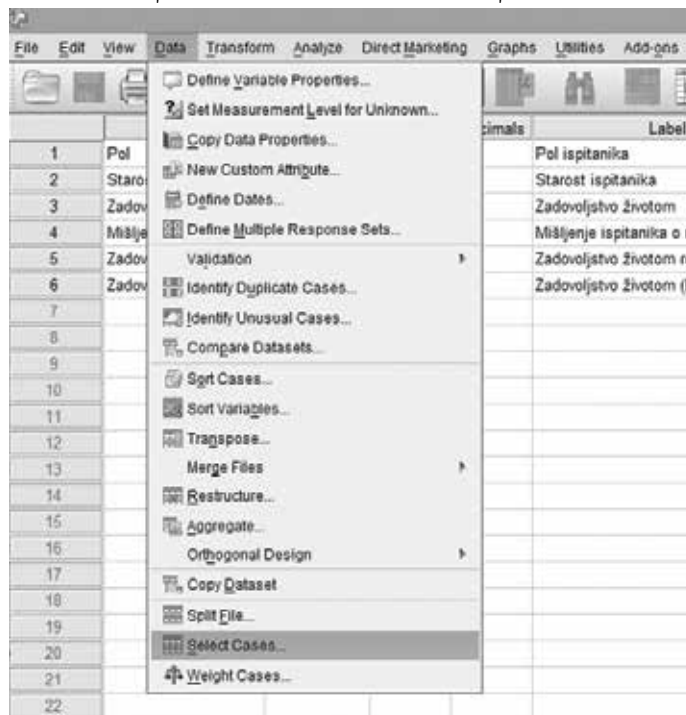
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Nisko zadovoljstvo	4	40.0	40.0	40.0
	Srednje zadovoljstvo	3	30.0	30.0	70.0
	Visoko zadovoljstvo	3	30.0	30.0	100.0
Total		10	100.0	100.0	

Slika 67: Tabele frakvencije za novoformiranu varijablu *Zadovoljstvo životom (Binned)*

I na ovom mjestu je jedini razlog zašto podjela nije „idealna“ (na tri ravnomjerna dijela po 33%), mali broj opservacija (*case-ova*). U velikim javnomnjenjskim istraživanjima, primjera radi, sa uzorkom reprezentativnim na nivou cijele populacije (1000+ opservacija/ispitanika), obično dobijamo zaista podjelu koju smo željeli, na jednake djelove, a odstupanja budu minimalna, i često ih mjerimo decimalama.

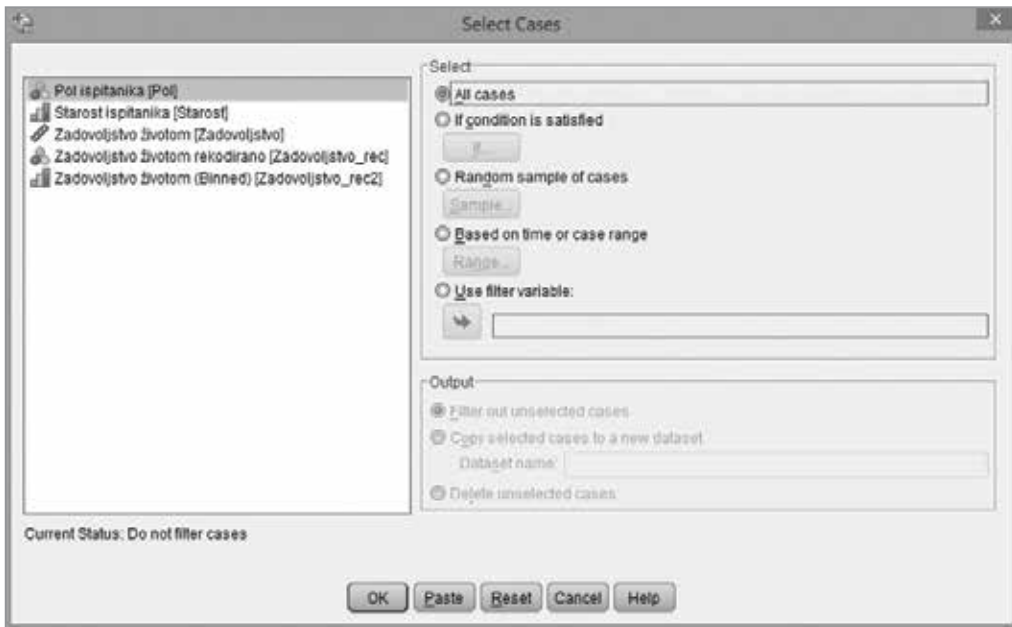
# IZBOR SLUČAJEVA

U nekim situacijama želimo da analiziramo samo određeni broj opservaija, a ne sve koje postoje u bazi. SPSS nam omogućava da aktiviramo filter, preko kojeg ćemo sami definisati koje opservacije želimo da posmatramo. To radimo preko polja *Select cases...* koje se nalazi u *taskbar* meniju *Data*.



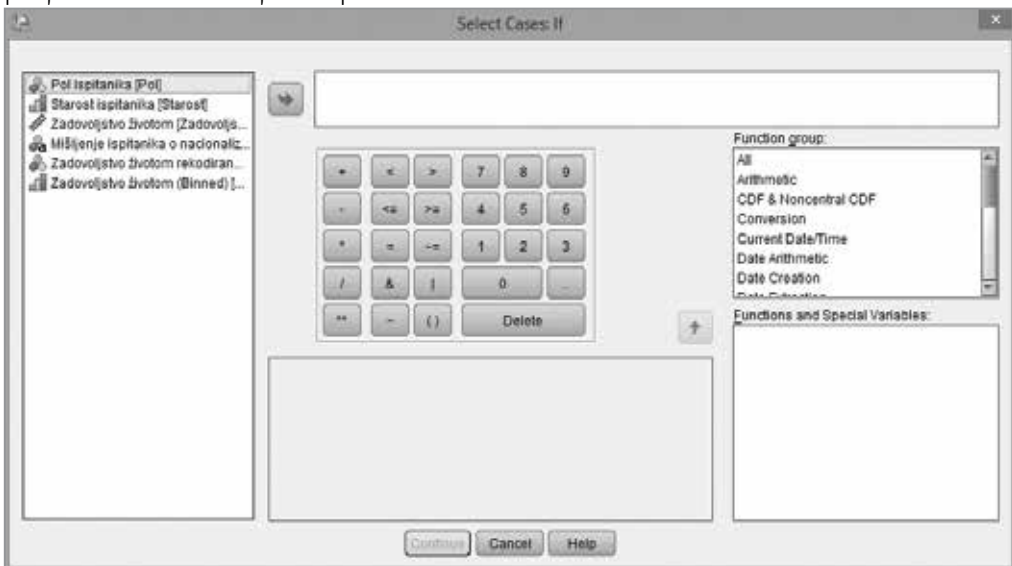
Slika 68: *Data* -> *Select Cases*

Pošto izaberemo ovu opciju, otvoriće nam se sljedeći prozor.



Slika 69: Select cases prozor u SPSS-u

U box-u *Select* na desnoj strani, vidimo da je po *default*-u štrahirana opcija *All cases*, što znači da sve analizi vršimo na svim slučajevima, to jest opservacijama koje imamo u bazi. Ako želimo to da promijenimo, to jest da vršimo analizu samo na onim slučajevima koje odaberemo, štrihiraćemo polje *If condition is satisfied* (ako je zadovoljen uslov). Da bismo definisali taj uslov, koji želimo, to jest da bismo definisali šta je naš filter, izaberemo polje *If* koje sada možemo selektovati, pošto smo štrihirali opciju *If condition is satisfied*. Kada izaberemo polje *If* dobićemo sljedeći prozor.



Slika 70: Polje If u Select Cases



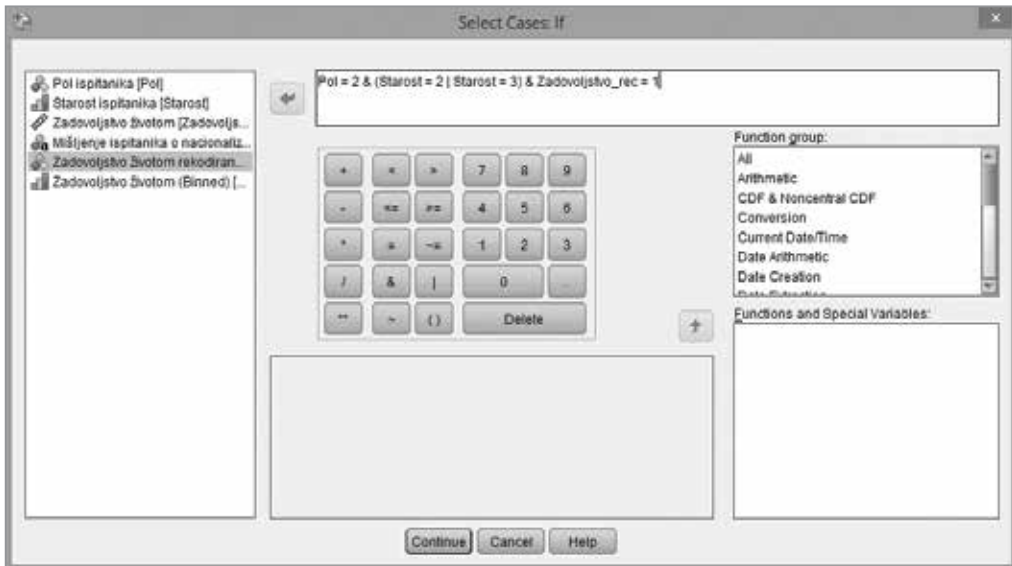
Najprije ćemo objasniti ulogu *box*-ova sa desne strane. U dijelu *Function group* su sve opcije koje SPSS u ovom dijelu može obaviti, grupisane po funkcijama. Kada bismo izabrali neku od njih, na primjer, *Arithmetic* u donjem *box*-u *Functions and Special Variables* vidjeli bismo pojedinačno sve opcije koje su grupisane ovom funkcijom. Ako bismo u tom *box*-u selektovali neku od funkcija, sa lijeve strane u plavoj površini dobili bismo tekstualno kratak opis šta ta opcija radi i čemu služi. Radoznali čitalac i sam će vidjeti koliko mnoštvo opcija i operacija je moguće uraditi sa varijablama samo unutar ovog prozora. Međutim, na ovom mjestu, nećemo koristiti ništa od tih opcija, već ćemo se zadržati na najosnovnijem izboru slučaja za analizu, po vrijednostima na varijabli/ama.

I prost izbor slučaja, po vrijednostima (*values*) na varijabli odnosno varijablama, pruža veliki broj mogućnosti, čak i kada radimo u tako maloj bazi kao što je naša, sa veoma malim brojem opservacija, i svega 4 originalne varijable, kojima smo dodali još dvije, dobijene transformacijom vrijednosti na jednoj od četiri početne.

Najprije ćemo proći kroz jedan primjer, a zatim pokušati da predstavimo mnoštvo opcija koje ovaj prozor nudi. Pretpostavimo da želimo da radimo analizu samo na onim opservacijama koje na varijabli „Pol ispitanika“ imaju vrijednost 2 (žene), pored toga na varijabli „Starost ispitanika“ imaju vrijednost 2 ili 3 (2 – između 35 i 54 godine; 3 – stariji od 55 godina), i dodatno na našoj transformisanoj varijabli „Zadovoljstvo\_rec“ imaju vrijednost 1 (optimisti). U ovom dijelu najvažnije je da stvari logički posložimo u svojoj glavi, a zatim je lakši dio zadati tražene opcije u SPSS-u.

Pored brojčane testature u sredini prozora imamo malu testaturu za razne opcije koje su nama potrebne. Sve što treba da uradimo je da redom prebacujemo varijable koje želimo, i određujemo koje vrijednosti će program da uzme u obzir u filter. Najprije, sun am potrebne sve žene. Dakle, izaberemo varijablu „Pol ispitanika“ iz lijevog menija i prebacimo je u prazni desni *box*. Pored „Pol“, ili ručno upisujemo ili selektujemo iz donje testature znak jednakosti i zatim upisujemo broj 2 (brojem 2 smo kodirali sve žene na nominalnoj varijabli „Pol ispitanika“). Za početak, u našem desnom *box*-u bi trebalo da piše: „Pol = 2“. Time smo izabrali sve žene, i možemo da pređemo na sljedeći korak. Među svim ženama, nama su dodatno potrebne one koje su starije od 35 godina. Postoje dva načina na koje možemo uraditi sljedeći korak. Najprije ćemo poslije „Pol = 2“ dodati znak „i“, to jest u programu „&“. Za prvi način, otvorićemo zagradu, prenijeti varijablu „Starost ispitanika“, upisati jednako 2, zatim staviti znak za „ili“, to jest u programu „|“ opet prenijeti varijablu „Starost ispitanika“ upisati jednako 3, i zatvoriti zagradu. Do sada smo kreirali filter, prema kojem će nam program izdovjiti sve one opservacije koje su žene koje pripadaju ili drugoj ili trećoj starosnoj kohorti. Naša formula do sada bi trebalo da izgleda

“Pol = 2 & (Starost = 2 | Starost = 3)”.<sup>6</sup> Preostalo nam je još da zadamo jedan filter. Izabrali smo sve žene starije od 35 godina (obje kohorte), i sada treba da među njima izaberemo sve one koje na transformisanoj varijabli “Zadovoljstvo životom rekodirano” (prvoj varijabli koju smo napravili, u rekodiranju varijabli, a ne u *Visual Binning*-u), izaberemo optimiste, to jest one koje imaju vrijednost 1. Na postojeću formula dodajemo jos jedno “i”, to jest “&”, prebacujemo varijablu “Zadovoljstvo\_rec” i upisujemo jednako 1. Tada bi naš *Select Cases: If* prozor trebalo da izgleda ovako.



Slika 71: *Select Cases –if sa filter formulom*

Prije nego što zaista potvrdimo ovaj filter preko opcije *Continue* na dnu prozora, razmotrimo još par opcija koje možemo koristiti prilikom kreiranja filtera u ovom prozoru:

- Možemo sabirati, oduzima, množiti, dijeliti i stepenovati varijable i vrijednosti na njima, koristeći se opcijama u prvoj koloni.
- Vrijednosti na varijablama možemo definisati kao “manje od”, “veće od”, ili “manje i jednako od” kao i “veće i jednako od”, preko opcije koje su i u SPSS-u označene kao i u matematici “<”, “>” ili “<=” odnosno “>=”.

Kako smo već definisali našu filter formulu, pritisnemo *Continue*, a zatim i *OK* u sljedećem prozoru, i SPSS će u *Output*-u pokazati šta smo uradili. Od ovog trenutka naš filter je aktiviran, i sada možemo raditi analizu samo sa izabranim slučajevima (to jest u našem primjeru ženama starijim od 35 godina, koje su prema transformisanoj varijabli zadovoljstva životom optimisti). Svi drugi

<sup>6</sup> Drugi način, s obzirom na to da varijabla „Starost ispitanika“ ima tri opcije, a mi tražimo dvije, i s obzirom na to da znamo da nema nedostajućih vrijednosti na njoj, jeste da opcije 2 i 3 definišemo preko negacije vrijednosti 1 na ovoj varijabli. To jest, zadamo programu da selektuje sve one slučajeve koji nijesu 1. Oznaka za to u SPSS-u je „~“. Tada bi naš desni box izgledao ovako “Pol = 2 & Starost ~ = 1”. To jest sve žene, koje nijesu u prvoj starosnoj kohorti.

slučajevi ne ulaze u analizu, i SPSS ih doslovno prekriži, što možemo vidjeti u *Data View sheet-u*.

	Pol	Šteta1	Zadovoljstvo	Mišljenje	Zadovoljstvo_rec	Zadovoljstvo_rec2	Šta_5
1	Muškari	predodređeni (od 35 do 64 godine)	7		Neutralni	Šećerje zadovoljstvo	Not Selected
2	Muškari	predodređeni (od 35 do 64 godine)	0	Racionalizam je veći problem za jedno društvo	Neutralni	Šećerje zadovoljstvo	Not Selected
3	Žene	stariji (55+ godine)	U potpunosti nezadovoljan	Racionalizam je isto što i potrošačizam		Šećerje zadovoljstvo	Not Selected
4	Muškari	mlađi (od 18 do 34 godine)	5		Neutralni	Šećerje zadovoljstvo	Not Selected
5	Žene	mlađi (od 18 do 34 godine)	U potpunosti nezadovoljan	Racionalizam je različito stanje svijeta		Šećerje zadovoljstvo	Not Selected
6	Žene	stariji (55+ godine)	U potpunosti nezadovoljan			Šećerje zadovoljstvo	Not Selected
7	Muškari	stariji (55+ godine)	U potpunosti zadovoljan		Optimisti	Višerje zadovoljstvo	Not Selected
8	Žene	predodređeni (od 35 do 64 godine)	U potpunosti zadovoljan	Racionalizam je super stvar	Optimisti	Višerje zadovoljstvo	Selected
9	Žene	predodređeni (od 35 do 64 godine)	U potpunosti zadovoljan	Racionalizam je naša stvar	Optimisti	Višerje zadovoljstvo	Selected
10	Žene	mlađi (od 18 do 34 godine)	6	Racionalizam je naša stvar	Neutralni	Šećerje zadovoljstvo	Not Selected
11							
12							

Slika 72: *Data View sheet* sa uključenim filterom

Sve analize podataka koje budemo radili, dok je uključen filter, uzimaće kao validne slučajeve/opservacije (*valid cases*) samo dve "neprekrižena" slučaja, to jest opservacije pod brojevima 8 i 9. U to se možemo uvjeriti i prostim pogledom na tabelu frekvencije na varijabli *pol*, na primjer.

#### Statistics

Pol ispitanika

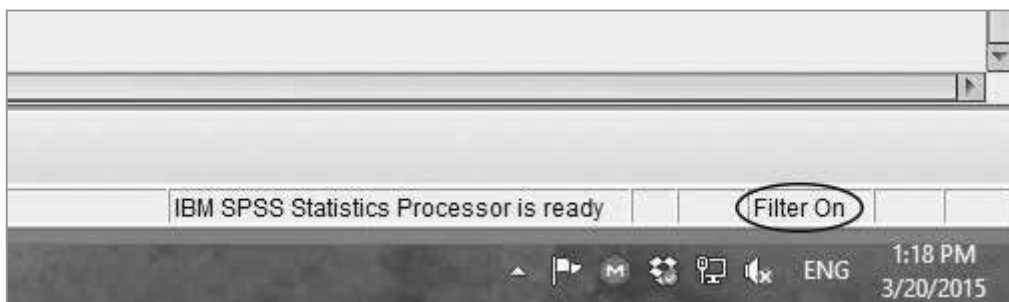
N	Valid	Missing
	2	0

#### Pol ispitanika

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid    Žene	2	100.0	100.0	100.0

Slika 73: Tabela frekvencije za varijablu *Pol ispitanika* pri uključenom filteru

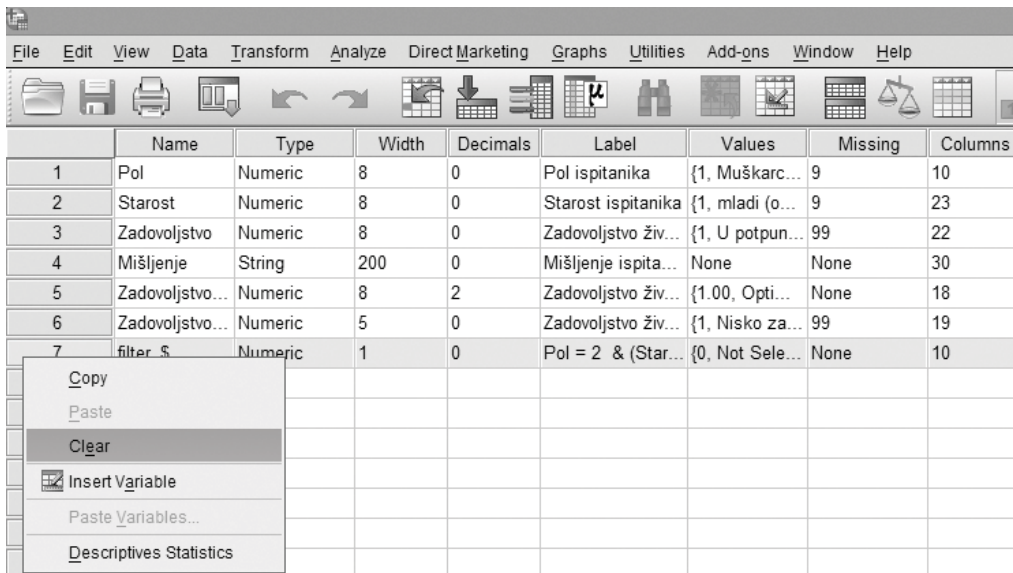
U našoj bazi, sa malim brojem slučajeva (svega 10), već sam pogled na *Data View sheet* otkriva da je uključen filter, s obzirom na to da su skoro svi case-ovi optisani. Međutim, u većim bazama, sa kakvim se najčešće radi u SPSS-u, sa velikim brojem opservacija to nije uvijek tako. Jedna od veoma korisnih stvari po ovom pitanju, koja nam govori da je filter uključen nalazi se u donjem desnom uglu.



Slika 74: *Filter On* polje u SPSS-u

Filter nam pruža mnogo mogućnosti za izbor slučajeva, i kao što smo ranije istakli sve što je potrebno je da istraživač logički posloži stvari u svojoj glavi, koje tačno slučajevu želi da izabere, i zatim unese to u traženo polje. Često, potreban nam je samo najosnovniji filter, po jednoj varijabli, na primjer da izaberemo samo muškarce sa varijable "Pol ispitanika". Tada bi, sa našom bazom, u filter polje, u meniju *Data* u *taskbar*-u bilo dovoljno samo da izaberemo tu varijablu i upišemo jednako 1, s obzirom na to da smo na nominalnoj varijabli "Pol ispitanika" muškarcima dodijelili numeričku vrijednost 1. Odnosno da u polju piše "Pol = 1".

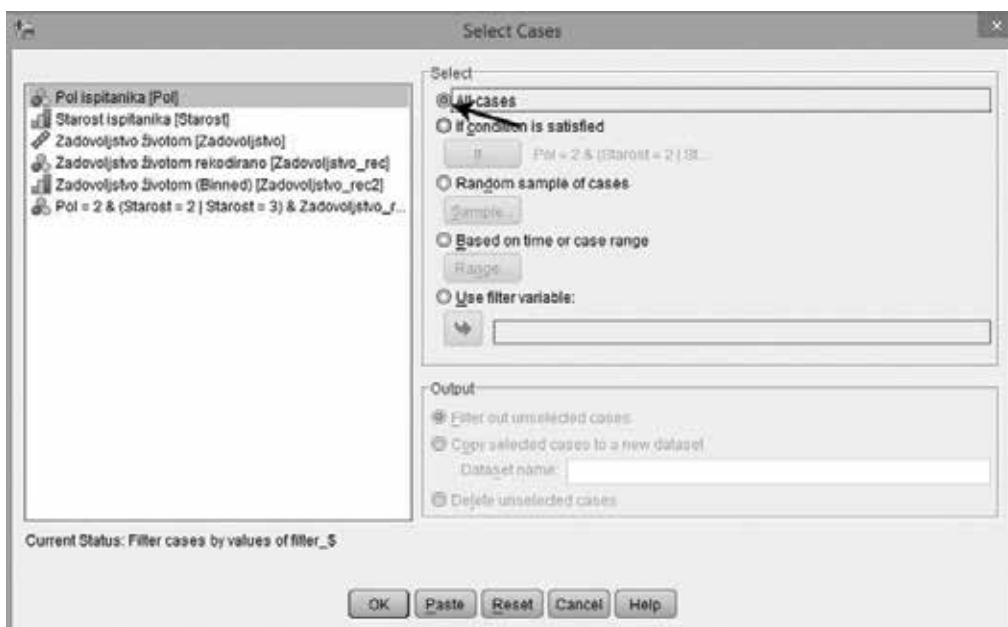
Na umu treba imati i da prilikom kreiranja filtera SPSS sam napravi dodatnu varijablu "filter\_\$", na kraju baze. Ta varijabla nam, što možemo vidjeti u *Data vies sheet*-u govori koji slučajevi su validni, kada je filter uključen. Na varijabli su dodatno automatski kreirani values (vrijednosti), pa su svi slučajevi koji su validni kada je uključen filter, označeni kao *Selected* (izabrani) i *Not Selected* (nije izabran). Ova varijabla, u velikoj bazi, sa mnogo varijabli, ponekad je višak, i najčešće je prosto izbrisemo u *Variable view sheet*-u, preko desnog klika na rednom broju varijable i opcije *Clear*.



	Name	Type	Width	Decimals	Label	Values	Missing	Columns
1	Pol	Numeric	8	0	Pol ispitanika	{1, Muškarc...	9	10
2	Starost	Numeric	8	0	Starost ispitanika	{1, mladi (o...	9	23
3	Zadovoljstvo	Numeric	8	0	Zadovoljstvo živ...	{1, U potpun...	99	22
4	Mišljenje	String	200	0	Mišljenje ispita...	None	None	30
5	Zadovoljstvo...	Numeric	8	2	Zadovoljstvo živ...	{1.00, Opti...	None	18
6	Zadovoljstvo...	Numeric	5	0	Zadovoljstvo živ...	{1, Nisko za...	99	19
7	filter_\$	Numeric	1	0	Pol = 2 & (Star...	{0, Not Sele...	None	10

Slika 75: Brisanje varijable filter

Kada želimo da ugasimo filter, i vratimo se na rad sa svim opservacijama, sve što treba da uradimo je da ugasimo filter. To radimo opet u meniju *Data* u *taskbar*-u, i polju *Select Cases*. U tom prozoru, sve što treba da uradimo je da štrahiramo ponovo polje *All cases* (svi slučajevi), u *box*-u *Select*.



Slika 76: Isključivanje filtera u SPSS-u

Output SPSS-a će opet prijaviti šta smo uradili, i kao što možemo vidjeti u *Data view sheet*-u svi slučajevi će ponovo biti validni. Dodatno, u donjem desnom uglu, neće više biti oznake *Filter On*.

CIP - Каталогизација у публикацији  
Национална библиотека Црне Горе, Цетиње

ISBN 978-9940-9655-0-1  
COBISS.CG-ID 26839824





# DEFACTO

CONSULTANCY

DeFacto Consultancy je agencija specijalizovana za istraživanje tržišta, ispitivanje javnog mnjenja, politički i biznis konsalting, osnovana 2008. godine u Podgorici. De Facto se izdvaja po tome što svakom projektu pristupa sa jedinstvenom pažnjom od dizajna istraživanja do njegove realizacije, pri tom posebno vodeći računa da se podaci prikupljaju i obrađuju u skladu sa uporednim iskustvima ili već definisanim indeksima i načinima mjerenja.

[www.defacto.me](http://www.defacto.me)