

3. Odabir i analiza skupina ispitanika

TEME

- | | |
|----|--------------------------------------------------------------------------------------|
| 1. | Čemu služe opcije za odabir i analizu skupina ispitanika? |
| 2. | Odabir uključivanjem ili isključivanjem kategorija nominalne ili ordinalne varijable |
| 3. | Odabir pomoću procedure Filter |

3.1. Čemu služe opcije za odabir i analizu skupina ispitanika?

Povremeno je potrebno napraviti analize (deskriptivne ili inferencijalne) samo na jednom dijelu uzorka, koji može biti definiran jednom karakteristikom / obilježjem (npr. samo muški spol) ili s više njih (samo muški spol, stariji od 35 godina, iz Dalmacije itd.).

U proceduri 'Descriptive Statistics' već smo objasnili funkciju 'Split', što je također funkcija za odabir i analizu skupina ispitanika, međutim, ta je funkcija ograničenog doseg a ako imamo više kriterija za definiranje našeg poduzorka.

3.2. Odabir uključivanjem ili isključivanjem kategorija nominalne ili ordinalne varijable

Radimo na primjerima u datoteci pod nazivom:
03_odabir_i_analiza_skupina_ispitanika.sav

Napomene:

- Odabir ispitanika pomoću kategorija nominalnih i ordinalnih varijabli funkcionira na osnovi naziva vrijednosti (labela), a ne pridružene numeričke vrijednosti.
- Cijeli je JASP temeljen na *open-source* programu i jeziku **R** te postupno nastalim modulima. R-ovi moduli (i sintaksa) ne funkcioniraju na glagoljici, egipatskim hijeroglifima, ali i kada se koriste slova s dijakritičkim znakovima (č, ć, đ, š, ž). Prema tome, slova u labelama s dijakritičkim znakovima potrebno je zamijeniti onima bez njih.

Filter	Value	Label
✓	1	Muški
✓	2	Ženski



Filter	Value	Label
✓	1	Muski
✓	2	Zenski

- Jako je važno ISKLJUČITI filter nakon njegova korištenja za određenu analizu, kako ne biste nehotično nastavili provoditi sve analize s uključenim filterom, iako vam je on trebao samo u primjerice jednom zadatku.

3.2.1. Odabir prema kategoriji jedne nominalne ili ordinalne varijable radi provedbe neke analize samo na određenoj skupini ispitanika

Primjer 1. Odabiremo samo ispitanike muškog spola jer želimo deskriptivno analizirati samo rezultate muškaraca na varijabli sreća.

1. Odaberemo varijablu Spol (kliknuti na naziv varijable)

The screenshot shows the SPSS menu bar with various analysis options: Descriptives, T-Tests, ANOVA, Mixed Models, Regre, Click here to change labels, Factor, and Bain. Below the menu bar is a data grid with columns for variables: nadbica1, nadbica2, nadbica3, nadbica4, nadbica5, spol, godina, and religja. The 'spol' column contains values 'muski' for all three rows. A red circle highlights the 'spol' column header.

	nadbica1	nadbica2	nadbica3	nadbica4	nadbica5	spol	godina	religja
1	vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	uvjereni vjernik
2	ne vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	uvjereni vjernik
3	vjerujem	vjerujem	vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	uvjereni vjernik

2. U prozoru koji se pojavio iznad podataka odabiremo koju kategoriju varijable Spol želimo "filtrirati", odnosno na njoj provoditi daljnje analize, dok je drugi dio podataka iz tih analiza isključen. Prvotno su sve kategorije ove varijable uključene (imaju oznaku kvačice), što znači da će sve buduće analize biti provedene i na rezultatima muškaraca i na rezultatima žena.

The screenshot shows the 'Filter' dialog box for the 'spol' variable. The 'Filter' column has checkmarks for both '1' (muski) and '2' (zenski). Below the dialog box is the same data grid as in the previous screenshot, but with a red circle around the 'spol' column header.

Filter	Value	Label
<input checked="" type="checkbox"/>	1	muski
<input checked="" type="checkbox"/>	2	zenski

	nadbica1	nadbica2	nadbica3	nadbica4	nadbica5	spol	godina	religja
1	vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	uvjereni vjernik
2	ne vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	uvjereni vjernik
3	vjerujem	vjerujem	vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	uvjereni vjernik

Ako želimo provesti neku buduću analizu samo na podacima muškaraca, tada treba "filtrirati" samo njihove podatke, tj. pod opcijom *Filter* isključiti kategoriju koja se odnosi na žene. To učinimo tako da kliknemo na kvačicu pokraj kategorije "ženski" kako bismo je isključili. Kad je pod opcijom *Filter* neka kategorija (ili više njih) označena križićem, to znači da će podaci svih ispitanika koji pripadaju ovoj kategoriji biti isključeni iz daljnjih analiza bilo koje varijable u tom data fileu.

The screenshot shows the 'Filter' dialog box for the 'spol' variable. The 'Filter' column has a checkmark for '1' (muski) and an 'X' for '2' (zenski). Below the dialog box is the same data grid as in the previous screenshots, but with a red circle around the 'spol' column header.

Filter	Value	Label
<input checked="" type="checkbox"/>	1	muski
<input checked="" type="checkbox"/>	2	zenski

	nadbica1	nadbica2	nadbica3	nadbica4	nadbica5	spol	godina	religja
1	vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	uvjereni vjernik
2	ne vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	uvjereni vjernik
3	vjerujem	vjerujem	vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	uvjereni vjernik

3. Kad smo odabrali na kojoj skupini ispitanika želimo provoditi daljnje analize, zatvorimo prozor za filtriranje i nastavimo s daljnjim analizama.

Uočite da su redci s podacima ispitanica (ženskog spola) svijetlo sive boje, što znači da su ti podaci trenutno neaktivni, odnosno da se sve daljnje analize koje zatražimo na njima neće provoditi. Ovako će ostati dok god ne isključimo filter.

Dodatno, uočite da se kraj imena varijable spol pojavio simbol za lijevak (kao i u Excelu) koji označava da je trenutno aktivan filter na toj varijabli.

	nadbica1	nadbica2	nadbica3	nadbica4	nadbica5	spol	godina	religija
22	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muški	1. godina	nije vjernik
23	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muški	1. godina	nije vjernik
24	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muški	1. godina	nije vjernik
25	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muški	1. godina	protivnik vjere
26	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muški	1. godina	protivnik vjere
27	vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	zenski	1. godina	uvjereni vjernik
28	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	zenski	1. godina	uvjereni vjernik
29	vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	zenski	1. godina	uvjereni vjernik
30	vjerujem	ne vjerujem	vjerujem	ne vjerujem	ne vjerujem	zenski	1. godina	uvjereni vjernik

4. Prema zadatku u ovom primjeru pripremili smo podatke na kojima sad možemo učiniti deskriptivnu analizu samo rezultata muškaraca na varijabli sreća. Vidimo da je riječ o 78 ispitanika.

Descriptive Statistics

Variables: sreća

Statistics:

- Percentile Values:
 - Quartiles
 - Cut points for: 4 equal groups
 - Percentiles:
- Central Tendency:
 - Mean
 - Median
 - Mode
 - Sum

Results

Descriptive Statistics

Descriptive Statistics	
sreća	
Valid	78
Missing	1
Mean	3.487
Std. Deviation	1.734
Minimum	1.000
Maximum	7.000

5. Nakon provedbe analize, ako nema više potrebe za filtriranjem, tj. odabirom određenih skupina ispitanika, potrebno je **isključiti filter** na način da se vratimo u prozor za odabir kategorija prema kojima se podaci filtriraju te ponovno uz svaku kategoriju označimo **kvačicu**.

The screenshot shows the SPSS 'Filter' dialog box for the variable 'spol'. The 'Filter' column has checkmarks for values 1 (muski) and 2 (zenski). Below the dialog box, a data table is displayed with columns for various variables and their values for three rows.

	nadbica1	nadbica2	nadbica3	nadbica4	nadbica5	spol	godina	religja
1	vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	uvjereni vjernik
2	ne vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	uvjereni vjernik
3	vjerujem	vjerujem	vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	uvjereni vjernik

3.2.2. Odabir prema kategoriji/ama dviju ili više nominalnih ili ordinalnih varijabli radi provedbe neke analize samo na određenoj skupini ispitanika

Primjer 2. Odabiremo samo ispitanice ženskog spola koje pohađaju tehničke ili prirodoslovne fakultete (varijabla *fakultet*) jer želimo deskriptivno analizirati samo njihove rezultate na varijabli *sreca*.

Odaberemo svaku varijablu zasebno prema kojoj ćemo filtrirati podatke. Ovdje su to nominalne varijable *spol* i *fakultet*. Redoslijed filtriranja varijabli nije relevantan, tj. svejedno je hoćete li prvo filtrirati podatke prema spolu ili prema fakultetu.

Filtriranje prema *spolu*:

Filter	Value	Label
X	1	muski
✓	2	zenski

	nadbica1	nadbica2	nadbica3	nadbica4	nadbica5	spol	godina	religja
22	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	nje vjernik
23	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	nje vjernik
24	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	nje vjernik
25	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	protivnik vjere
26	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	protivnik vjere
27	vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	zenski	1. godina	uvjereni vjernik
28	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	zenski	1. godina	uvjereni vjernik
29	vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	zenski	1. godina	uvjereni vjernik

Filtriranje prema *fakultetu*:

Filter	Value	Label
X	1	društ.-humanistički
✓	2	tehnički
✓	3	prirodoslovni
X	4	medicinsko-bioteh.
X	5	umjetnički

	nadbica1	nadbica2	nadbica3	nadbica4	nadbica5	spol	godina	religja	sreca	fakultet
29	vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	zenski	1. godina	uvjereni vjernik	5	društ.-humanistički
30	vjerujem	ne vjerujem	vjerujem	ne vjerujem	ne vjerujem	zenski	1. godina	uvjereni vjernik	2	društ.-humanistički
31	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	zenski	1. godina	uvjereni vjernik	2	društ.-humanistički
32	vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	zenski	1. godina	uvjereni vjernik	izrazito sretan	društ.-humanistički
33	ne vjerujem	vjerujem	vjerujem	vjerujem	vjerujem	zenski	1. godina	uvjereni vjernik	6	tehnički
34	vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	zenski	1. godina	vise vjernik	izrazito sretan	medicinsko-bioteh.
35	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	zenski	1. godina	vise vjernik	izrazito sretan	medicinsko-bioteh.
36	vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	zenski	1. godina	vise vjernik	izrazito sretan	medicinsko-bioteh.
37	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	zenski	1. godina	vise vjernik	3	medicinsko-bioteh.

Uočite da su ovog puta svijetlo sive boje i redci svih ispitanika (muškog spola) bez obzira na to koji fakultet pohađaju, ali i redci ispitanica (ženskog spola) koje pohađaju društveno-humanističke, medicinsko-biotehnoške ili umjetničke fakultete. To ponovno znači da su ti podaci trenutno neaktivni, odnosno da se sve daljnje analize koje zatražimo na njima neće provoditi.

Dodatno, uočite da se kraj imena obje varijable (spol i fakultet) pojavio simbol za filter koji označava da je trenutno aktivan filter na obje varijable.

Prema zadatku u ovom primjeru pripremili smo podatke na kojima sad možemo učiniti deskriptivnu analizu samo rezultata ispitanica ženskog spola koje pohađaju tehničke ili prirodoslovne fakultete na varijabli *sreca*. Vidimo da je ukupno 15 takvih ispitanica, za koje odabiremo željene deskriptive pokazatelje rezultata na varijabli *sreca*:

The image shows the SPSS software interface. On the left is the 'Descriptive Statistics' dialog box, and on the right is the 'Results' window.

Descriptive Statistics Dialog:

- Variables:** 'sreca' is selected.
- Statistics:**
 - Percentile Values:** Quartiles, Cut points for: 4 equal groups, Percentiles.
 - Dispersion:** S. E. mean, Std. deviation, MAD, MAD Robust, IQR, Variance, Range, Minimum, Maximum.
 - Central Tendency:** Mean, Median, Mode, Sum.
 - Distribution:** Skewness, Kurtosis, Shapiro-Wilk test.

Results Window - Descriptive Statistics:

Descriptive Statistics	
	sreca
Valid	15
Missing	0
Mean	4.333
Std. Deviation	1.877

3.3. Odabir pomoću procedure Filter

Primjer uključivanja jednostavnog filtera (kako provesti analizu samo na podacima žena) na stranicama JASP-a:

https://jasp-stats.org/wp-content/uploads/2018/06/02_drag_cat.gif

Napomene:

- a) Kriterij za odabir ispitanika definiramo logičkim odnosima (=, ≠, <, ≤, >, ≥,) između željene varijable i labele ili vrijednosti.
Npr. spol = ženski ili visina > 180
Potrebno je **točno upisati labelu** (prethodno redefinirati labele tako da ne sadrže dijakritičke znakove).
- b) Kompleksne filtere (oni koji sadrže dvije ili više varijabli) gradimo uz pomoć Booleovih operatora:
 - a. Operator ' ' znači 'ili' (engl. OR)
 - b. Međutim, operator '&' (znači 'i', engl. AND) nije predviđen. Dovoljno je u novi redak upisati dodatni kriterij i time koristimo ovaj operator.
- c) Zbog korištenja labele u proceduri Filter, preporučljivo je redefinirati labele u skraćene (i razumljive) varijante (npr. 'Viša škola / Preddiplomski studij' preimenovati u 'BA')
- d) Ako Filter ne funkcionira (npr. pogrešno navedena labela, labela sadrži dijakritičke znakove, upisana nepostojeća vrijednost kvantitativne varijable), pojavit će se sljedeća poruka:

Filter applied

Filtered out all data..

To znači da su svi podaci filtrirani, odnosno da niti jedna skupina podataka neće biti isključna iz budućih analiza. Dakle, filter zapravo nije uključen i potrebno je provjeriti ispravnost formulacije filtera.

3.3.1. Mogućnosti odabira skupina ispitanika preko opcije Filter


Pomoću opcije Filter moguće je uključivati jednostavne filtere (filtriranje prema vrijednostima jedne varijable), ali i složene filtere (filtriranje prema kombinacijama vrijednosti više varijabli), pri čemu su moguće kombinacije nominalnih, ordinalnih i kvantitativnih varijabli.

Npr.

- žene s rezultatom na varijabli *sreća* većim od 5 (dvije varijable: nominalna i kvantitativna)
- studenti (muškog spola) s tehničkih fakulteta koji vjeruju u *nadbica1* (vanzemaljce) (tri nominalne varijable)
- uvjereni vjernici s prve godine studija društveno-humanističkih fakulteta (tri varijable: ordinalna, ordinalna i nominalna)

3.3.2. Pokretanje procedure Filter

Procedura Filter pokreće se odabirom simbola za filter u gornjem lijevom uglu:



	nadbica1	nadbica2	nadbica3	nadbica4	nadbica5	spol	godina	religija
1	vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	uvjereni vjernik
2	ne vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	uvjereni vjernik
3	vjerujem	vjerujem	vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	uvjereni vjernik
4	vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	uvjereni vjernik
5	vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	uvjereni vjernik
6	vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	uvjereni vjernik
7	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	uvjereni vjernik
8	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	vise vjernik
9	ne vjerujem	vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	vise vjernik
10	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	ne vjerujem	muski	1. godina	vise vjernik

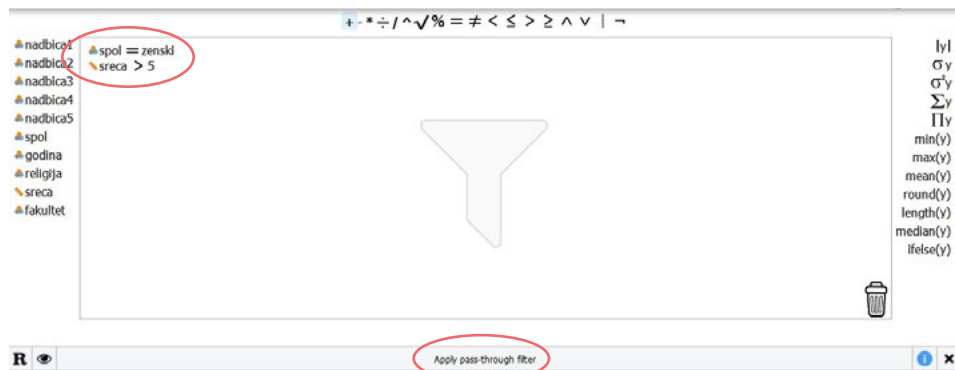
Prije postavljanja filtera važno je provjeriti jesu li svim varijablama u datoteci dodijeljene ispravne vrste varijabli. Ovdje tome nije tako pa je prvo potrebno varijable *godina* i *religija* pretvoriti u ordinalne, a varijablu *sreca* u kvantitativnu.

3.3.3. Primjeri složenih filtriranja

Primjer 1.

Navedite koliki udio studentica s rezultatom na varijabli *sreca* većim od 5 vjeruje u duhove (*nadbica2*).

Prvo je potrebno filtrirati studentice (ženskog spola) s rezultatom na varijabli *sreca* većim od 5 te potom zatražiti tablicu frekvencija varijable *nadbica2*.



Varijable prebacujemo iz popisa s lijeve strane u prozor za konstrukciju filtera, operatore ('=' i '>') biramo iz niza operatora iznad tog prozora, a nazive kategorija ili iznose vrijednosti pojedinih varijabli sami upisujemo ili koristimo opciju Copy/Paste za pojedinu labelu. Kursor miša potrebno je staviti na tri točkice nakon znaka logičkog odnosa između varijable i njezine labele/vrijednosti, a nakon unosa pritisnuti Enter.



Potom je potrebno odabrati *Apply pass-through filter* kako bismo uključili filter.

Ako je sve u redu, pisat će *Filter applied*. Filter će ostati uključen sve dok ga ne isključimo.

Sukladno zadatku u primjeru, pokrenut ćemo proceduru *Descriptives* i zatražiti tablicu frekvencija varijable *nadbica2*.

Rješenje zadatka: 81% studentica s rezultatom na varijabli *sreca* većim od 5 vjeruje u duhove.

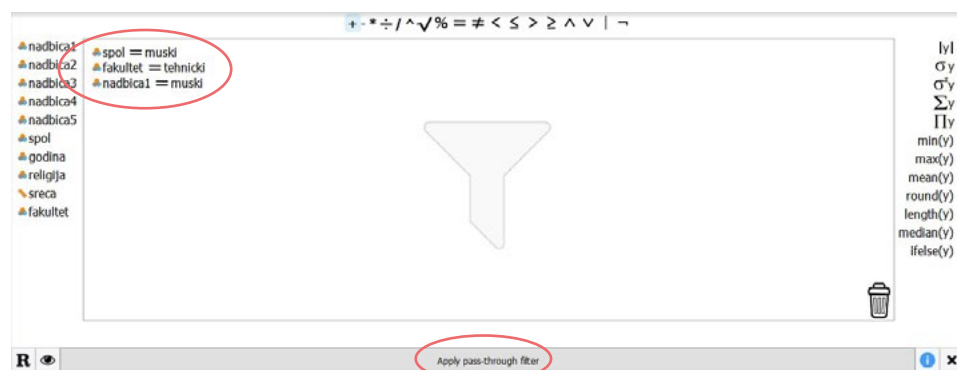
Potom je potrebno **isključiti filter**. Filter se isključuje tako da se vratimo u prozor s filterima, na jedan od dva načina:

1. Dupli klik na "kantu za smeće"  u donjem desnom kutu tog prozora. Time je automatski pokrenuto i poništavanje filtera te podaci više nisu filtrirani.
2. Metodom *drag-and-drop* sve filtere premjestimo u "kantu za smeće"  u donjem desnom kutu tog prozora te potom ponovno odaberemo *Apply pass-through filter* kako bismo isključili filter. Premještanje pojedinih varijabli u "kantu za smeće" najlakše je napraviti tako da odaberemo operator koji je postavljen između naziva varijable i naziva kategorije, odnosno vrijednosti te varijable koji služi za filtriranje. Ako je sve u redu, pisat će *Filter cleared*. Podaci više nisu filtrirani.

Primjer 2.

Napravite tablicu frekvencija varijable *religija* za studente (muškog spola) s tehničkih fakulteta koji vjeruju u vanzemaljce (*nadbica1*).

Prvo je potrebno filtrirati studente (muškog spola) s tehničkih fakulteta koji vjeruju u vanzemaljce (*nadbica1*) te potom zatražiti tablicu frekvencija varijable *religija*.



Ako je sve u redu, pisat će *Filter applied*. Filter će ostati uključen sve dok ga ne isključimo.

Sukladno zadatku u primjeru, pokrenut ćemo proceduru *Descriptives* i zatražiti tablicu frekvencija varijable *religija*.

Rješenje zadatka (tablica je dodatno uređena):

Distribucija frekvencija varijable *religija* (samo studenti s tehničkih fakulteta koji vjeruju u vanzemaljce, N=32)

religija	f	%
uvjereni vjernik	4	12,50
vise vjernik	12	37,50
ravnodušan	6	18,75
nije vjernik	7	21,88
protivnik vjere	3	9,38
UKUPNO	32	100,00

Nakon provedbe tražene analize potrebno je isključiti filter.

Primjer 3.

Odredite vrijednost medijana na varijabli *sreca* za osobe s prve godine studija društveno-humanističkih fakulteta koje se ne izjašnjavaju kao uvjereni vjernici.

Prvo je potrebno filtrirati osobe s prve godine studija (*godina*) društveno-humanističkih fakulteta (*fakultet*) koje se ne izjašnjavaju kao uvjereni vjernici (*religija*) te potom zatražiti tablicu frekvencija varijable *religija*.



Ako je sve u redu, pisat će *Filter applied*. Filter će ostati uključen sve dok ga ne isključimo.

Sukladno zadatku u primjeru, pokrenut ćemo proceduru *Descriptives* i zatražiti medijan varijable *sreca*.

Rješenje zadatka: Medijan na varijabli *sreca* za osobe s prve godine studija društveno-humanističkih fakulteta koje se ne izjašnjavaju kao uvjereni vjernici iznosi 3.

Nakon provedbe tražene analize potrebno je isključiti filter.