

Nakladnik:
Sveučilište u Splitu
Filozofski fakultet

Za nakladnika:
prof. dr. Josip Milat, dekan

Biblioteka:
Suvremena nastava
Knjiga br. 1

Recenzentice:
prof. dr. Elza Jurun
prof. dr. Anči Leburić
doc. dr. Mirjana Čižmešija

Lektorica:
Sonja Krvavica, prof.

Dizajn naslovnice:
Baldo Leburić

Grafička obrada:
doc. dr. Snježana Pivac

CIP - Katalogizacija u publikaciji
SVEUČILIŠNA KNJIŽNICA U SPLITU

UDK 311.1(075.8)

PIVAC, Snježana

Statistika za sociološka istraživanja / Snježana Pivac, Ante
Rozga. - Split : Filozofski fakultet Sveučilišta, 2006. - (Biblioteka
Suvremena nastava ; knj. 1)

Bibliografija.

ISBN 953-7395-01-4

1. Rozga, Ante

120625059

ISBN 953-7395-01-4

**SNJEŽANA PIVAC
ANTE ROZGA**

**STATISTIKA
za sociološka istraživanja**

Split, 2006.

SADRŽAJ

| | |
|--|-----|
| Predgovor | 7 |
| 1. Uvod | 9 |
| 1.1. Pojam i zadaća statistike | 9 |
| 1.2. Predmet proučavanja statistike | 10 |
| 1.3. Vrste varijabli - statistička obilježja | 11 |
| 1.4. Izvori podataka i metode njihova pribavljanja | 12 |
| 1.5. Sekundarni podaci | 13 |
| 1.6. Primarni podaci | 14 |
| 1.7. Načini prikupljanja podataka | 23 |
| 1.8. Priprema podataka za statističku analizu | 24 |
| 1.9. Uzorak | 25 |
| 2. Uređivanje i prikazivanje podataka | 35 |
| 2.1. Grupiranje podataka | 35 |
| 2.1.1. Nizovi kvalitativnih podataka | 36 |
| 2.1.2. Numerički niz | 37 |
| 2.2. Tabeliranje | 40 |
| 2.3. Relativni brojevi strukture | 43 |
| 2.4. Grafičko prikazivanje statističkih nizova | 46 |
| 2.5. Relativni brojevi koordinacije i indeksi | 71 |
| 3. Srednje vrijednosti | 103 |
| 3.1. Srednje vrijednosti | 103 |
| 3.2. Aritmetička sredina | 104 |
| 3.3. Mod | 107 |
| 3.4. Harmonijska sredina | 118 |
| 3.5. Medijan | 121 |
| 3.5.1. Kvantili | 124 |
| 4. Mjere disperzije, asimetrija i zaobljenost | 133 |
| 4.1. Mjere disperzije | 133 |
| 4.1.1. Apsolutne mjere disperzije | 133 |
| 4.1.2. Relativne mjere disperzije | 139 |

| | |
|--|------------|
| 4.1.3. Momenti numeričkih nizova | 151 |
| 4.2. Mjere asimetrije | 152 |
| 4.3. Mjera zaobljenosti | 156 |
| 5. Vjerojatnost | 167 |
| 5.1. Klasična definicija vjerojatnosti | 167 |
| 5.2. Teorem o zbrajanju vjerojatnosti | 168 |
| 5.3. Teorem o množenju vjerojatnosti | 169 |
| 5.4. Uvjetna vjerojatnost | 170 |
| 5.5. Diskontinuirana slučajna varijabla | 176 |
| 5.6. Teorijske distribucije diskontinuirane slučajne varijable | 177 |
| 5.6.1. Binomna distribucija | 177 |
| 5.6.2. Poissonova distribucija | 183 |
| 5.6.3. Dvodimenzionalna diskontinuirana distribucija ... | 189 |
| 5.6.4. Marginalne distribucije diskontinuirane slučajne varijable | 190 |
| 5.6.5. Uvjetne distribucije diskontinuirane slučajne varijable | 191 |
| 5.7. Kontinuirana slučajna varijabla | 195 |
| 5.8. Teorijske distribucije kontinuirane slučajne varijable | 196 |
| 5.8.1. Normalna distribucija | 196 |
| 5.8.2. Studentova distribucija | 198 |
| 6. Ocjene parametara osnovnog skupa na osnovi uzorka ... | 201 |
| 6.1. Uzorak | 201 |
| 6.2. Procjena aritmetičke sredine osnovnog skupa | 202 |
| 6.3. Procjena proporcije osnovnog skupa | 209 |
| 7. Analitički primjeri za vježbu | 217 |
| 8. Tablice odabranih statističkih distribucija | 257 |
| A. Površine ispod normalne krivulje | 257 |
| B. Kritične vrijednosti t, Studentove distribucije | 258 |
| 9. Literatura | 259 |
| Bilješka o autorici i autoru | 264 |

PREDGOVOR

Ovaj udžbenik namijenjen je prvenstveno studentima Odsjeka za sociologiju Filozofskog fakulteta u Splitu, ali i svim drugim zainteresiranim korisnicima koji u svom stručnom i znanstvenom radu, baveći se društvenim istraživanjima, primjenjuju statističke metode i tehnike.

Rad obuhvaća teorijske osnove i objašnjenja za svako u ovaj rad uključeno područje statistike. Kroz rješavanje velikog broja konkretnih analitičkih primjera daju se objašnjenja dobivenih rezultata i njihovo kritičko vrednovanje. Na kraju udžbenika u sklopu posljednjeg poglavlja nalazi se određeni broj analitičkih primjera koji su ostavljeni korisniku za samostalno rješavanje. Svaki primjer sadrži i rješenja u svrhu kontrole valjanosti usvojenoga gradiva. Na taj način se studentima ne ostavlja dvojbenim ni način pismene provjere znanja te ih se osposobljava da samostalno statistički analiziraju određene pojave na stručno zadovoljavajući način.

Primjeri u udžbeniku obrađeni su u SPSS - Statistical Package for the Social Sciences, koji je upravo prilagođen sociološkim istraživanjima. Ovaj program pruža puno mogućnosti za provođenje statičkih metoda i tehnika na konkretnim analizama. Naime, upotrebom statističkih paketa, počevši već od pripremne faze statističkog istraživanja, znatno se skraćuje i pojednostavljuje vrijeme potrebno za primjenu statističkih metoda i tehnika. Na taj se način statistički postupci približavaju mnogim korisnicima. Svaki analitički primjer riješen u SPSS-u u tekstu udžbenika ima oznaku u obliku slike CD-a. Takovi primjeri su priloženi u elektroničkom obliku u dokumentu SPSS i svaki od njih ima i konkretno rješenje u dokumentu Output programa SPSS.

Potrebno je napomenuti da će studenti koji nastave obrazovanje na poslijediplomskim studijima biti pripremljeni za stručni i znanstveni rad u svojim istraživanjima, gdje se redovito koristi SPSS. Dio studenata koji će se nakon diplomiranja neposredno uključiti u poslovnu praksu, također će imati koristi, jer je danas SPSS redovito najzastupljenija programska potpora u društvenim analizama.

Potrebno je napomenuti da je SPSS kompatibilan s Microsoft Excelom, pa se već postojeći podaci iz jednog programa mogu jednostavno kopirati u drugi. Microsoft Excel je jedan od najraširenijih i programa za tablične izračune te je lako dostupan većini korisnika.

U radu su korištene oznake i simboli, koji su preuzeti iz standardne statističke literature. Ako se, pak, u literaturi koriste različite oznake, upotrijebljena je češće spominjana verzija.

Na kraju istaknimo, ovaj udžbenik je nastao kao rezultat višegodišnjeg iskustva autora kod primjene statističkih metoda u izradi brojnih znanstvenih i stručnih radova, studija i analiza te kao rezultat predavačkog iskustva pri prenošenju znanja iz područja statistike na mnoge generacije studenata splitskoga Ekonomskog fakulteta i drugih društvenih znanosti.

Za suradnju i niz korisnih sugestija zahvaljujemo recenzenticama prof. dr. Elzi Jurun, prof. dr. Anči Leburić i doc. dr. Mirjani Čižmešija.

Split, rujan 2006.

Autori



1. UVOD

1.1. Pojam i zadaća statistike

Statistika je posebna znanstvena disciplina koja u svrhu realizacije postavljenih ciljeva istraživanja na organiziran način prikuplja, odabire, grupira, prezentira i vrši analizu informacija ili podataka te interpretira rezultate provedene analize.

Statistika se kao znanstvena disciplina **može podijeliti na:**

1. deskriptivnu statistiku i
 2. inferencijalnu statistiku.
1. **Deskriptivna ili opisna statistika temelji se na potpunom obuhvatu statističkog skupa čiju masu podataka organizirano prikuplja, odabire, grupira, prezentira i interpretira dobivene rezultate analize.** Na taj način se, izračunavanjem različitih karakteristika statističkog skupa, sirova statistička građa svodi na lakše razumljivu i jednostavniju formu. Ako se statističke metode i tehnike primjenjuju na **čitav statistički skup**, dakle ako su istraživanjem obuhvaćeni svi elementi skupa, oni tvore **statističku populaciju**.
 2. **Inferencijalna statistika temelji se na dijelu (uzorku) jedinica izabranih iz cjelovitoga statističkog skupa pomoću kojeg se uz primjenu odgovarajućih statističkih metoda i tehnika donose zaključci o čitavom statističkom skupu.**

Uvijek je prisutan odgovarajući stupanj rizika kada se koriste rezultati iz uzorka, za koji je poželjno da bude izabran na slučajan način i da bude reprezentativan. Inferencijalna statistika pripada skupini **induktivnih metoda**, kojima se izvode zaključci polazeći **od općega prema posebnome**.

1.2. Predmet proučavanja statistike

Statističke metode i tehnike temelj su za provođenje statističke analize **društvenih** (i prirodnih) **pojava**.

- **Predmet proučavanja statistike su određene zakonitosti koje se javljaju u masovnim pojavama. Zadaća statistike je da uoči zakonitosti u masovnim i slučajnim pojavama te da ih iskaže broјčano.**
- **Masovne pojave su skupine istovrsnih elemenata koji imaju jedno ili više zajedničkih svojstava. Takvu skupinu nazivamo statističkom masom ili statističkim skupom.**

Statistički skup potrebno je definirati pojmovno, prostorno i vremenski.

Pojmovno odrediti statistički skup podrazumijeva odrediti pojam ili svojstvo svakog elementa promatranog skupa.

Prostorno odrediti statistički skup znači odrediti prostor na koji se odnosi ili kojemu pripadaju elementi statističkog skupa.

Vremenski odrediti statistički skup znači odrediti vremenski trenutak ili razdoblje kojim će se obuhvatiti svi elementi koji ulaze u statistički skup.

Analitički primjer 1.1:

Statistički skup, "studenti 1. godine Filozofskoga fakulteta - Odsjeka za sociologiju u Splitu u Republici Hrvatskoj u akademskoj godini 2006./07., na dan 10. 10. 2006. godine", vrlo je precizno određen i iz takve njegove definicije mogu se odrediti svi njegovi elementi.

Svojstvo svakog elementa odnosno studenta definirano je najprije **pojmovno**, tj. jasno je da se radi o studentima 1. godine Filozofskoga fakulteta - Odsjeka za sociologiju.

Elementi ovog skupa određeni su i **prostorno**, tj. nalaze se u Splitu na području Republike Hrvatske.

Vremenska definicija toga statističkog skupa upućuje na vremenski interval s kojim su elementi skupa obuhvaćeni, tj. akademsku godinu 2006./07. i trenutak u kojem su studenti selektirani, tj. 10. 10. 2006. godine.

1.3. Vrste varijabli - statistička obilježja

Statistička obilježja su opće karakteristike elemenata statističkog skupa po kojima su ti elementi međusobno slični i po kojima se međusobno razlikuju.

Analitički primjer 1.2:

Statistički skup, "studenti 1. godine Filozofskoga fakulteta - Odsjeka za sociologiju u Splitu u Republici Hrvatskoj u akademskoj godini 2006./07., na dan 10. 10. 2006. godine", promatran je prema **obilježju** "spol".

Sasvim je jasno da je spol jedna od karakteristika svakog studenta. Postoji muški i ženski spol. Neki studenti su jednakog spola, a neki se razlikuju po spolu.

Potrebno je napomenuti da je u programskom paketu **SPSS** oznaka za statistička obilježja: **Variable Type**. Općenito se statistička obilježja mogu podijeliti: na

1. **kvalitativna statistička obilježja**, koja mogu poprimiti različite oblike, ali *se izražavaju opisno*. Ako se modalitetima ovog obilježja slučajno pridruže brojevi, s njima *nisu dopuštene nikakve računske operacije (u SPSS-u: Variable Type: String)*.

- 2. kvantitativna statistička obilježja** koja se izražavaju brojčano. (u SPSS-u: *Variable Type: Numeric*).

Kvalitativna statistička obilježja mogu se podijeliti: na a) **nominalna statistička obilježja** i na b) **redoslijedna statistička obilježja**.

- a) **Nominalna statistička obilježja** izražavaju se opisno.
- b) **Redoslijedna statistička obilježja** mijenjaju se prema intenzitetu ili rang. Na primjer: uspjeh na ispitu iz predmeta Statistika u sociologiji I (ovdje se javljaju modaliteti obilježja od 1 do 5, ali se njima ne vrše nikakve računске operacije, već se pomoću njih elementi skupa, odnosno studenti, mogu rangirati).

Kvantitativna statistička obilježja još se nazivaju i **numerička statistička obilježja**, a mogu se podijeliti: na a) **neprekidna ili kontinuirana statistička obilježja** i na b) **prekidna ili diskontinuirana statistička obilježja**.

- a) **Neprekidna ili kontinuirana statistička obilježja** takva su numerička obilježja koja mogu poprimiti neprebrojivo beskonačno mnogo vrijednosti. Primjeri takvog obilježja su: visina, težina, duljina, dob itd.
- b) **Prekidna ili diskontinuirana statistička obilježja** takva su numerička obilježja koja mogu poprimiti prebrojivo beskonačno mnogo vrijednosti. Primjeri takvog obilježja su: broj djece, broj učenika u razredu, starost u godinama, visina plaće u kunama itd.

1.4. Izvori podataka i metode njihova pribavljanja

Osnovne faze statističkog istraživanja su:

- a) statističko promatranje
- b) grupiranje (tabelarno i grafičko prikazivanje statističkih podataka)

c) statistička analiza i interpretacija rezultata provedene analize.

Statističko promatranje je organizirano prikupljanje statističkih podataka.

Nakon precizne definicije zadatka, cilja i predmeta istraživanja, tj. statističkog skupa, pristupa se organiziranom prikupljanju statističkih podataka. Uspješnost i objektivnost ovog prvog koraka uvjetuje kvalitetu rezultata ostalih faza statističkog istraživanja. Nepotpune i neistinite prikupljene informacije do kojih bi se došlo u ovoj fazi značile bi da konačan rezultat statističkog istraživanja sadrži grešku. Pri tom **greška može biti sistematska i slučajna**. Sistematsku grešku je lakše uočiti jer se ona ponavlja (npr. neispravnost određenoga mjernog instrumenta, neistinito izjašnjavanje ispitanika). Slučajnu grešku je teško precizno identificirati, jer se ona ne javlja kod svakog mjerenja i ne javlja se istim intenzitetom. Stoga se kod slučajne greške često veže pretpostavka o poništavanju njezin utjecaja na globalnoj razini promatranja.

U ovisnosti o **karakteru izvora podataka**, statistički podaci se dijele: na

- a) **sekundarne podatke** i na
- b) **primarne podatke**.

1.5. Sekundarni podaci

Sekundarni podaci su oni koji se pribavljaju iz već postojećih baza podataka različitih državnih ustanova. Takvi se podaci prikupljaju sustavno na odgovarajući način, a njihov opseg ne ovisi o donošenju neke poslovne odluke ili zadanom cilju nekakvog istraživanja.

Takve podatke u Hrvatskoj prikupljaju: Državni zavod za statistiku, Hrvatska narodna banka, Hrvatska gospodarska komora, te neke druge specijalizirane agencije. Jedan od najčešće korištenih sekundarnih izvora podataka u Hrvatskoj je Statistički ljetopis Hrvatske u izdanju Hrvatskog zavoda za statistiku. U svjetskim okvirima poznat je World Statistical Yearbook, a internetom su danas dostupne mnoge baze sekundarnih podataka (na primjer: Eurostat, U.S. Census Bureau i

slično). Na razini poduzeća raznovrsna specifična izvješća o poslovanju i zaposlenicima imaju sekundarni karakter.

Sekundarni podaci su uglavnom broježani. Predočeni su tablicama, a vrlo često i grafičkim prikazima.

1.6. Primarni podaci

Primarni podaci prikupljaju se neposrednim promatranjem svojstava elemenata statističkog skupa u skladu s unaprijed definiranim ciljevima statističkog istraživanja. Prikupljanje tih podataka zahtijeva definiranje statističkog skupa, izbor obilježja koja se žele istražiti, određivanje modaliteta promatranog obilježja, pripremanje anketnih upitnika i/ili pratećih formulara te organiziranje i provođenje samog prikupljanja podataka.

Ako se podaci prikupljaju za sve članove nekog skupa, takvo promatranje naziva se **census**. Primjer takvog prikupljanja je popis stanovništva. Vrlo često su takva istraživanja skupa, pa se podaci prikupljaju za podskup osnovnog skupa, odnosno za uzorak. Takvo promatranje se naziva **reprezentativno promatranje**.

Prema vremenu promatranje može se podijeliti na **jednokratno, periodično i tekuće**. Jednokratno promatranje provodi se jedanput i nema ponavljanja. Periodično promatranje se ponavlja nakon jednakih perioda (npr. bilanca uspjeha u poduzeću). Za tekuće promatranje mjerenje je kontinuirano (npr. proizvodnja mlijeka).

Prikupljanje primarnih podataka može se vršiti i pomoću **statističkih pokusa**. Statističkim pokusom se vrši mjerenje vrijednosti obilježja nastalih u kontroliranim uvjetima. Vrlo čestu primjenu statistički pokus ima u marketinškom istraživanju tržišta. Na primjer, želi se istražiti kako boja pakiranja deterdženta za rublje utječe na kupnju. U takvom slučaju istraživač će u pokusu odrediti različite boje pakiranja (vrijednosti kontroliranog obilježja) i u njima će izložiti proizvod. U određenom vremenu vršiti će se mjerenje opsega prodaje. Na taj način dobiveni statistički podaci su primarni.

Statistički podaci prikupljeni nekom od navedenih metoda predstavljaju "**sirovu**" **statističku građu** koju je potrebno na odgovarajući način urediti i pripremiti za analizu. Danas se statistički podaci uglavnom obrađuju prikladnim programima za računalo. Postoje suvremeni statistički programski paketi koji omogućavaju, uz vrlo jednostavno rukovanje, unos, pripremu i obradu prikupljenih statističkih podataka. Primjeri suvremenih statističkih paketa su: STATISTICA, SAS, SPSS. Kratica **SPSS** odgovara engleskom nazivu programa **Statistical Package for the Social Sciences**, koji je upravo prilagođen sociološkim istraživanjima. Spomenuti statistički paketi se koriste za složenije statističke analize.

Potrebno je napomenuti da je jedan od najpopularnijih softwarea (programskih jezika) za tablične proračune u provođenju različitih aspekata statističke analize i Microsoft Excel. U području programa koji su namijenjeni prvenstveno tabličnim proračunima su Lotus 1-2-3 i QuatroPro, ali upotreba MS Excela je najraširenija.

MS Excel je u funkciji svih bitnih faza procesa statističke analize od formiranja baze podataka, sređivanja i grupiranja podataka, grafičkog prikazivanja statističkih nizova, izračuna temeljnih karakteristika statističkog niza pa sve do složenijih statističkih analiza i procedura i analize vremenskih nizova. Ovaj program je kompatibilan sa SPSS-om, pa se već postojeći podaci iz jednog programa mogu jednostavno kopirati u drugi.

Upotrebom statističkih paketa, počevši već od pripremne faze statističkog istraživanja, znatno se skraćuje i pojednostavljuje vrijeme potrebno za primjenu statističkih metoda i tehnika. Na taj se način statistički postupci približavaju mnogim korisnicima.

Prikupljene statističke podatke potrebno je unijeti i pohraniti u odgovarajuću datoteku odabranog statističkog paketa. Kada se unose podaci kvalitativnog ili opisnog karaktera, potrebno je izvršiti šifriranje ili kodiranje vrijednosti takvih obilježja.

U bilo kojem statističkom programskom paketu, pa tako i u **SPSS**-u, potrebno je za svaku varijablu odabrati odgovarajuću skalu mjerenja. U **SPSS**-u oznaka za skalu mjerenja je **Measure** i postoje 3 vrste skale mjerenja:

- **Nominal:** *nominalne skale*, za koje se umjesto imena predmeta navodi njegov broj. Brojevi služe samo za identifikaciju.
- **Ordinal:** *ordinalne skale* koje služe za označavanje redoslijeda (određuju je li nešto veće ili manje od drugoga).
- **Scale:** *numerička statistička obilježja* su vezana za *intervalnu i omjernu skalu*.

Intervalna skala nema apsolutnu nulu (nula u ovom slučaju ne znači nepostojanje promatrane pojave). Na primjer: obilježje "temperatura zraka" 0°C ne upućuje na nepostojanje temperature. *Vrijednosti ove skale ne mogu se dijeliti*, (5°C na nekom području u odnosu na 15°C na nekom drugom području ne znači da je na jednom mjestu bilo tri puta hladnije u odnosu na to drugo).

Omjerna skala ima apsolutnu nulu (nula podrazumijeva nepostojanje svojstva). *Omjeri znače i jednake odnose u mjerenom svojstvu*. Na primjer: dužina, težina, visina uštedevine na računu u banci.

Analitički primjer 1.3:

Zaposleni u poduzeću "P" u mjestu "M" na dan 30. 09. 2006. godine statistički su skup. Promatrani su prema sljedećim obilježjima:

- dob, godine staža, spol, stručna sprema, plaća u kunama, visina, težina, bračno stanje, stambena situacija, broj članova u obitelji, mjesto rođenja.

a) Zadatak je utvrditi kojim vrstama statističkih obilježja pripadaju navedena obilježja?

Rješenje 1.3:

Navedena obilježja pripadaju sljedećim vrstama:

- dob: numeričko (kontinuirano) obilježje
- godine staža: numeričko (diskontinuirano) obilježje
- spol: nominalno obilježje
- stručna sprema: redoslijedno obilježje
- plaća: numeričko (diskontinuirano) obilježje

- visina: numeričko (kontinuirano) obilježje
- težina: numeričko (kontinuirano) obilježje
- bračno stanje: nominalno obilježje
- stambena situacija: nominalno obilježje
- broj članova u obitelji: numeričko (diskontinuirano) obilježje
- mjesto rođenja: nominalno (zemljopisno) obilježje.

Analitički primjer 1.4:

Na odabranom području izvršena je analiza zainteresiranosti učenika i studenata za poduke iz različitih predmeta upitnikom u kojem su između ostalih bila sljedeća pitanja:

1. Iz kojih predmeta ste pohađali poduke?
 - a) Matematika
 - b) Fizika
 - c) Engleski jezik
 - d) Hrvatski jezik
 - e) Ostalo
2. Željeli biste pohađati:
 - a) Individualne poduke
 - b) Grupne poduke
 - c) Svejedno mi je
3. Spol
 - a) žensko
 - b) muško

- a) Zadatak je izvršiti kodiranje odabranih pitanja?

Rješenje 1.4:

Tablica 1.1.

Kodna lista

| OBIJEŽJE | OBLICI OBIJEŽJA | KOD |
|---------------------------------|--|-----------------------|
| poduke iz željenog predmeta(V1) | Matematika Fizika Engleski jezik Hrvatski jezik Ostalo | A B C D E |
| željena vrsta poduka (V2) | Individualne poduke Grupne poduke Nije bitno | A B C |
| spol (V3) | ženski muški | 1 2 |

Izvor: Simulirani podaci.

U tablici 1.1. vidi se da obilježje "poduka iz željenog predmeta" može poprimiti 5 oblika: Matematika, Fizika, Engleski jezik, Hrvatski jezik i ostalo. Svakom obliku obilježja dodijeljena je šifra ili kod: A, B, C, D i E. Na taj način pri unosu podataka umjesto da se pišu čitave riječi upisuje se šifra, čime se postiže velika ušteda vremena.

Kodovi ne moraju biti slova, pa se kod obilježja spol za "ženski" unosi 1, a za "muški" 2.

Moderni statistički paketi imaju mogućnosti kreiranja i grupiranja unesenih podataka u različitim željenim varijantama u tablice, što opet ovisi o vrsti istraživanja i postavljenim ciljevima statističkog istraživanja.

Analitički primjer 1.5:

Kodirana pitanja iz 1.4 potrebno je za 10 ispitanika unijeti u tablicu!

Rješenje 1.5:

U tablici 1.2. prikazana su kodirana pitanja za 10 ispitanika:

Tablica 1.2.

Tablica s kodiranim podacima (ili matrica podataka)

| Ispitanici | V1 | V2 | V3 |
|------------|----|----|----|
| 1 | A | A | 1 |
| 2 | C | A | 2 |
| 3 | A | B | 1 |
| 4 | B | C | 1 |
| 5 | A | A | 2 |
| 6 | A | B | 1 |
| 7 | D | A | 2 |
| 8 | E | C | 1 |
| 9 | B | A | 2 |
| 10 | A | A | 2 |

Izvor: Simulirani podaci.

U tablici 1.2 vrste statističkih obilježja smještene su u stupce. U drugom stupcu tablice obilježje V1, prema kodnoj listi u tablici 1.1., znači "poduke iz željenog predmeta". Stupac V2 podrazumijeva "željenu vrstu poduke", a V3 znači "spol".

Ako se pogledaju podaci za 3. ispitanika u tablici 2, A u stupcu obilježja V1 (poduke iz željenog predmeta) znači da taj ispitanik želi poduku iz Matematike. Oznaka B u stupcu V2 (željena vrsta poduka) istog ispitanika podrazumijeva da on preferira grupnu poduku, a njegova oznaka 1 u stupcu V3 (spol) kaže nam da je taj ispitanik ženskog spola.



Analitički primjer 1.6:

Na odabranom području izvršena je analiza stanovništva upitnikom u kojem su između ostalih bila sljedeća pitanja:

1. Spol:
 - a) Muški
 - b) Ženski

2. Visina:

.....

Tablica 1.3:

Kodna lista

| OBIJEŽJE | OBLICI OBIJEŽJA | KOD |
|------------------|-----------------|-------|
| spol (V1) | Muški | 1 |
| | Ženski | 2 |
| visina u cm (V2) | | |

Izvor: Simulirani podaci.

- a) Kodirana pitanja iz tablice 1.3. potrebno je unijeti u dokument programa SPSS!
- b) Za 20 ispitanika potrebno je u SPSS unijeti konkretne podatke o spolu i visini u cm!

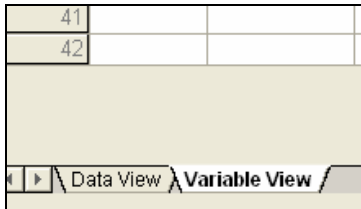


Rješenje 1.6:

- a) U programu **SPSS** pitanja, odnosno varijable se definiraju u dijelu **Variable view** koji se bira u donjem lijevom kutu ekrana (slika 1.1.)

Slika 1.1.

Variable view za definiranje varijabli u programu SPSS



Izvor: Iz programa SPSS.

Slika 1.2.

Prikaz definiranih varijabli u programu SPSS

| 1Vjezbe Sociologija.sav - SPSS Data Editor | | | | | | | | | | |
|--|------|---------|-------|----------|-------------|---------------|---------|---------|-------|---------|
| File Edit View Data Transform Analyze Graphs Utilities Window Help | | | | | | | | | | |
| | Name | Type | Width | Decimals | Label | Values | Missing | Columns | Align | Measure |
| 1 | V1 | Numeric | 3 | 0 | Spol | {1, musko}... | None | 8 | Right | Nominal |
| 2 | V2 | Numeric | 3 | 0 | Visina u cm | None | None | 8 | Right | Scale |

Izvor: Simulirani podaci.

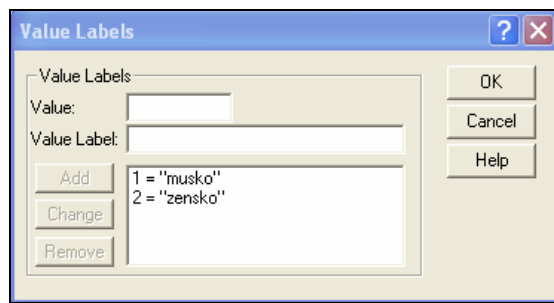
Prema slici 1.2 vidi se da su varijable nazvane (**Name**) V1 i V2. Radi lakšeg rada s podacima naznačena je vrsta podataka (**Type**) **Numeric**. U stupcu **Label** označeno je da je varijabla V1: Spol, a varijabla V2: Visina u cm. Zadnji stupac definira skalu mjerenja (**Measure**), gdje je za V1-Spol označena opisna, tj. **Nominal** skala, a za V2-Visinu u cm numerička, tj. **Scale** skala.

U stupcu **Values** za varijablu V1-Spol uneseni su kodovi prema tablici 1.3. Klikom miša u prvo polje stupca Values otvara se prozor Value Labels, u kojem se definira konkretno kodiranje. To je za ovaj primjer prikazano na slici 1.3.

Na slici 1.3 vidi se da muški spol ima kod 1, a ženski spol kod 2. Potrebno je napomenuti da pri unosu objašnjenja i vrijednosti treba izbjegavati slova koja ne pripadaju engleskoj abecedi jer ih **SPSS** neće prepoznati.

Slika 1.3.

Prikaz definiranih varijabli u programu SPSS

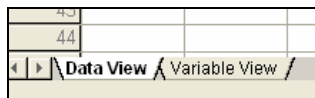


Izvor: Prema podacima autora.

b) U programu SPSS podaci se prema definiranim varijablama unose u dijelu **Data view** koji se bira u donjem lijevom kutu ekrana (slika 1.4).

Slika 1.4.

Data view za unošenje podataka u programu SPSS



Izvor: Iz programa SPSS.

Na slici 1.5 prikazani su podaci o spolu i visini u cm za 20 ispitanika (podaci su simulirani).

Naravno, nakon što su se podaci unijeli u program **SPSS**, potrebno ih je snimiti na standardan način kako se to radi kod većine klasičnih programa za računalo. Na glavnom izborniku odabire se **File**, a u njegovu padajućem izborniku **Save As** te se uneseni podaci snimaju pod željenim imenom.

Slika 1.5.

Prikaz podataka o varijablama V1 i V2 za 20 ispitanika u programu SPSS

| | V1 | V2 | var | var |
|----|----|-----|-----|-----|
| 1 | 1 | 185 | | |
| 2 | 2 | 177 | | |
| 3 | 1 | 188 | | |
| 4 | 1 | 190 | | |
| 5 | 2 | 176 | | |
| 6 | 2 | 160 | | |
| 7 | 1 | 173 | | |
| 8 | 2 | 176 | | |
| 9 | 2 | 186 | | |
| 10 | 2 | 173 | | |
| 11 | 2 | 170 | | |
| 12 | 2 | 159 | | |
| 13 | 2 | 171 | | |
| 14 | 2 | 172 | | |
| 15 | 1 | 185 | | |
| 16 | 1 | 186 | | |
| 17 | 1 | 193 | | |
| 18 | 1 | 181 | | |
| 19 | 1 | 175 | | |
| 20 | 2 | 174 | | |
| 21 | | | | |
| 22 | | | | |
| 23 | | | | |

Izvor: Simulirani podaci.

Moderni statistički paketi imaju mogućnosti kreiranja i grupiranja unesenih podataka u različitim željenim varijantama u tablice, što opet ovisi o vrsti istraživanja i postavljenim ciljevima statističkog istraživanja.

Takav način obrade podataka olakšava rad s masom statističkih podataka i omogućava veliku uštedu vremena.

1.7. Načini prikupljanja podataka

Pojedinačne metode statističkog promatranja su:

- a) mjerenje
- b) brojenje
- c) ocjenjivanje
- d) evidentiranje i
- e) anketiranje.

Jedan od načina statističkog promatranja i prikupljanja podataka je **mjerenje**. Mjeri se na primjer urod pšenice po jedinici poljoprivredne površine. Može se mjeriti težina proizvoda, kao i visina stanovništva nekog područja.

Brojenjem se može doći do podataka o broju zaposlenih u pojedinim organizacijskim jedinicama poduzeća, o broju upisanih učenika u srednje škole. Broje se i noćenja turista u turističkoj sezoni pa se na taj način dobivaju podaci relevantni za analizu uspješnosti sezone.

Ocjenjivanjem kao metodom prikupljanja statističkih podataka određuje se kvaliteta provođenja određenih radnji, stoga se taj postupak obično veže za redoslijedna obilježja statističkog skupa. Ocjenjuju se i rangiraju studenti na testu iz statistike, ocjenjuje se usluga nekog hotelskog poduzeća i slično.

Evidentiranje podataka podrazumijeva kontinuirano praćenje kretanja neke pojave u duljem ili kraćem razdoblju. Za potrebe evidencije često se uređuju odgovarajući obrasci. Na primjer, za promatranje izostanaka i broja ostvarenih radnih sati djelatnika u nekom poduzeću postoji obrazac za evidenciju gdje zaposlenici upisuju: ime i prezime, vrijeme dolaska i odlaska s radnog mjesta, vrijeme izostanka, razlog izostanka. Evidentiranje se može vršiti i tehničkim uređajima. To su umrežena računala, optički čitači, uređaji za brojenje npr. putnika i slično.

Anketa i/ili intervju je metoda kojom se prikupljaju podaci uz pomoć unaprijed pripremljenih upitnika na kojima ispitanici svojim

odgovorima daju informacije o promatranim obilježjima statističkog skupa. Da bi anketa uspjela, potrebno je veliku pozornost obratiti sastavljanju upitnika. Anketu sastavlja statističar, koji se često konzultira i sa psihologom. Upiti moraju biti kratki, precizni i jasni. Moraju biti postavljeni tako da ne sugeriraju odgovor. Broj pitanja ne smije biti velik da ne zamara one koji odgovaraju. Pri provođenju ankete pristup ispitaniku, odnosno jedinici statističkog skupa može biti izravan i neizravan.

Izravan pristup ostvaruje se kada osoba ili tim koji anketiraju izlaze na teren i u direktnom kontaktu s ispitanicima prikupljaju odgovore na pitanja iz upitnika. **Neizravan pristup** ostvaruje se poštom, telefonom, elektroničkom poštom i web-om. Na taj način smanjuju se troškovi prikupljanja podataka (npr. putni troškovi osoba koje provode anketu, tj. anketara). Iako se i neizravnim načinom ispitanicima prezentira tko istražuje i koja mu je svrha, praksa je pokazala da je tim pristupom anketiranja prisutan velik postotak neodaziva, kao i često velik postotak nevaljano i nepotpuno popunjenih upitnika. Naravno, razloge treba tražiti u činjenici da ispitaniku nije na raspolaganju osoba koja će mu pojasniti nejasna pitanja.

Ovisno o obimu istraživanja **organizaciju prikupljanja podataka može provoditi jedna osoba, skupina istraživača ili osoblje specijalizirane ustanove** kojoj je to osnovna djelatnost. Ako se radi o manjem istraživanju, u prikupljanju podataka će sudjelovati manji broj istraživača, dok će veće istraživanje zahtijevati rad većeg broja istraživača.

1.8. Priprema podataka za statističku analizu

Prije početka obrade prikupljenih podataka potrebno je izvršiti kontrolu sirove statističke građe. Kontrola se može vršiti u tijeku ili na kraju postupka prikupljanja podataka, što ovisi i o različitim metodama prikupljanja.

Preventivna kontrola obavlja se već u tijeku samog postupka prikupljanja statističkih podataka. Kod provođenja ankete to

podrazumijeva kontrolu upitnika pri njegovu preuzimanju od ispitanika. Kontrolira se jesu li dani odgovori na sva pitanja i jesu li pravilno popunjena predviđena mjesta za tražene odgovore.

Naknadna kontrola prikupljenih podataka obavlja se nakon postupka prikupljanja. **Formalnom kontrolom** se uspoređuje realizirani broj prikupljenih podataka s onim planiranim obuhvatom statističkog skupa. Ako je anketa vršena neizravnim putem, na primjer poštom, uspoređuje se broj odaslanih upitnika s brojem vraćenih upitnika. **Materijalnom kontrolom** ispituje se potpunost i točnost sadržaja prikupljenih podataka. Na primjeru ankete to podrazumijeva kontrolu potpunosti i logičnosti danih odgovora.

1.9. Uzorak

Uzorak je podskup osnovnoga statističkog skupa, a uzima se u svrhu ispitivanja obilježja elemenata osnovnog skupa (ili populacije).

Poželjno je da uzorak bude **što veći**, ali u konkretnim istraživanjima može se dogoditi da:

- povećavanjem uzorka istraživanje postaje sve skuplje
- katkada se uzorci uništavaju (npr. kemijska analiza nekih prehrambenih artikala).

Kad god je moguće, poželjno je izabrati **slučajni uzorak**. U takvom uzorku svaka jedinica populacije (osnovnog skupa) ima jednaku vjerojatnost da bude izabrana. Uzorak **treba biti reprezentativan**, a **ne selekcioniran** (npr. potrebno je obuhvatiti istraživanjem i gradsko i seosko stanovništvo, zatim u nekom medicinskom istraživanju nije cilj samo analizirati i donositi zaključke za "dobrovoljce").

Ako neki članovi populacije imaju veću šansu od drugih da budu izabrani, takav uzorak se naziva **pristrani uzorak** (biased sample).

Slučajni uzorak sastavlja se prema određenim principima, koji odgovaraju zakonu slučaja. Najbolji način je upotreba "tablice slučajnih

brojeva" ili korištenje računalnog sustava slučajnog izbora (generator slučajnih brojeva).

Pri izradi **tablice slučajnih brojeva** koristi se deseterostrana prizma s brojevima od 0 do 9.

Analitički primjer 1.7:

Nakon 40 bacanja deseterostrane prizme dobije se na slučajan način 40 brojeva; skupljeni u skupine od po 4:

7766 7520 1607 6048 2771 4733 8558 8681 5204 3806

a) Zadatak je izabrati uzorak veličine 350 iz osnovnog skupa veličine 790. (Prije početka odabira svakoj osobi iz osnovnog skupa dodjeljuje se broj od 1 do 790!).

Rješenje 1.7:

a) Najprije je potrebno otvoriti tablicu slučajnih brojeva *slučajno!* Pretpostavka je da se tablica otvorila na stranicu na kojoj je u prvom retku prikazan niz od 40 brojeva.

Kako je u osnovnom skupu ukupno 790 osoba, potrebno je čitati troznamenaste brojeve iz tablice (u redovima, u stupcima ili dijagonalno).

Ako se brojevi čitaju po redovima, prvi troznamenasti broj iz prezentiranog retka, tj. prva odabrana osoba je ona označena brojem 776, zatim 675, pa 201 itd.

Ako se naiđe na broj veći od 790, on se odbacuje, jer u osnovnom skupu nema više od 790 osoba. Ako se dogodi da se neki broj ponavlja, on se opet odbacuje, jer se jedna osoba u uzorak ne može odabrati dva puta.

Uzorak može biti i **sistematski** ako se jedinice iz osnovnog skupa biraju sistematski. Na primjer, ako se po redu u uzorak bira svaki 10. element iz osnovnog skupa.

Stratificirani ili slojeviti uzorak je takav uzorak koji se dobije na način da se populacija podijeli u slojeve ili stratumе prema nekim

karakteristikama, te da se iz svake od grupa uzme slučajni uzorak. Na primjer, u sociološkim istraživanjima stratumi se mogu birati prema dobnim skupinama.

Klaster uzorci su lošija varijanta slučajnog uzorka i koriste se u velikim tržišnim, ekonomskim ili političkim istraživanjima. Na primjer, pri ispitivanju mišljenja stanovnika nekog grada o nekoj problematici, grad se može prema planu podijeliti na 50-ak blokova, odnosno kvartova. Tada se na slučaj biraju neki blokovi u kojima anketari *detaljno* intervjuiraju sve stanovnike. Čak se vraćaju na adrese dok ne dobiju intervju od svakog stanovnika u odabranim blokovima.

Kvotni uzorci su još lošiji jer predstavljaju neslučajni stratificirani uzorak. Koriste se kod ad hoc organiziranih istraživanja za potrebe tržišta, za prikupljanje mišljenja građana o nekom pitanju i slično. Istraživač unaprijed izabere broj ljudi (kvotu) svakog pojedinog stratuma koje mora intervjuirati, stoga se ti uzorci nazivaju kvotni.

Prigodni uzorak je onaj koji se "nađe pri ruci" jer je drugi nedostupan. Na primjer, dostupni bolesnici na odjelu u bolnici, prisutni studenti neke godine studija i slično.

Pri istraživanju se mogu dogoditi **pogreške** koje mogu biti:

- a) **pogreške izvan uzorka** (zbog pogrešno odabrane metode istraživanja, zbog pogrešne obrade rezultata i slično)
- b) **pogreške uzorka.**

Veličinu uzorka nije moguće općenito definirati, jer to ovisi o varijabilnosti pojave koja se mjeri i o preciznosti kojom se pojava želi izmjeriti. Za uzorak je najvažnije da bude reprezentativan, tj. kad god je to moguće slučajno sakupljen.

Frakcija odabiranja (f) omjer je jedinica u uzorku i broja jedinica u osnovnom skupu:

$$f = \frac{n}{N} \quad (1.1)$$

gdje je :

n - broj jedinica u uzorku

N - broj jedinica u osnovnom skupu.

Korak izbora je recipročna vrijednost frakcije odabiranja ($\frac{1}{f}$) i upotrebljava se kod sistematskog izbora jedinica u uzorak. To znači da ako je korak izbora jednak 20, da se u uzorak iz osnovnog skupa odabire svaki 20. element.

Broj svih mogućih uzoraka (bez ponavljanja) veličine n iz osnovnog skupa veličine N jednak je broju kombinacija bez ponavljanja n -tog razreda:

$$K = \binom{N}{n}. \quad (1.2)$$

Pomoću uzorka vrši se procjena određenih parametara osnovnog skupa i testiraju se hipoteze o nepoznatim parametrima osnovnog skupa.



Analitički primjer 1.8:

Na odabranom području izvršena je analiza studentske populacije upitnikom u kojem su između ostalih bila sljedeća pitanja:

Anketni upitnik:

1. Spol:
 1. Muški
 2. Ženski
2. Visina u cm:
3. Težina u kg:
4. Prosječna ocjena u srednjoj školi:
5. Koliki džeparac dobivate na mjesec u kunama?
6. Financijske prilike u obitelji:
 1. Nedovoljno novca za normalni standard
 2. Zadovoljavamo osnovne potrebe
 3. Nema materijalnih problema
 4. Živimo odlično.

a) Kodirana pitanja iz anketnog upitnika potrebno je unijeti u dokument programa SPSS!

b) Za svakog ispitanika potrebno je u SPSS unijeti konkretne podatke o postavljenim pitanjima iz ankete!



Rješenje 1.8:

a) U programu **SPSS** pitanja, odnosno varijable se definiraju u dijelu **Variable view** koji se bira u donjem lijevom kutu ekrana (slika 1.1.).

Slika 1.6.

Prikaz definiranih varijabli u programu SPSS

| | Name | Type | Width | Decimals | Label | Values | Missing | Columns | Align | Measure |
|---|------|---------|-------|----------|--------------------------------|---------------|---------|---------|-------|---------|
| 1 | v1 | Numeric | 3 | 0 | Spol | (1, muško)... | None | 8 | Right | Nominal |
| 2 | v2 | Numeric | 3 | 0 | Visina u cm | None | None | 8 | Right | Scale |
| 3 | v3 | Numeric | 11 | 2 | Tezina u kg | None | None | 8 | Right | Scale |
| 4 | v4 | Numeric | 11 | 2 | Prosječna ocjena u srednjo | None | None | 8 | Right | Ordinal |
| 5 | v5 | Numeric | 11 | 0 | Financijske prilike u obitelji | (1, Nema novc | None | 8 | Right | Ordinal |
| 6 | v6 | Numeric | 11 | 2 | Koliko džeparac dobivate | None | None | 8 | Right | Scale |
| 7 | | | | | | | | | | |

Izvor: Simulirani podaci.

Prema slici 1.6. vidi se da su varijable nazvane (**Name**) od v1 do v2. Radi lakšeg rada s podacima, naznačena je vrsta podataka za sve varijable (**Type**) Numeric. U stupcu **Label** označeno je da je varijabla V1: Spol, a varijabla V2: Visina u cm, itd. Zadnji stupac definira skalu mjerenja (**Measure**).

Slika 1.7.

Prikaz definirane varijable v6-Financijske prilike u obitelji u programu SPSS

Value Labels

Value Labels: [Empty field]

Value: [Empty field]

Value Label: [Empty field]

Buttons: Add, Change, Remove, OK, Cancel, Help

List of labels:

- 1 = "Nema novca za normalan standar"
- 2 = "Zadovoljavamo osnovne potrebe"
- 3 = "Nema materijalnih problema"
- 4 = "Zivimo odlicno"

Izvor: Simulirani podaci.

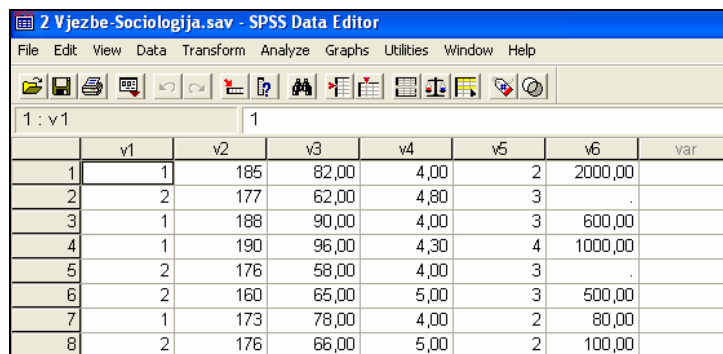
U stupcu *Values* za varijable v1-Spol i v6-Financijske prilike u obitelji uneseni su odgovarajući kodovi prema odgovarajućim pitanjima iz ankete. Na slici 1.7 prikazani su kodovi za varijablu v6-Financijske prilike u obitelji.

Na slici 1.7 vidi se da odgovor "Nema novca za normalan standard" kod 1, odgovor "Zadovoljavamo osnovne potrebe" kod 2 itd..

b) U programu *SPSS* podaci prema definiranim varijablama unose se u dijelu *Data view* koji se bira u donjem lijevom kutu ekrana (slika 1.4).

Slika 1.8.

Prikaz dijela podataka o varijablama v1 - v6 za odabrane ispitanike u programu SPSS



| | v1 | v2 | v3 | v4 | v5 | v6 | var |
|---|----|-----|-------|------|----|---------|-----|
| 1 | 1 | 185 | 82,00 | 4,00 | 2 | 2000,00 | |
| 2 | 2 | 177 | 62,00 | 4,80 | 3 | . | |
| 3 | 1 | 188 | 90,00 | 4,00 | 3 | 600,00 | |
| 4 | 1 | 190 | 96,00 | 4,30 | 4 | 1000,00 | |
| 5 | 2 | 176 | 58,00 | 4,00 | 3 | . | |
| 6 | 2 | 160 | 65,00 | 5,00 | 3 | 500,00 | |
| 7 | 1 | 173 | 78,00 | 4,00 | 2 | 80,00 | |
| 8 | 2 | 176 | 66,00 | 5,00 | 2 | 100,00 | |

Izvor: Simulirani podaci.

Na slici 1.8 prikazan je dio podataka o ispitanicima za varijable v1 - v6 (podaci su simulirani).

Naravno, nakon što su se podaci unijeli u program *SPSS*, potrebno ih je snimiti na standardan način kako se to radi kod većine klasičnih programa za računalu. Na glavnom izborniku odabire se *File*, a u njegovom padajućem izborniku *Save As* te se uneseni podaci snimaju pod željenim imenom.



Analitički primjer 1.9:

Prema podacima iz analitičkog primjera 1.8:

- a) Za sve ispitanike potrebno je izračunati indeks kojim se određuje idealna tjelesna težina u skladu s visinom ispitanika. To je Body Mass Index (BMI) - (v7) koji se računa prema sljedećem izrazu:

$$BMI = \frac{\text{težina u kg}}{\text{visina u m}^2}.$$

- b) Zadatak je za sve ispitanike od numeričke varijable džeparac (v6) formirati novu varijablu grupirani džeparac (v8) po principu koji je prikazan u tablici 1.4.

Tablica 1.4.

Princip formiranja nove varijable grupirani džeparac (v8)

| Grupe | Visina džeparca | Nova vrijednost |
|-------|-----------------|-----------------|
| 1 | 0 - 200 | 200 |
| 2 | 201 - 400 | 400 |
| 3 | 401 - 600 | 600 |
| 4 | 601 - 800 | 800 |
| 5 | 800 - 1000 | 1000 |
| 6 | 1001 - 3500 | 3500 |

Izvor: Simulirani podaci.

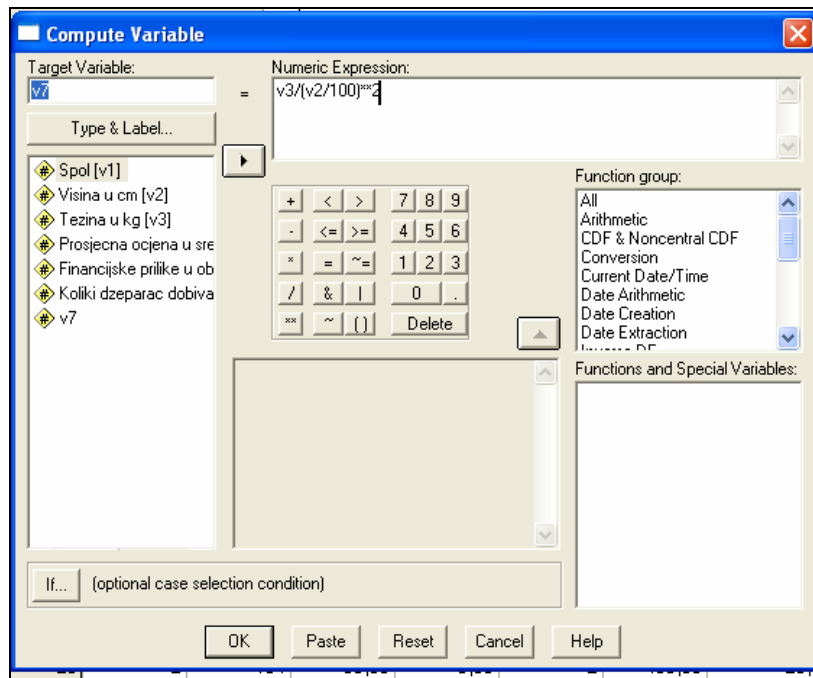


Rješenje 1.9:

- a) Da bi se pomoću postojećih numeričkih varijabli izračunala nova numerička varijabla, potrebno je na glavnom izborniku odabrati **Transform**, a u njegovu padajućem izborniku **Compute**. Na ekranu se pojavi prozor **Compute Variable** koji je prikazan na slici 1.9.

Slika 1.9.

Prozor Compute Variable u kojem se kreira nova numerička varijabla v7



Izvor: Simulirani podaci.

Prema slici 1.9 vidi se da je u polje **Target Variable** potrebno unijeti ime nove varijable BMI -v7, dok se u **Numeric Expression** unosi izraz po kojem se iz postojećih varijabli kreira nova varijabla. Pritom treba voditi računa o tome da su postojeći podaci o visini ispitanika izraženi u cm te da bi se dobila visina izražena u metrima, potrebno je centimetre podijeliti sa 100. To se može vidjeti na slici 1.9 pod **Numeric Expression**. Nakon što se klikne ikona **OK**, nova varijabla v7 automatski se formira u dodatnom stupcu u **Data View** programa **SPSS**. Nova varijabla se može vidjeti na slici 1.10.

U **Variable View** potrebno je na odgovarajući način definirati tu novu varijablu te u stupcu **Label** naznačiti da se radi o Body Mass Index. Ta nova varijabla je numerička i ima skalu mjerenja **Scale**.

Slika 1.10.

Data view s novom varijablom v7

| v4 | v5 | v6 | v7 | var |
|------|----|---------|-------|-----|
| 4,00 | 2 | 2000,00 | 23,96 | |
| 4,80 | 3 | . | 19,79 | |
| 4,00 | 3 | 600,00 | 25,46 | |
| 4,30 | 4 | 1000,00 | 26,59 | |
| 4,00 | 3 | . | 18,72 | |
| 5,00 | 3 | 500,00 | 25,39 | |
| 4,00 | 2 | 80,00 | 26,06 | |
| 5,00 | 2 | 100,00 | 21,31 | |
| 5,00 | 3 | 100,00 | 20,23 | |
| 5,00 | 2 | 200,00 | 19,38 | |
| 4,30 | 2 | 00 | 17,38 | |

Izvor: Simulirani podaci.

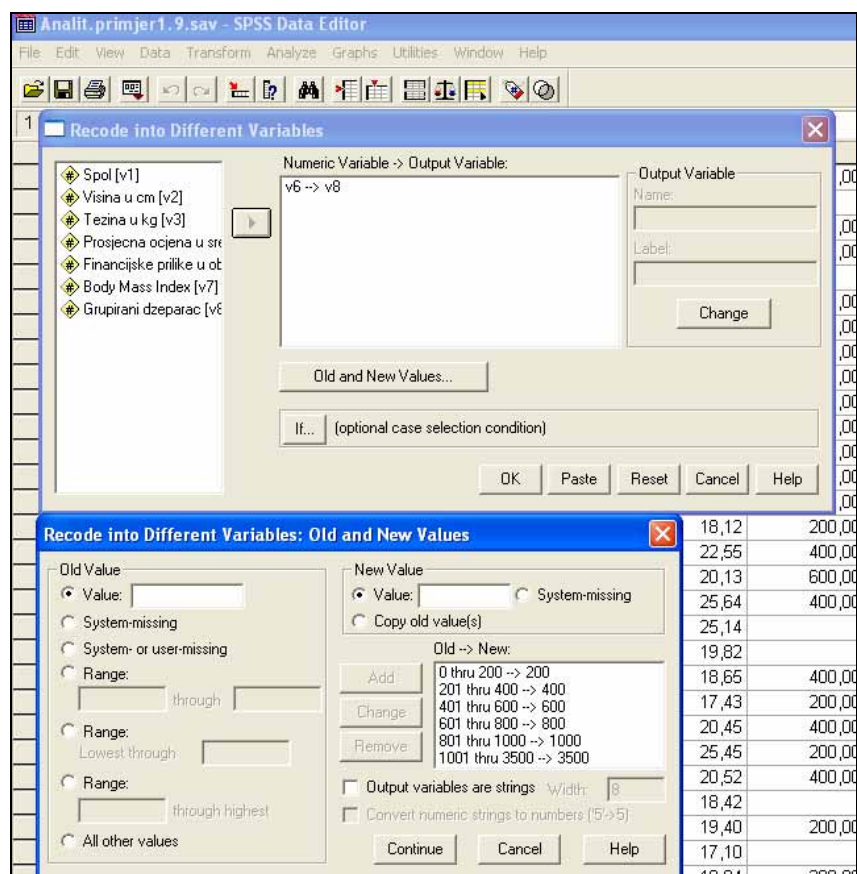
b) Da bi se pomoću varijable v6 - džeparac formirala nova varijabla grupirani džeparac po principu koji je prikazan u tablici 1.4, potrebno je na glavnom izborniku odabrati **Transform**, a u njegovu padajućem izborniku odabrati **Recode**. U njegovom izborniku bira se **Into Different Variables** (jer bi se odabirom **Into Same Variables** izgubila postojeća varijabla). U polje **Numeric Variable** bira se v6 - "Koliki džeparac dobivate", a u **Output Variable, Name** je v8, a **Label** Grupirani džeparac.

Odabirom ikone **Old and New Values** otvara se novi prozor **Recode into Different Variables**. Aktiviranjem opcije **Range** upisuju se odgovarajuće grupe visine džeparca, kako je zadano u tablici 1.4. Nakon upisivanja svakog pojedinog **Range**, upisuje se odgovarajuća **New Value**, koja se svaki put kao nova grupa dodaje klikom na ikonu **Add**. Nakon što su unesene sve grupe ikonom **Continue**, zatim **Change** i **OK**, u novom stupcu **Data View** kreirat će se nova varijabla Grupirani džeparac, odnosno v8.

Na slici 1.11. prikazani su prozori **Recode into Different Variables** i **Old and New Values**.

Slika 1.11.

Prozori Recode into Different Variables pomoću kojih se kreira nova varijabla v8 - Grupirani džeparac



Izvor: Simulirani podaci.

2. UREĐIVANJE I PRIKAZIVANJE PODATAKA

2.1. Grupiranje podataka

Ako se urede prikupljeni statistički podaci prema nekom obilježju ili karakteristici, dobiva se statistički niz.

Grupiranje statističkih podataka je postupak diobe statističkog skupa na određeni broj podskupova prema prethodno utvrđenim modalitetima promatranog obilježja i uz poštivanje principa isključivosti i iscrpnosti.

Princip isključivosti podrazumijeva da svaki element osnovnog skupa istovremeno može pripadati samo jednoj grupi, tj. podskupu. Princip iscrpnosti podrazumijeva da postupkom grupiranja trebaju biti obuhvaćeni svi elementi osnovnog skupa.

Statistička obilježja, odnosno opće karakteristike elemenata statističkog skupa općenito se označavaju: sa

$$X_i, i = 1, 2, \dots, N. \quad (2.1)$$

Prema tome postoji N modaliteta, tj. pojava oblika promatranih obilježja.

Broj elemenata statističkog skupa koji pripadaju određenoj grupi, tj. jednom modalitetu ili pojavnom obliku promatranog obilježja, naziva se apsolutna frekvencija.

Oznaka za apsolutnu frekvenciju je

$$f_i, i = 1, 2, \dots, N. \quad (2.2)$$

Zbroj svih elemenata statističkog skupa naziva se opseg statističkog skupa. Zbog principa isključivosti i iscrpnosti taj broj

odgovara i **zbrotu svih apsolutnih frekvencija** kako je prikazano izrazom (2.3).¹

$$f_1 + f_2 + f_3 + \dots + f_N = \sum_{i=1}^N f_i \quad (2.3)$$

Može se definirati da je **skup uređenih parova modaliteta promatranog obilježja i njima pripadajućih apsolutnih frekvencija statistički niz**.

2.1.1. Nizovi kvalitativnih podataka

Kvalitativni statistički nizovi su **nominalni** (atributivni i zemljopisni) i **redosljedni**.

¹ U izrazu (2.3) dan je znak zbrajanja $\sum_{i=1}^N f_i$ koji se čita: zbroj apsolutnih frekvencija f_i , gdje i ide od 1 do N. Neke od karakteristika znaka zbrajanja su sljedeće:

1. $\sum_{i=1}^N X_i = X_1 + X_2 + \dots + X_N$, gdje je i = indeks zbrajanja;
2. $\sum_{i=1}^N (X_i \pm Y_i \pm Z_i \pm \dots) = \sum_{i=1}^N X_i \pm \sum_{i=1}^N Y_i \pm \sum_{i=1}^N Z_i \pm \dots$;
3. $\sum_{i=1}^N X_i Y_i = X_1 Y_1 + X_2 Y_2 + \dots + X_N Y_N \neq \sum_{i=1}^N X_i \cdot \sum_{i=1}^N Y_i$ (ovdje se za desni dio nejednakosti, primjenom izraza 1., može zaključiti da suma produkta nije jednaka produktu suma);
4. $\sum_{i=1}^N \frac{X_i}{Y_i} = \frac{X_1}{Y_1} + \frac{X_2}{Y_2} + \dots + \frac{X_N}{Y_N} \neq \frac{\sum_{i=1}^N X_i}{\sum_{i=1}^N Y_i}$ (ovdje se može zaključiti da je suma kvocijenta različita od kvocijenta suma);
5. $\sum_{i=1}^N aX_i = aX_1 + aX_2 + \dots + aX_N = a(X_1 + X_2 + \dots + X_N) = a \sum_{i=1}^N X_i$, gdje je "a" konstanta, jer ne ovisi o indeksu zbrajanja i;
6. $\sum_{i=1}^N (aX_i \pm bY_i \pm cZ_i \pm \dots) = a \sum_{i=1}^N X_i \pm b \sum_{i=1}^N Y_i \pm c \sum_{i=1}^N Z_i \pm \dots$, gdje su a, b, c, ... konstante.

Nominalni statistički nizovi formiraju se grupiranjem elemenata osnovnog skupa prema modalitetima odgovarajućeg nominalnog obilježja.

Modaliteti nominalnog obilježja izražavaju se pomoću atributa, kategorija i slovnih oznaka, a mogu se navoditi abecednim redom, prema veličini apsolutnih frekvencija, te dogovorno ili zakonski utvrđenim popisima. Na primjer, službena statistika koristi: SMTK ili Standardnu međunarodnu trgovinsku klasifikaciju i NKD ili Nacionalnu klasifikaciju djelatnosti.

Zemljopisni statistički nizovi formiraju se grupiranjem elemenata osnovnog skupa prema modalitetima odgovarajućega zemljopisnog ili prostornog obilježja.

Modaliteti prostornog obilježja najčešće odgovaraju teritorijalno-administrativnoj podjeli određenog geografskog prostora, kao na primjer: gradovi, županije, regije, države, kontinenti. Grupe toga obilježja mogu biti poredane po abecednom redu, po veličinama apsolutnih frekvencija ili po geografskom položaju. Ako se zbog veće preglednosti nekoliko modaliteta stavi zajedno u grupu "ostalo", ona obično dolazi na kraju statističkog niza.

Redoslijedni statistički nizovi formiraju se grupiranjem elemenata osnovnog skupa prema modalitetima odgovarajućeg redoslijednog obilježja.

Grupiranje podataka statističkog skupa prema redoslijednom obilježju vrši se slično kao kod nominalnih obilježja, ali je redoslijed grupa ovdje određen rangom koji pojedina grupa obilježja predstavlja. Rangiranje se može vršiti polazeći od najnižeg ranga prema najvišem ili obrnuto.

2.1.2. Numerički niz

Grupiranjem elemenata statističkog skupa prema numeričkom ili kvantitativnom obilježju nastaje numerički statistički niz.

Grupiranje i uređivanje elemenata numeričkog statističkog skupa ovisi o broju podataka i o njihovoj vrsti, odnosno o tome radi li se o neprekidnim (kontinuiranim) ili prekidnim (diskontinuiranim) numeričkim statističkim obilježjima.

Ako numeričko obilježje može poprimiti samo mali broj numeričkih vrijednosti, tada ta svaka vrijednost može biti posebna grupa u nizu. To je najčešće slučaj kod prekidnog ili diskontinuiranog numeričkog obilježja.

Uređen negrupiran prekidni numerički statistički niz ima karakteristiku da su njegove vrijednosti složene po veličini počevši, najčešće, od najmanje:

$$X_1, X_2, \dots, X_N; X_{i-1} < X_i, i = 2, 3, \dots, N. \quad (2.4)$$

Općenito je pojedini oblik (modalitet) obilježja: $X_i, i = 1, 2, \dots, k$, a pripadajuća apsolutna frekvencija: $f_i, i = 1, 2, \dots, k$.

Skup uređenih parova:

$$(X_i, f_i), i = 1, 2, 3, \dots, k \quad (2.5)$$

je distribucija frekvencija promatranog prekidnog numeričkog obilježja.

Također vrijedi da je **suma svih apsolutnih frekvencija jednaka opsegu statističkog odnosno osnovnog skupa:**

$$\sum_{i=1}^k f_i = N. \quad (2.6)$$

Za vrijednosti koje se pojavljuju manje puta formiraju se zajedničke grupe ili "razredi". U razredima se nalaze elementi skupa s vrijednostima obilježja koje se nalaze između donje i gornje granice razreda.

Na taj se način formiranjem razreda postiže veća preglednost, ali se gubi na preciznosti i potpunosti informacija. U slučaju **diskontinuiranog numeričkog obilježja donja granica sljedećeg ($i+1$) razreda razlikuje se od gornje granice prethodnog (i) razreda za jedinicu mjere promatranog obilježja.**

Te originalne granice razreda nazivaju se još i **nominalne granice**. Prije provođenja statističke analize potrebno je nominalne granice zamijeniti **preciznim ili pravim granicama**, što ovisi o načinu mjerenja promatranog svojstva. To se najčešće radi na način da se donja granica sljedećeg ($i+1$) nominalnog razreda umanjuje, a gornja granica prethodnog (i) nominalnog razreda uveća za polovicu razlike između tih granica. Ponekad se precizne granice formiraju jednostavno na način da se sve gornje granice uvećaju za 1 jedinicu obilježja i tako izjednače sa svojom sljedećom donjom granicom ili da se sve donje granice umanje za 1 jedinicu obilježja i tako izjednače sa svojom prethodnom gornjom granicom.

Kontinuirano ili neprekidno statističko obilježje može poprimiti vrijednosti iz nekog intervala, pa se **njegovi modaliteti formiraju kao razredi**. Skup uređenih parova razreda obilježja i njima pripadajućih frekvencija²

$$(DG_i \leq X_i < GG_i, f_i), i = 1, 2, 3, \dots, k \quad (2.7)$$

je distribucija frekvencija promatranog neprekidnog numeričkog obilježja. U slučaju kontinuiranog numeričkog obilježja donja granica sljedećeg ($i+1$) razreda jednaka je gornjoj granici prethodnog (i) razreda, pa ovdje nije potrebno računati precizne ili prave granice.

Ako nije definirana najmanja vrijednost obilježja za neki skup, **donja granica početnog razreda** pri grupiranju se može ostaviti **otvorena**. Isto tako ako nije definirana najveća vrijednost obilježja za neki skup, **gornja granica posljednjeg razreda** može se ostaviti **otvorena**. Međutim prije početka same statističke analize, te je otvorene granice potrebno zatvoriti takozvanim **procijenjenim granicama**, koje se stavljaju u zagrade, jer njima istraživač procjenjuje da nijedan element skupa nema vrijednost manju od prve donje ili veću od zadnje gornje procijenjene granice.

Veličina razreda (i) razlika je između gornje i donje granice razreda (kada je riječ o diskontinuiranom numeričkom obilježju radi se s preciznim ili pravim granicama).

² gdje je DG_i donja granica i -tog razreda, a GG_i gornja granica i -tog razreda.

$$i = GG_i - DG_i, \quad i = 1, 2, \dots, k. \quad (2.8.)$$

Sredina razreda ($sredX_i$) je prosjek (odnosno jednostavna aritmetička sredina) između donje i gornje (prave ili precizne) granice razreda:

$$sredX_i = \frac{DG_i + GG_i}{2}, \quad i = 1, 2, \dots, k. \quad (2.9)$$

Ako vrijednosti promatranog numeričkog niza nisu ravnomjerno raspoređene, veličine koje se izračunavaju pomoću sredine razreda sadrže tzv. **grešku grupiranja**.

2.2. Tabeliranje

Tabeliranje je postupak svrstavanja grupiranih prikupljenih statističkih podataka u tablice.

Statističke tablice kao jedan od oblika prikazivanja statističkih podataka prisutne su u literaturi svuda oko nas.

Tablica nastaje crtanjem okomitih i vodoravnih linija prema određenim pravilima. Svaka **statistička tablica mora imati: naslov, broj tablice (ako ih ima više), tekstualni dio, numerički ili brojčani dio i izvor podataka te po potrebi napomenu.**

Tablica 2.1.

| Naslov tablice | | | | | |
|---------------------------------|--------|------------------|---|-------|----------------------------------|
| | | Zaglavlje | | | Zbirni stupac: $\sum_{j=1}^n$ |
| | | Oznake stupca | | | |
| 0 | | 1 | 2 | ... | n |
| Pred- stu- pac | Oznake | redak | | | |
| | retka | stu- pac | | polje | |
| | | | | polje | |
| Zbirni redak: $\sum_{i=1}^m$ | | | | | |

Izvor podataka:

Tablica 2.1. prikazuje opći oblik statističke tablice.

Naslov tablice mora biti jasan i kratak, a istovremeno u sebi mora sadržavati pojmovnu, prostornu i vremensku definiciju statističkog skupa da bi onaj tko je čita mogao precizno odrediti njezine elemente.

Tekstualni dio statističke tablice sastoji se od dva dijela: zaglavlja i predstupca. U zaglavlju ili tumaču stupaca opisuje se i objašnjava sadržaj stupaca. U predstupcu ili tumaču redaka opisuje se i objašnjava sadržaj redaka. To su najčešće oblici statističkog obilježja po kojemu je promatran statistički niz. Uz tekstualni dio svaki stupac može biti označen i odgovarajućim brojem. Kako se vidi u tablici 2.1, predstupac se često označava nulom, a numeriranje dalje ide po redu: 1, 2, ..., n.

Brojčani ili numerički dio tablice sastoji se od polja u koja se unose frekvencije, odnosno rezultati grupiranja statističkih podataka. Zbirni ili marginalni stupac sadrži zbrojeve pojedinih redaka. Njegova suma $\sum_{j=1}^n$ zbraja elemente svakog retka po j - stupcima. Zbirni ili marginalni redak sadrži zbrojeve pojedinih stupaca. Njegova suma $\sum_{i=1}^m$ zbraja elemente svakog stupca po i- redcima.

Izvor podataka se navodi ispod tablice. On omogućuje provjeru ispravnosti prikupljenih podataka u tablici, kao i eventualnu dopunu podataka, ako to zahtijeva statističko istraživanje (posebno ako su podaci u tablici preuzeti iz sekundarnih izvora).

Zbog veće preglednosti tablice, često se zaglavlje, predstupac, zbirni redak i zbirni stupac odvajaju debljim crtama od brojčanog dijela kao što se može vidjeti u tablici 2.1.

Dakle, **statistička tablica mora poštivati 3 načela: preglednost, potpunost i jasnoću.**

Statističke tablice mogu se **podijeliti: na**

- 1) **opće ili izvještajne statističke tablice** i na
- 2) **analitičke ili sumarne statističke tablice.**

Opće statističke tablice prikazuju jako velik broj statističkih podataka o nekome promatranom statističkom skupu.

Analitičke statističke tablice se konstruiraju za neku posebnu analizu i one su u pravilu preglednije.

U ovisnosti o tome prikazuju li se prikupljeni statistički podaci u tablici prema jednom ili više statističkih obilježja i prikazuje li se jedan ili više statističkih nizova **razlikuju se**:

- 1) **jednostavne statističke tablice** i
- 2) **složene statističke tablice**, koje se mogu podijeliti: na
 - **skupne statističke tablice** i na
 - **kombinirane statističke tablice**.

Jednostavne statističke tablice prikazuju jedan statistički skup grupiran u jedan statistički niz, tj. grupiran prema jednom obilježju.

Skupne statističke tablice prikazuju dva ili više statistička skupa grupirana u dva ili više statistička niza prema jednom obilježju.

Tablica 2.2.

Naslov tablice: Dvostruka (složena) statistička tablica

| | | Zaglavlje | | | | Zbirni stupac: $\sum_{j=1}^n$ |
|---------------------------------|--------|-----------------------|-----------------------|-----|-----------------------|------------------------------------|
| | | Oznake stupca | | | | |
| 0 | | 1 | 2 | ... | n | |
| Pred- stu- pac | Oznake | x_{11} | x_{12} | ... | x_{1n} | $\sum_{j=1}^n x_{1j}$ |
| | retka | x_{21} | x_{22} | ... | x_{2n} | $\sum_{j=1}^n x_{2j}$ |
| | | ... | ... | | | |
| | | x_{m1} | x_{m2} | ... | x_{mn} | $\sum_{j=1}^n x_{mj}$ |
| Zbirni redak: $\sum_{i=1}^m$ | | $\sum_{i=1}^m x_{i1}$ | $\sum_{i=1}^m x_{i2}$ | ... | $\sum_{i=1}^m x_{in}$ | $\sum_{i=1}^m \sum_{j=1}^n x_{ij}$ |

Izvor podataka:

Kombinirane statističke tablice (tablice kontigence) prikazuju jedan statistički skup grupiran u dva ili više statističkih nizova prema dva ili više obilježja. Ako kod kombinirane tablice s dva obilježja jedno obilježje ulazi u tablicu iz predstupca, a drugo iz zaglavlja, kaže se da je to tablica s dva ulaza. U skladu s tim postoje kombinirane tablice s više ulaza, ali pri njihovom kreiranju gubi se jasnoća i preglednost.

Tablica 2.2 prikazuje jednu opću dvostruku statističku tablicu s odgovarajućim sumama u zbirnom retku i zbirnom stupcu. Jednadžba (2.10) pokazuje da je zbroj zbirnog retka jednak zbroju zbirnog stupca, odnosno ako kod dvostruke sume one zamijene svoja mjesta, ništa se neće promijeniti:

$$\sum_{i=1}^m \sum_{j=1}^n x_{ij} = \sum_{j=1}^n \sum_{i=1}^m x_{ij} . \quad (2.10)$$

2.3. Relativni brojevi strukture

Relativni brojevi su "neimenovani", stoga se pomoću njih mogu uspoređivati i analizirati pojave koje imaju različitu jedinicu mjere ili različit broj elemenata. Na taj način dobije se relativna važnost dijela ili cjeline statističkog niza.

Relativni brojevi nastaju dijeljenjem dviju veličina. Veličina s kojom se dijeli zove se osnova relativnog broja i po njoj se relativni brojevi međusobno razlikuju.

Relativni brojevi strukture pokazuju odnos dijela prema cjelini, i njima se olakšava analiza rasporeda podataka prema modalitetima obilježja u jednom statističkom nizu, odnosno njihova struktura. Najčešće se izražavaju u postocima, a mogu i u promilima.

Ako relativni brojevi strukture (P_i) pokazuju odnos apsolutnih frekvencija prema opsegu statističkog skupa (tj. ukupnom broju elemenata statističkog skupa), zovu se relativne frekvencije (fr_i).

$$P_i = \frac{\text{dio}}{\text{cjelina}}; \quad fr_i = \frac{f_i}{\sum_{i=1}^n f_i} \quad (2.11)$$

$$P_i = \frac{\text{dio}}{\text{cjelina}} \cdot 100 \text{ (u \%)}; \quad fr_i = \frac{f_i}{\sum_{i=1}^n f_i} \cdot 100 \text{ (u \%)} \quad (2.12)$$

$$P_i = \frac{\text{dio}}{\text{cjelina}} \cdot 1000 \text{ (u \text{‰})}; \quad fr_i = \frac{f_i}{\sum_{i=1}^n f_i} \cdot 1000 \text{ (u \text{‰})} \quad (2.13)$$

Zbroj svih relativnih brojeva u jednom statističkom nizu je 1 ili 100, ili 1000, u ovisnosti o tome kako je relativni broj izražen.

Relativni brojevi strukture se **grafički** mogu prikazivati pomoću strukturnih stupaca, strukturnih krugova, polukrugova ili nekim drugim geometrijskim likom. Pri tom se za usporedbu dvaju ili više statističkih nizova konstruiraju geometrijski likovi jednakih površina, jer je zbroj relativnih frekvencija uvijek isti.

Analitički primjer 2.1.

Statistički nizovi "prosječan broj nezaposlenih osoba u RH, 1995. i 2002. godine" **prema redoslijednom obilježju** "stručna sprema", prikazani su u tablici 2.2.

- a) Zadatak je izračunati strukturu nezaposlenih osoba prema stručnoj spremi za 1995. i 2003. godinu.
- b) Zadatak je strukturu nezaposlenih osoba prema stručnoj spremi za 1995. i 2003. godinu prikazati na grafikonu pomoću strukturnih stupaca uz sve potrebne oznake.

Rješenje 2.1.

- a) U tablici 2.3. u posljednja 2 stupca prikazana je struktura nezaposlenih osoba prema stručnoj spremi za 1995. i 2003. godinu.

$$\text{Na primjer: } fr_1 = \frac{f_1}{\sum_{i=1}^6 f_i} \cdot 100 = \frac{51367}{249070} \cdot 100 = 20,62\%$$

što znači da je 1995. godine udio niskokvalificiranih radnika (NKV) u ukupnom broju nezaposlenih bio 20,62%.

Tablica 2.3.

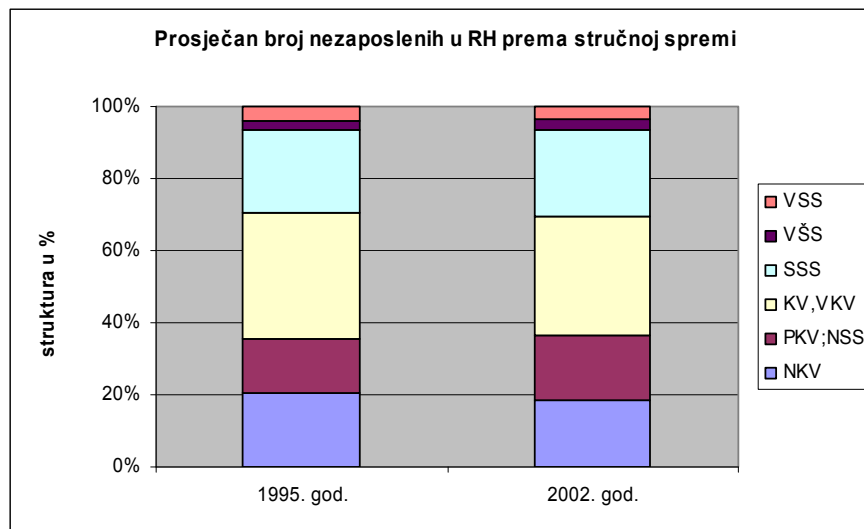
Prosječan broj nezaposlenih osoba u RH prema stručnoj spremi, 1995. i 2003. godine

| Stručna sprema | Prosječan broj nezaposlenih osoba | | Struktura u % | |
|----------------|-----------------------------------|-----------------|------------------|------------------|
| | 1995. god. | 2003. god. | 1995. god. | 2002. god. |
| | f1 _i | f2 _i | fr1 _i | fr2 _i |
| NKV | 51367 | 58939 | 20,62352 | 18,4945 |
| PKV;NSS | 36453 | 56957 | 14,63564 | 17,87256 |
| KV,VKV | 87172 | 105728 | 34,999 | 33,17644 |
| SSS | 57276 | 76087 | 22,99594 | 23,87537 |
| VŠS | 7346 | 9891 | 2,949372 | 3,103701 |
| VSS | 9456 | 11082 | 3,796523 | 3,477426 |
| Ukupno: | 249070 | 318684 | 100 | 100 |

Izvor: Statistički ljetopis Republike Hrvatske, 2005. godina, str. 177.

b) Na slici 2.1. prikazana je struktura nezaposlenih osoba prema stručnoj spremi za 1995. i 2003. godinu.

Slika 2.1.



Izvor: Statistički ljetopis Republike Hrvatske, 2005. godina, str. 177.

2.4. Grafičko prikazivanje statističkih nizova

Uz statističke tablice pomoćno sredstvo u analizi statističkih nizova su grafički prikazi.

Grafikonima se na jednostavan i pregledan način uz pomoć različitih geometrijskih likova prezentiraju osnovne karakteristike statističkih nizova.

Grafički prikazi statističkih podataka su pregledniji i razumljiviji u odnosu na njihovo prikazivanje statističkom tablicom. Grafikoni omogućuju jednostavnije uočavanje glavnih karakteristika promatranih pojava, ali vrlo često ta preglednost ide na štetu preciznosti statističkih informacija. Stoga je poželjno uz grafički prikaz prezentirati i tablicu s originalnim vrijednostima statističkog niza. Suvremeni statistički programski paketi, a između njih i SPSS, naravno u skladu sa statističkom teorijom, imaju mnoštvo mogućnosti kreiranja grafičkih prikaza. Pomoću njih se mogu odabrati različite boje, oblici i linije na grafikonu, što omogućuje još zorniji prikaz promatrane pojave.

Oznake na grafikonu moraju biti takve da onaj tko čita grafikon može jasno raspoznati koji su elementi i koja je pojava prikazana. Stoga i grafikon mora imati **naslov, jedinice mjere promatranog obilježja, oznake modaliteta obilježja, izvor podataka i po potrebi kazalo ili tumač oznaka.**

Postoje tri skupine grafičkih prikaza:

- 1) površinski grafikoni
- 2) linijski grafikoni i
- 3) kartogrami.

Analitički primjer 2.2.

Statistički skup "Zaposleni u pravnim osobama u RH, stanje 31. ožujka 2004. god.", **prema redoslijednom obilježju** "stupanj stručnog obrazovanja" prikazan je u tablici 2.3.

a) Zadatak je zaposlene u pravnim osobama u RH prema stupnju stručnog obrazovanja za 2004. godinu prikazati na grafikonu pomoću jednostavnih stupaca uz sve potrebne oznake.

Tablica 2.4.

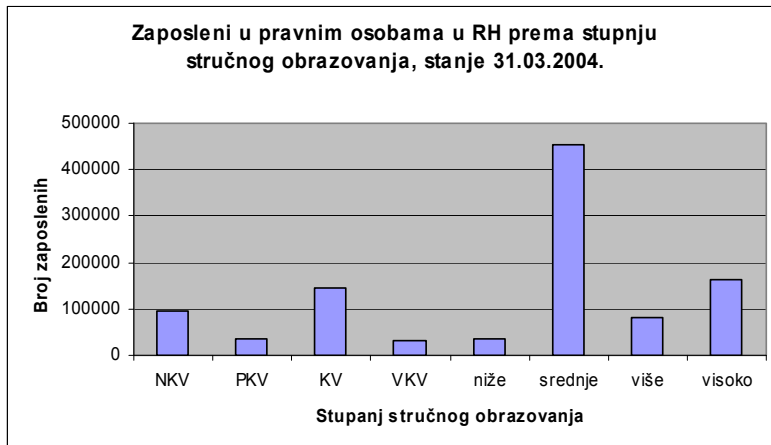
Zaposleni u pravnim osobama u RH prema stupnju stručnog obrazovanja, stanje 31. ožujka 2004. god.

| Stupanj stručnog obrazovanja | Broj zaposlenih |
|------------------------------|-----------------|
| NKV | 97516 |
| PKV | 35948 |
| KV | 146161 |
| VKV | 30626 |
| niže | 36510 |
| srednje | 453710 |
| više | 82283 |
| visoko | 163417 |
| Ukupno: | 1046171 |

Izvor: Statistički ljetopis Republike Hrvatske, 2005. godina, str. 136.-137.

Rješenje 2.2.

a) Na slici 2.2 prikazani su zaposleni u pravnim osobama u RH prema stupnju stručnog obrazovanja za 2004. godinu.

Slika 2.2.

Izvor: Statistički ljetopis Republike Hrvatske, 2005. godina, str. 136.-137.

Na slici 2.2 jasno se vidi da je u Republici Hrvatskoj u pravnim osobama, na dan 31. 03. 2004., najviše bilo zaposlenih sa srednjim

stupnjem obrazovanja, dok je onih u kategoriji VKV, odnosno visokokvalificiranih radnika bilo najmanje.

Analitički primjer 2.3.

Statistički skup, "Stanovništvo Republike Hrvatske, popis 2001. godine" prema obilježjima "spol i aktivnost" **podijeljen je na grupe** i zadan je u **tablici 2.5.**

a) Zadatak je napraviti **strukturne krugove** za stanovništvo muškog i ženskog spola prema obilježju "aktivnost".

Tablica 2.5.

Muškarci i žene u RH prema aktivnosti, popis 2001. godine

| Aktivnost | Stanovništvo (u 000) | | Stupnjevi struktur. kruga | |
|----------------------------|----------------------|--------|---------------------------|----------|
| | Muškarci | Žene | Muškarci | Žene |
| | $f1_i$ | $f2_i$ | S^01_i | S^02_i |
| Aktivno stanovništvo | 1005,4 | 819,8 | 178,3854 | 132,9885 |
| Osobe ss osobnim prihodima | 490,6 | 639 | 87,04584 | 103,659 |
| Udržavano stanovništvo | 533 | 760,4 | 94,56875 | 123,3526 |
| Ukupno | 2029 | 2219,2 | 360 | 360 |

Izvor: Statistički ljetopis Republike Hrvatske, 2005. godina, str. 100.

Rješenje 2.3.

a) U tablici 2.5 uz originalne apsolutne frekvencije dani su i stupnjevi strukturnoga kruga. Stupnjevi strukturnoga kruga računaju se na sljedeći način:

$$S^0_i = \frac{f_i}{\sum_{i=1}^N f_i} \cdot 360^0. \quad (2.14)$$

Na slici 2.3 pomoću strukturnih krugova prikazani su muškarci i žene u Republici Hrvatskoj prema aktivnosti (popis 2001.).

U stupcu tablice 2.5. za muški spol:

$$S^0 1_1 = \frac{f_{1_1}}{\sum_{i=1}^3 f_{1_i}} \cdot 360^0 = \frac{1005,4}{2029} \cdot 360^0 = 178,385^0,$$

što odgovara određenom dijelu 1. strukturnoga kruga na lici 2.3.

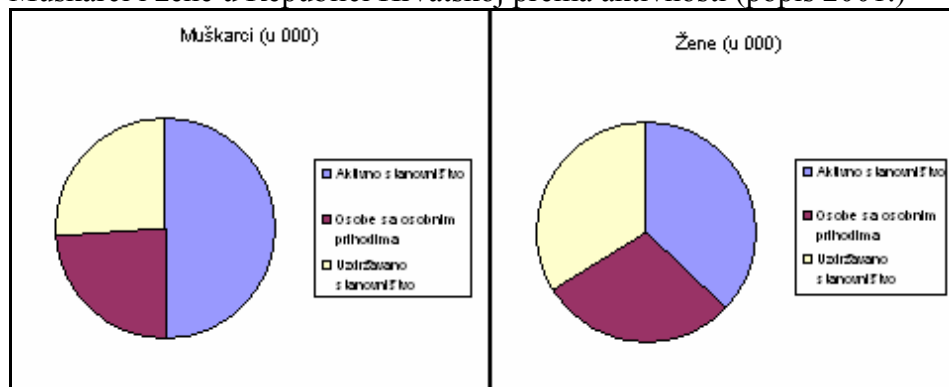
U stupcu tablice 2.5. za ženski spol:

$$S^0 2_2 = \frac{f_{2_2}}{\sum_{i=1}^3 f_{2_i}} \cdot 360^0 = \frac{639}{2219,2} \cdot 360^0 = 103,659^0,$$

što odgovara određenom dijelu 2. strukturnoga kruga na slici 2.3.

Slika 2.3.

Muškarci i žene u Republici Hrvatskoj prema aktivnosti (popis 2001.)



Izvor: Statistički ljetopis Republike Hrvatske, 2005. godina, str. 100.

Iz slike 2.3. jasno se vidi da je kod muškaraca bilo najviše aktivnog stanovništva, dok je kod žena aktivno stanovništvo samo malo veće od uzdržavanog stanovništva.

Analitički primjer 2.4.

Zadan je **statistički skup**, "Zasijane površine u oranicama" prema **prostornom obilježju** "Županije u RH".

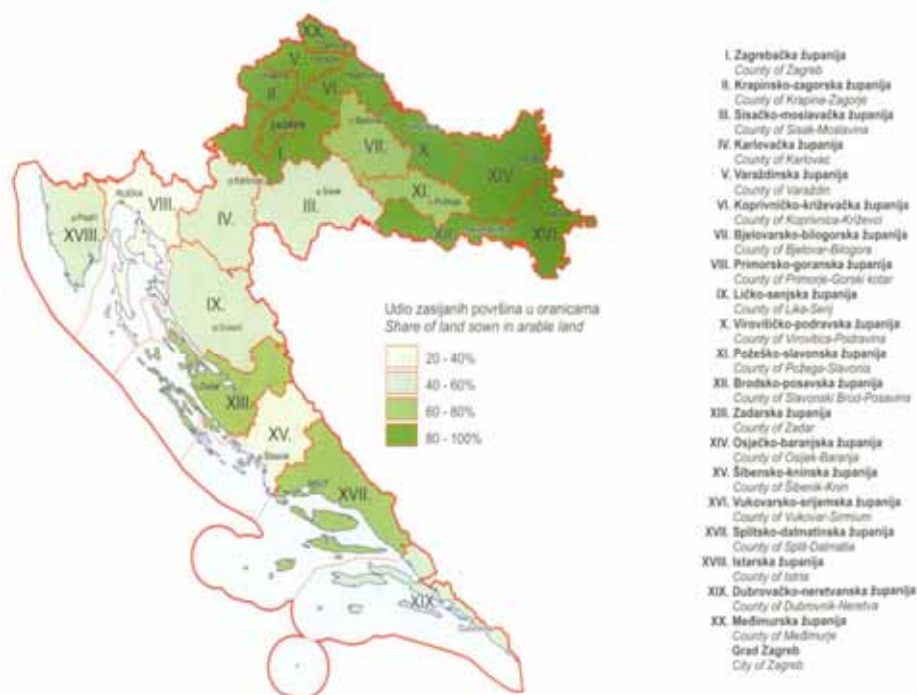
- a) Zadatak je zadani skup prikazati pomoću **statističke karte**.

Rješenje 2.4.

a) Statistički skup, "Zasijane površine u oranicama" prema prostornom obilježju "Županije u RH" prikazan je na slici 2.4.

Slika 2.4.

Udio zasijanih površina u oranicama po županijama u RH, 2002. god.



Izvor: Statistički ljetopis Republike Hrvatske, 2003. godina, str. 673.

Na slici 2.4 vidi se udio zasijanih površina u oranicama u Republici Hrvatskoj prikazan po županijama. Na zemljovidu su sjenčanjem prikazani intenziteti vrijednosti pojave po modalitetima obilježja, tj. po županijama. Tako su one županije s manjim udjelom zasijane površine u oranicama slabije osjenčane, dok su one s većim udjelom zasijane površine jače osjenčane.



Analitički primjer 2.5.

Zadani su podaci o spolu i džeparcu uzorka studentske populacije na jednom sveučilištu.

- a) Zadatak je formirati i komentirati apsolutne frekvencije varijabli:
- spol,
 - "grupirani" džeparac.



Rješenje 2.5.

- a) U programskom paketu **SPSS** apsolutne frekvencije se mogu formirati na 2 načina.

I. način:

Potrebno je na glavnom izborniku odabrati **Analyze**, a na njegovu padajućem izborniku odabrati **Tables**. Dalje se bira **Tables of Frequencies**. Najprije se varijabla v1 - Spol prebaci u polje **Frequencies for:**. Klikom na **OK** u **Outputu** programa SPSS dobije se tablica frekvencija uzorka prema obilježju, odnosno varijabli spol. Rezultat je prikazan u tablici 2.6.

Tablica 2.6.

Ispitanici prema spolu

| | Spol Count |
|--------|---------------|
| Musko | 94 |
| Zensko | 144 |

Izvor: Simulirani podaci.

Prva frekvencija u tablici 2.6. znači da je među ispitanicima bilo 94 studenta. Druga frekvencija u tablici 2.6. znači da je među ispitanicima bilo 144 studentice.

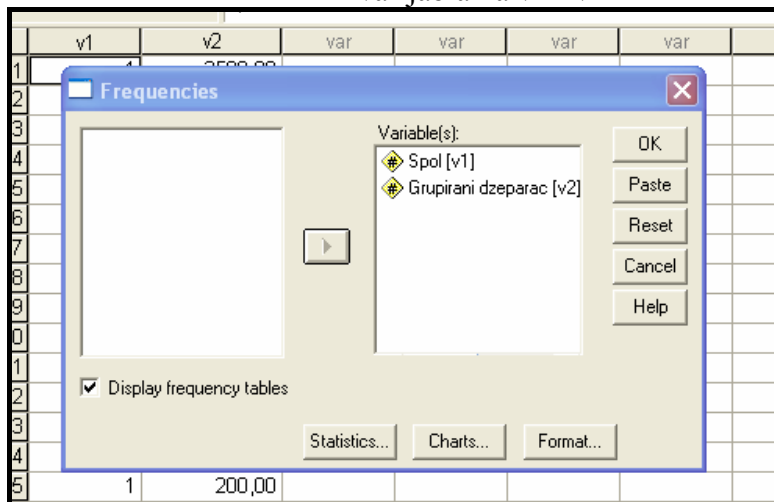
II. način:

Potrebno je na glavnom izborniku odabrati **Analyze**, a na njegovu padajućem izborniku odabrati **Descriptive Statistics**. Dalje se bira **Frequencies**. U ovom slučaju se mogu obje varijable v1 - Spol i v2 -

Grupirani džeparac prebaciti u polje **Variable(s)**. Konačni izgled prozora **Frequencies** s odabranim varijablama prikazan je na slici 2.5.

Slika 2.5.

Prozor "Frequencies" iz izbornika "Descriptive Statistics" s odabranim varijablama v1 i v2



Izvor: Simulirani podaci.

Klikom na **OK** u **Outputu** programa **SPSS** dobije se tablica frekvencija uzorka prema odabranim obilježjima, odnosno varijablama. Rezultat je prikazan u tablicama 2.7 i 2.8.

Tablica 2.7.

Ispitanici prema obilježju: **Spol**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------|--------|-----------|---------|---------------|--------------------|
| Valid | Musko | 94 | 39,3 | 39,5 | 39,5 |
| | Zensko | 144 | 60,3 | 60,5 | 100,0 |
| | Total | 238 | 99,6 | 100,0 | |
| Missing | System | 1 | ,4 | | |
| Total | | 239 | 100,0 | | |

Izvor: Simulirani podaci.

Prva frekvencija u stupcu **Frequency** u tablici 2.7. znači da je među ispitanicima bilo 94 studenta. Druga frekvencija istog stupca znači da je među ispitanicima bilo 144 studentice. Na postavljeno pitanje

odgovorilo 238 ispitanika, dok frekvencija 1 iz retka **Missing System** znači da 1 ispitanik nije dao odgovor na pitanje o njegovom spolu. Stupac **Percent** prikazuje relativne frekvencije u postotcima za valjane podatke uključujući i one koji nedostaju. Stupac **Valid Percent** prikazuje originalne valjane frekvencije u postotcima. Prvi podatak iz tog stupca pokazuje da je među onima koji su odgovorili na pitanje o spolu bilo 39,5% muškog spola, a 60,5 ženskog spola.

Tablica 2.8.

Ispitanici prema obilježju: **Grupirani džeparac**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------|---------|-----------|---------|---------------|--------------------|
| Valid | 200,00 | 73 | 30,5 | 42,2 | 42,2 |
| | 400,00 | 52 | 21,8 | 30,1 | 72,3 |
| | 600,00 | 30 | 12,6 | 17,3 | 89,6 |
| | 800,00 | 6 | 2,5 | 3,5 | 93,1 |
| | 1000,00 | 7 | 2,9 | 4,0 | 97,1 |
| | 3500,00 | 5 | 2,1 | 2,9 | 100,0 |
| | Total | 173 | 72,4 | 100,0 | |
| Missing | System | 66 | 27,6 | | |
| Total | | 239 | 100,0 | | |

Izvor: Simulirani podaci.

Prva frekvencija u stupcu **Frequency** u tablici 2.8 znači da je među ispitanicima bilo 73 studenta s džeparcem od 200 kn ili manje. Druga frekvencija istog stupca znači da je među ispitanicima bilo 52 studenta s džeparcem između 201 kn i 400 kn itd. Na postavljeno pitanje je odgovorilo 173 ispitanika, dok frekvencija 66 iz retka **Missing System** znači da 66 ispitanika nije dalo odgovor na pitanje o njihovom džeparcu. Stupac **Percent** prikazuje relativne frekvencije u postotcima za valjane podatke uključujući i one koji nedostaju. Stupac **Valid Percent** prikazuje originalne valjane frekvencije u postotcima. Prvi podatak iz tog stupca pokazuje da je među onima koji su odgovorili na pitanje o džeparcu bilo 42,2% studenata s džeparcem od 200 kn ili manje itd.

Ovaj II način prikaza frekvencija u programu **SPSS** je detaljniji, jer osim apsolutnih prikazuje i relativne frekvencije.

**Analitički primjer 2.6.**

Zadatak je za sve ispitanike od varijable "Prosječna ocjena u srednjoj školi" (v1) formirati novu varijablu "Grupirana ocjena u srednjoj školi" (v2) po principu koji je prikazan u tablici 2.9.

Tablica 2.9.

Princip formiranja nove varijable grupirana ocjena u srednjoj školi (v2)

| Grupe | Prosje. ocjena u sred. školi | Nova vrijednost |
|-------|---------------------------------|--------------------|
| 1 | 2,00-2,49 | 2 |
| 2 | 2,50-3,49 | 3 |
| 3 | 3,50-4,49 | 4 |
| 4 | 4,50-5,00 | 5 |

Izvor: Simulirani podaci.

**Rješenje 2.6:**

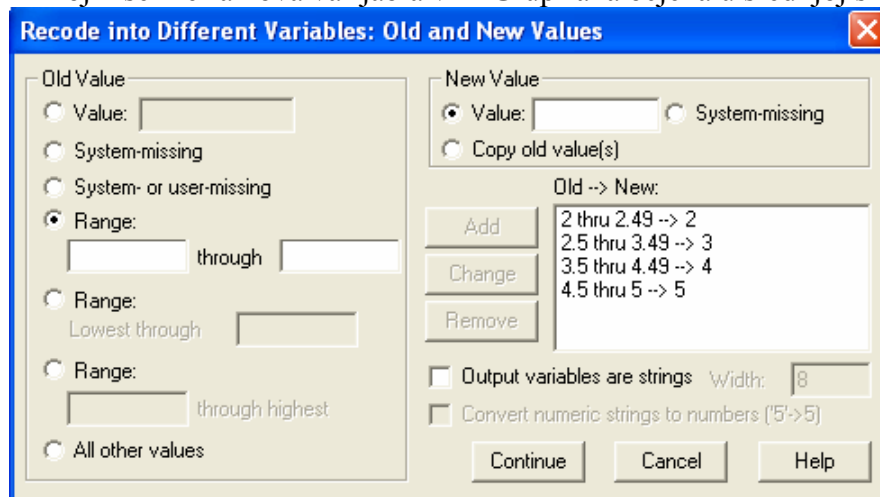
Da bi se pomoću varijable v1 - "Prosječna ocjena u srednjoj školi" formirala nova varijabla v2 - "Grupirana ocjena u srednjoj školi" po principu koji je prikazan u tablici 2.8, potrebno je na glavnom izborniku odabrati **Transform**, a u njegovu padajućem izborniku odabrati **Recode**. U njegovu izborniku bira se **Into Different Variables** (jer bi se odabirom **Into Same Variables** izgubila postojeća varijabla). U polje **Numeric Variable** bira se v1 - "Prosječna ocjena u srednjoj školi", a u **Output Variable, Name** je v2, a **Label** "Grupirana ocjena u srednjoj školi".

Odabirom ikone **Old and New Values** otvara se novi prozor **Recode into Different Variables**. Aktiviranjem opcije **Range** upisuju se odgovarajuće grupe ocjena, kako je zadano u tablici 2.8. Nakon upisivanja svakog pojedinog **Range**, upisuje se odgovarajuća **New Value**, koja se svaki put kao nova grupa dodaje klikom na ikonu **Add**. Nakon što su unesene sve grupe ikonom **Continue**, zatim **Change** i **OK**, u novom stupcu **Data View**, kreirat će se nova varijabla "Grupirana ocjena u srednjoj školi", odnosno v2. (Princip je sličan kao u analitičkom primjeru 1.9.)

Na slici 2.6 prikazan je prozor **Recode into Different Variables: Old and New Values**.

Slika 2.6.

Prozor "Recode into Different Variables: Old and New Values", pomoću kojih se kreira nova varijabla v2 - Grupirana ocjena u srednjoj školi



Izvor: Simulirani podaci.



Analitički primjer 2.7.

Zadatak je komentirati apsolutne i relativne frekvencije varijable: Grupirana ocjena u srednjoj školi (v2).

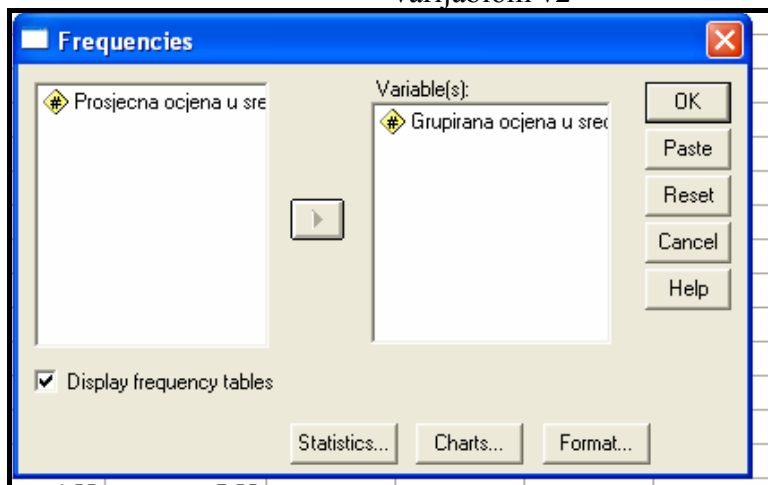


Rješenje 2.7:

Potrebno je na glavnom izborniku odabrati **Analyze**, a na njegovu padajućem izborniku odabrati **Descriptive Statistics**. Dalje se bira **Frequencies**, gdje se u dio prozora **Variable(s)** klikom na strelicu prebaci varijabla v2 - Grupirana ocjena u srednjoj školi. Konačni izgled prozora **Frequencies** s odabranom varijablom prikazan je na slici 2.7.

Slika 2.7.

Prozor "Frequencies" iz izbornika "Descriptive Statistics" s odabranom varijablom v2



Izvor: Simulirani podaci.

Klikom na **OK** u **Outputu** programa **SPSS** dobije se tablica frekvencija uzorka prema odabranom obilježju, odnosno varijabli. Rezultat je prikazan u tablici 2.10.

Tablica 2.10.

Ispitanici prema: **Grupirana ocjena u srednjoj školi**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------|--------|-----------|---------|---------------|--------------------|
| Valid | 2,00 | 1 | ,4 | ,4 | ,4 |
| | 3,00 | 34 | 14,2 | 14,3 | 14,7 |
| | 4,00 | 113 | 47,3 | 47,5 | 62,2 |
| | 5,00 | 90 | 37,7 | 37,8 | 100,0 |
| | Total | 238 | 99,6 | 100,0 | |
| Missing | System | 1 | ,4 | | |
| Total | | 239 | 100,0 | | |

Izvor: Simulirani podaci.

Prva frekvencija u stupcu **Frequency** u tablici 2.10 znači da je među ispitanicima bio 1 student s ocjenom 2 u srednjoj školi. Druga frekvencija istog stupca znači da je među ispitanicima bilo 34 studenta s

ocjenom 3, ... itd. Na postavljeno pitanje je odgovorilo 238 ispitanika, dok frekvencija 1 iz retka **Missing System** znači da 1 ispitanik nije dao odgovor na pitanje o ocjeni u srednjoj školi. Stupac **Percent** prikazuje relativne frekvencije u postotcima za valjane podatke uključujući i one koji nedostaju. Stupac **Valid Percent** prikazuje originalne valjane frekvencije u postotcima. Treći podatak iz tog stupca pokazuje da je 47,5% studenata imalo prosječnu ocjenu u srednjoj školi 4.



Analitički primjer 2.8.

Zadatak je između svih ispitanika odabrati samo one koji su ženskog spola.

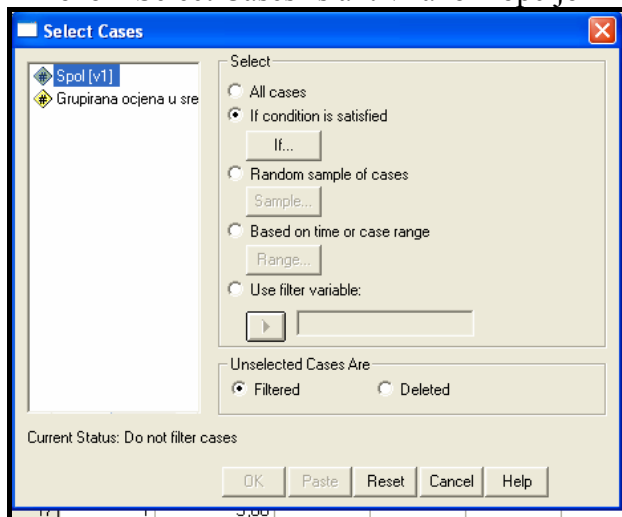


Rješenje 2.8:

U glavnom izborniku bira se **Data**, a na njegovu padajućem izborniku **Select Cases**, gdje je potrebno aktivirati opciju **If condition is satisfied** kako je prikazano na slici 2.8.

Slika 2.8.

Prozor "Select Cases" s aktiviranom opcijom "If condition is satisfied"

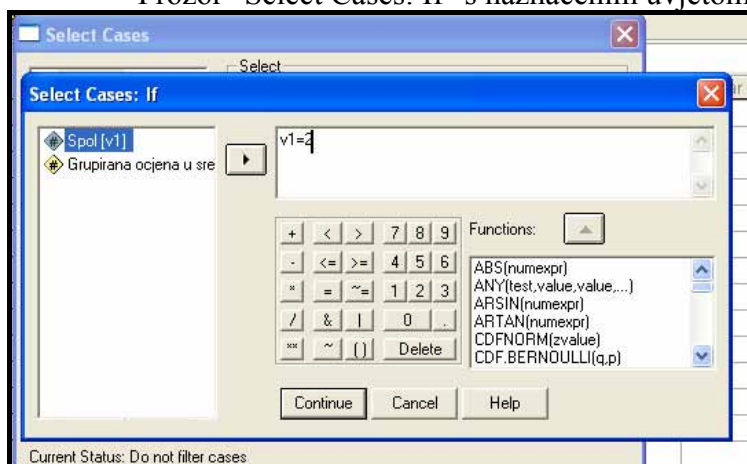


Izvor: Simulirani podaci.

Klikom na ikonu **If...** otvara se novi prozor **Select cases: If**, kako je prikazano na slici 2.9.

Slika 2.9.

Prozor "Select Cases: If" s naznačenim uvjetom $v1=2$



Izvor: Simulirani podaci.

Strelicom se varijabla $v1$ prebaci u odgovarajuće polje i izjednači s 2, jer je uvjet da se biraju samo oni ispitanici ženskog spola, a ženski spol je pri kodiranju označen s 2. Klikom na **Continue** i **OK**, u dodatnom stupcu **Data View** javlja se nova varijabla **filter_\$**, gdje su s 1 označena polja ispitanika ženskog spola, a s 0 polja ispitanika muškog spola. Naravno, **dok je filter aktivan program SPSS u svim svojim analizama uvažava samo polja označena u filteru s 1**, odnosno ispitanike ženskog spola. Stupac **filter_\$** može se vidjeti na slici 2.10.

Slika 2.10.

"Data View" sa stupcem "filter_\$"

| | v1 | v2 | filter_\$ | var |
|---|----|------|-----------|-----|
| 1 | 1 | 4,00 | 0 | |
| 2 | 2 | 5,00 | 1 | |
| 3 | 1 | 4,00 | 0 | |
| 4 | 1 | 4,00 | 0 | |
| 5 | 2 | 4,00 | 1 | |

Izvor: Simulirani podaci.



Analitički primjer 2.9.

Zadatak je komentirati apsolutne i relativne frekvencije varijable: Grupirana ocjena u srednjoj školi (v2) samo za ženski spol.



Rješenje 2.9:

Nakon što je izvršena selekcija ispitanika ženskog spola (kako je objašnjeno u 2.8), potrebno je na glavnom izborniku odabrati **Analyze**, a na njegovu padajućem izborniku **Descriptive Statistics**. Dalje se bira **Frequencies**, gdje se u dio prozora **Variable(s)** klikom na strelicu prebaci varijabla v2 - Grupirana ocjena u srednjoj školi. Klikom na **OK** u **Outputu** programa **SPSS** dobije se tablica frekvencija uzorka prema odabranom obilježju, odnosno varijabli. Rezultat je prikazan u tablici 2.11 gdje su dani rezultati samo za ženski spol.

Tablica 2.11.

Ispitanici prema: **Grupirana ocjena u srednjoj školi (ženski spol)**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|-------|-----------|---------|---------------|--------------------|
| Valid | 3,00 | 20 | 13,9 | 13,9 | 13,9 |
| | 4,00 | 59 | 41,0 | 41,0 | 54,9 |
| | 5,00 | 65 | 45,1 | 45,1 | 100,0 |
| | Total | 144 | 100,0 | 100,0 | |

Izvor: Simulirani podaci.

Prva frekvencija u stupcu **Frequency** u tablici 2.11 znači da je među ispitanicima ženskog spola bilo 20 njih s ocjenom 3 u srednjoj školi. Druga frekvencija istog stupca znači da je među ispitanicima bilo 59 studentica s ocjenom 4, a njih 65 s ocjenom 5. U analizi je ukupno bilo 144 ispitanika ženskog spola. Stupac **Percent** prikazuje relativne frekvencije u postotcima za valjane podatke. Kako se vidi iz tablice, njegove vrijednosti odgovaraju onima iz stupca **Valid Percent**. Treća proporcija ili relativna frekvencija pokazuje da su 45,1% studentica u odnosu na sve 144 studentice imala prosječnu ocjenu u srednjoj školi 5.



Analitički primjer 2.10.

Zadatak je iz postojeće baze podataka svih ispitanika odabrati slučajni uzorak od 60% ispitanika.

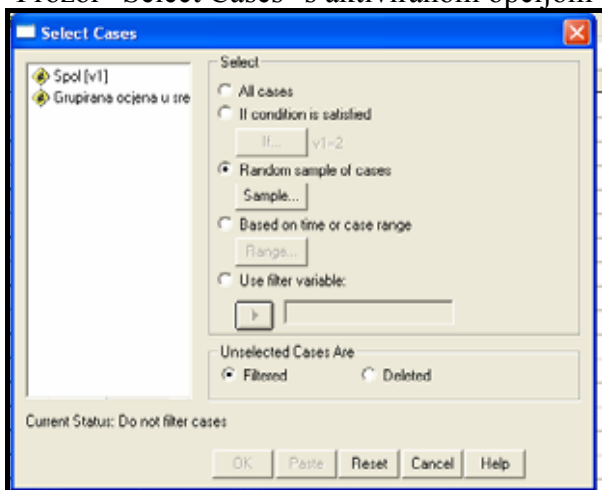


Rješenje 2.10:

U glavnom izborniku bira se **Data**, a na njegovu padajućem izborniku **Select Cases**, gdje je potrebno aktivirati opciju **Random sample of cases** kako je prikazano na slici 2.11.

Slika 2.11.

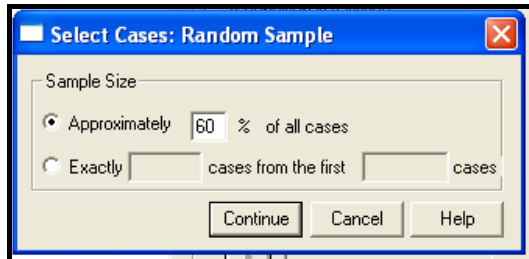
Prozor "Select Cases" s aktiviranom opcijom "Random sample of cases"



Izvor: Simulirani podaci.

Klikom na ikonu **Sample...** otvara se novi prozor **Select Cases: Random Sample**, kako je prikazano na slici 2.12. U tom prozoru potrebno je aktivirati **Approximately** i naznačiti **60% of all cases**. Klikom na **Continue** i **OK**, u dodatnom stupcu **Data View** javlja se nova varijabla **filter_\$**, gdje su s 1 označena polja slučajno odabranih 60% ispitanika, a sa 0 polja 40% ispitanika koji neće biti uključeni u daljnju analizu. Naravno, **dok je filter aktivan program SPSS u svim svojim analizama uvažava samo polja označena u filteru s 1.**

Slika 2.12.

Prozor "Select Cases: If" s naznačenim uvjetom $v1=2$ 

Izvor: Simulirani podaci.

**Analitički primjer 2.11.**

Zadatak je komentirati apsolutne i relativne frekvencije varijable : Grupirana ocjena u srednjoj školi (v2) za 60% slučajno odabranih ispitanika.

**Rješenje 2.11:**

Nakon što je izvršena slučajna selekcija 60% ispitanika (kako je objašnjeno u 2.10), potrebno je na glavnom izborniku odabrati **Analyze**, a na njegovu padajućem izborniku **Descriptive Statistics**. Dalje se bira **Frequencies**, gdje se u dio prozora **Variable(s)** klikom na strelicu prebaci varijabla v2 - Grupirana ocjena u srednjoj školi. Klikom na **OK** u **Outputu** programa SPSS dobije se tablica frekvencija uzorka prema odabranom obilježju, odnosno varijabli. Rezultat je prikazan u tablici 2.10, gdje su dani rezultati samo za 60% slučajno odabranih ispitanika.

Prva frekvencija u stupcu **Frequency** u tablici 2.12 znači da je između 60% slučajno odabranih ispitanika bio 1 s ocjenom 2 u srednjoj školi. Druga frekvencija istog stupca znači da je među tim ispitanicima bilo 17 s ocjenom 3, njih 75 s ocjenom 4, a 49 s ocjenom 5. Slučajno odabranih 60% ispitanika ima 143, a od toga 1 nije odgovorio na postavljeno pitanje.

Tablica 2.12.Ispitanici prema: **Grupirana ocjena u srednjoj školi**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------|--------|-----------|---------|---------------|--------------------|
| Valid | 2,00 | 1 | ,7 | ,7 | ,7 |
| | 3,00 | 17 | 11,9 | 12,0 | 12,7 |
| | 4,00 | 75 | 52,4 | 52,8 | 65,5 |
| | 5,00 | 49 | 34,3 | 34,5 | 100,0 |
| | Total | 142 | 99,3 | 100,0 | |
| Missing | System | 1 | ,7 | | |
| Total | | 143 | 100,0 | | |

Izvor: Simulirani podaci.

Stupac **Percent** prikazuje relativne frekvencije u postotcima za valjane podatke. Druga proporcija ili relativna frekvencija pokazuje da je od 60% ispitanika njih 12% imalo prosječnu ocjenu u srednjoj školi 3.

**Analitički primjer 2.12.**

Zadatak je formirati grafikon jednostavnih stupaca za varijablu: Spol (v1).

**Rješenje 2.12:**

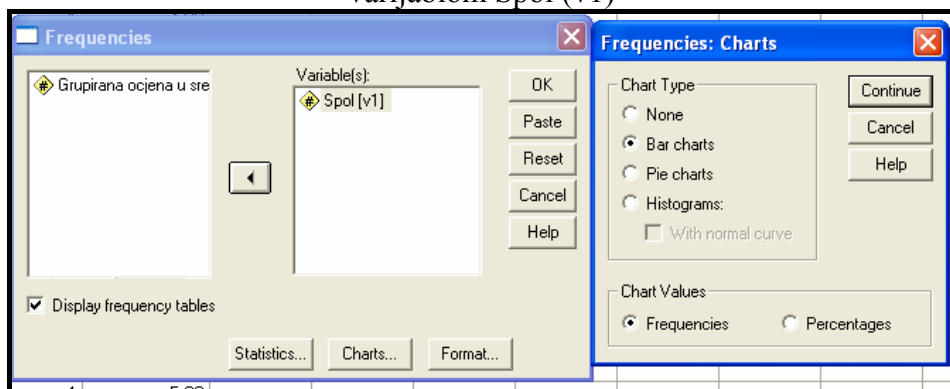
Grafikon jednostavnih stupaca za odabranu varijablu u programu **SPSS** može se nacrtati na dva načina:

I. način:

Potrebno je na glavnom izborniku odabrati **Analyze**, a na njegovu padajućem izborniku **Descriptive Statistics**. Dalje se bira **Frequencies**, gdje se u dio prozora **Variable(s)** klikom na strelicu prebaci varijabla v1 - Spol. Odabirom ikone **Charts** otvara se novi prozor **Frequencies: Charts**, gdje je potrebno aktivirati **Bar charts**. Klikom na **Continue** i **OK** u **Outputu** programa **SPSS** uz sve frekvencije dobije se i traženi grafikon.

Slika 2.13.

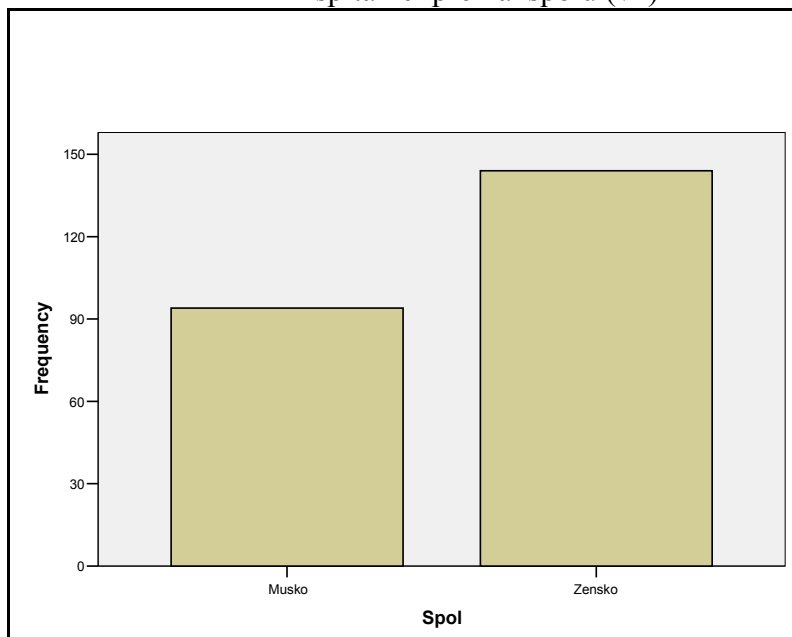
Prozori "Frequencies" i "Frequencies: Charts" s odabranim "Bar Charts" i varijablom Spol (v1)



Izvor: Simulirani podaci.

Slika 2.14.

Ispitanici prema spolu (v1)



Izvor: Simulirani podaci.

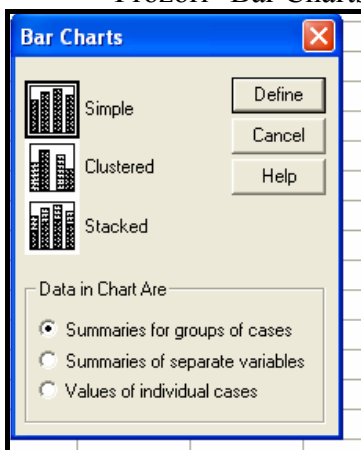
Prema slici 2.14 vidi se da je među ispitanicima više onih ženskog spola, preciznije njih 144, a onih muškog spola ima 94.

II. način:

Potrebno je na glavnom izborniku odabrati **Graphs**, a na njegovu padajućem izborniku **Bar chart**, na kojem se aktivira ikona **Simple**. To je prikazano na slici 2.15.

Slika 2.15.

Prozori "Bar Charts" s aktiviranim "Simple" grafikonima



Izvor: Simulirani podaci.

Odabirom ikone **Define** otvara se prozor **Define Simple Bar: Summaries for Groups of Cases**, gdje se u prostor **Category Axis** prebaci željena varijabla, što je u ovom slučaju Spol (v1). Grafikonu se može definirati i naslov na način da se klikne na ikonu **Titles...** Klikom na **OK** u **Outputu** programa **SPSS** dobije se traženi grafikon koji je identičan onome prikazanom na slici 2.14.



Analitički primjer 2.13.

Zadatak je formirati grafikon strukturni krug za varijablu: Spol (v1).



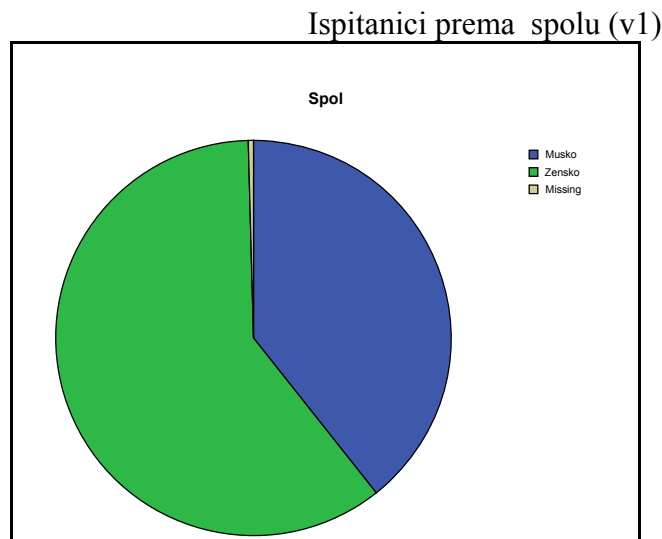
Rješenje 2.13:

Grafikon strukturnoga kruga za odabranu varijablu u programu **SPSS** može se nacrtati na dva načina:

I. način:

Potrebno je na glavnom izborniku odabrati **Analyze**, a na njegovu padajućem izborniku **Descriptive Statistics**. Dalje se bira **Frequencies**, gdje se u dio prozora **Variable(s)** klikom na strelicu prebaci varijabla v1 - Spol. Odabirom ikone **Charts** otvara se novi prozor **Frequencies: Charts**, gdje je potrebno aktivirati **Pie charts**. Klikom na **Continue** i **OK** u **Outputu** programa **SPSS** uz sve frekvencije dobije se i traženi grafikon.

Slika 2.16.



Izvor: Simulirani podaci.

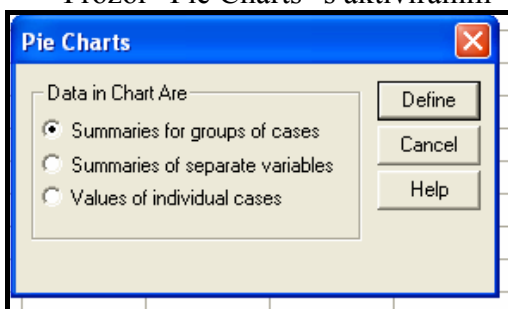
Prema slici 2.16 vidi se da je među ispitanicima više onih ženskog spola. Može se vidjeti da su na ovom grafičkom prikazu uključeni svi ispitanici, pa i oni koji nisu dali odgovor na pitanje o spolu.

II. način:

Potrebno je na glavnom izborniku odabrati **Graphs**, a na njegovu padajućem izborniku **Pie chart**, na kojem se aktivira **Summaries of groups of case**. To je prikazano na slici 2.17.

Slika 2.17.

Prozor "Pie Charts" s aktiviranim "Summaries of groups of case"



Izvor: Simulirani podaci.

Odabirom ikone **Define** otvara se prozor **Define Pie: Summaries for Groups of Cases**, gdje se u prostor **Define Slices by** prebaci željena varijabla, što je u ovom slučaju **Spol (v1)**. Grafikonu se može definirati i naslov, na način da se klikne na ikonu **Titles...** Klikom na **OK** u **Outputu** programa **SPSS** dobije se traženi grafikon koji je sličan onome prikazanom na slici 2.16, samo što ne sadrži ispitanike koji nisu odgovorili na pitanje o spolu (**Missing**).



Analitički primjer 2.14.

Za sve ispitanike zadatak je:

- a) Izračunati ukupan džeparac po obitelji (v4) ako sva djeca u jednoj obitelji dobivaju jednak iznos novca.
- b) Prikazati varijablu grupiranog džeparca (v2) grafički pomoću jednostavnih stupaca.
- c) Objasniti grafikon pod (b)!



Rješenje 2.14:

a) Za sve ispitanike potrebno je izračunati ukupan džeparac po obitelji uz pretpostavku da sva djeca u jednoj obitelji imaju jednak džeparac. Dakle, da bi se dobio ukupan džeparac po obitelji za svakog ispitanika, potrebno je džeparac ispitanika ($v1$) pomnožiti s brojem djece u svakoj obitelji ($v3$).

Da bi se pomoću postojećih numeričkih varijabli izračunala nova numerička varijabla, potrebno je na glavnom izborniku odabrati **Transform**, a u njegovu padajućem izborniku **Compute**. Na ekranu se pojavi prozor **Compute Variable**.

U polje **Target Variable** potrebno je unijeti ime nove varijable $v4$, dok se u **Numeric Expression** unosi izraz po kojem se iz postojećih varijabli kreira nova varijabla na sljedeći način:

$$v4 = v1 \times v3 . \quad (2.15)$$

Nakon što se klikne ikona **OK**, nova varijabla $v4$ automatski se formira u dodatnom stupcu u **Data View** programa **SPSS**. Nova varijabla se može vidjeti na slici 2.18.

Slika 2.18.

"Data view" s novom varijablom $v4$

| | v1 | v2 | v3 | v4 | var |
|---|---------|---------|----|---------|-----|
| 1 | 2000,00 | 3500,00 | 2 | 4000,00 | |
| 2 | . | . | 3 | . | |
| 3 | 600,00 | 600,00 | 2 | 1200,00 | |
| 4 | 1000,00 | 1000,00 | 3 | 3000,00 | |
| 5 | . | . | 2 | . | |
| 6 | 500,00 | 600,00 | 2 | 1000,00 | |
| 7 | 80,00 | 200,00 | 1 | 80,00 | |
| 8 | 100,00 | 200,00 | 3 | 300,00 | |
| 9 | 100,00 | 200,00 | 1 | 100,00 | |

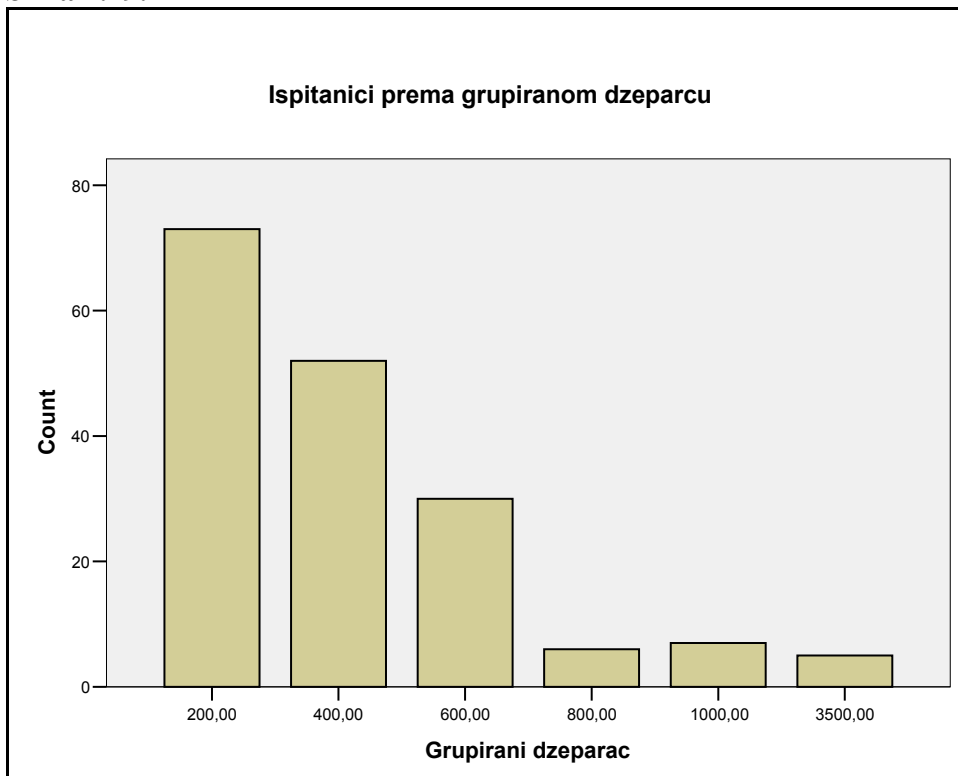
Izvor: Simulirani podaci.

U **Variable View** potrebno je na odgovarajući način definirati tu novu varijablu te u stupcu **Label** naznačiti da je to "Ukupan džeparac po obitelji". Ta nova varijabla je numerička i ima skalu mjerenja **Scale**.

b) Potrebno je na glavnom izborniku odabrati **Graphs**, a na njegovu padajućem izborniku **Bar char**, na kojem se aktivira ikona **Simple**.

Odabirom ikone **Define** otvara se prozor **Define Simple Bar: Summaries for Groups of Cases**, gdje se u prostor **Category Axis** prebaci željena varijabla, što je u ovom slučaju Grupirani džeparac (v2). Grafikonu se može definirati i naslov na način da se klikne na ikonu **Titles...** Klikom na **OK** u **Outputu** programa **SPSS** dobije se traženi grafikon koji je prikazan na slici 2.19.

Slika 2.19.



Izvor: Simulirani podaci.

- c) Na grafikonu se može vidjeti da najveći broj ispitanika ima džeparac do 200 kn. Kako raste visina džeparca, smanjuje se broj ispitanika koji ima toliki džeparac.



Analitički primjer 2.15.

Za sve ispitanike potrebno je:

- a) Grupirati varijablu Visina u cm (v2) u Grupiranu visinu (v3) po principu:
- | | | |
|-------|---|------|
| - 160 | ⇒ | 160 |
| - 170 | ⇒ | 170 |
| - 180 | ⇒ | 180 |
| - 190 | ⇒ | 190 |
| - 200 | ⇒ | 200 |
| - 210 | ⇒ | 210. |
- b) Prikazati varijablu grupirane visine (v3) grafički pomoću linijskog grafikona.
- c) Objasniti grafikon pod (b)!



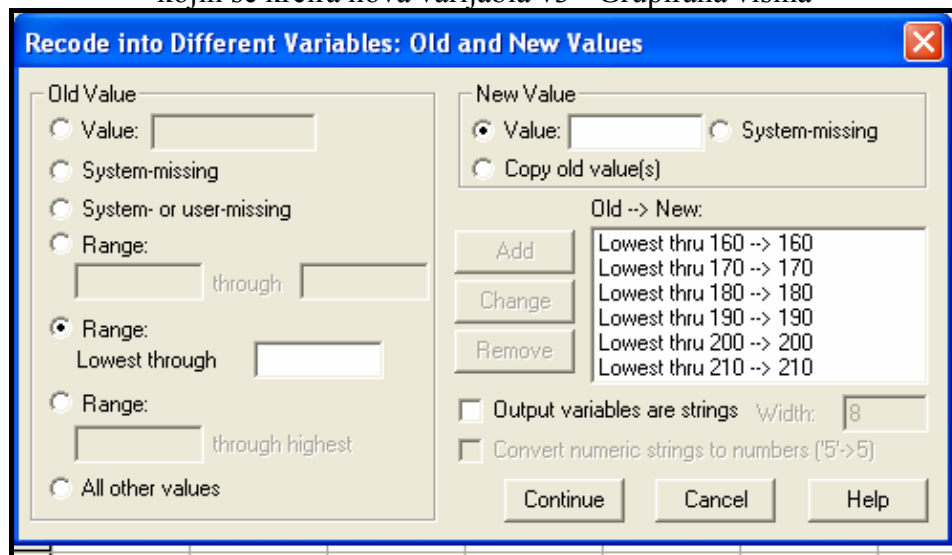
Rješenje 2.15:

- a) Da bi se pomoću varijable v2 - "Visina u cm" formirala nova varijabla v3 - "Grupirana visina " po principu koji je prikazan u zadatku pd (a), potrebno je na glavnom izborniku odabrati **Transform**, a u njegovu padajućem izborniku odabrati **Recode**. U njegovu izborniku bira se **Into Different Variables** (jer bi se odabirom **Into Same Variables** izgubila postojeća varijabla). U polje **Numeric Variable** bira se v2 - "Visina u cm", a u **Output Variable, Name** je v3, a **Label** "Grupirana visina".

Odabirom ikone **Old and New Values** otvara se novi prozor **Recode into Different Variables**. Aktiviranjem opcije **Range: Lowest through** upisuju se odgovarajuće grupe ocjena kako je prikazano na slici 2.20.

Slika 2.20.

Prozor "Recode into Different Variables: Old and New Values", pomoću kojih se kreira nova varijabla v3 - Grupirana visina

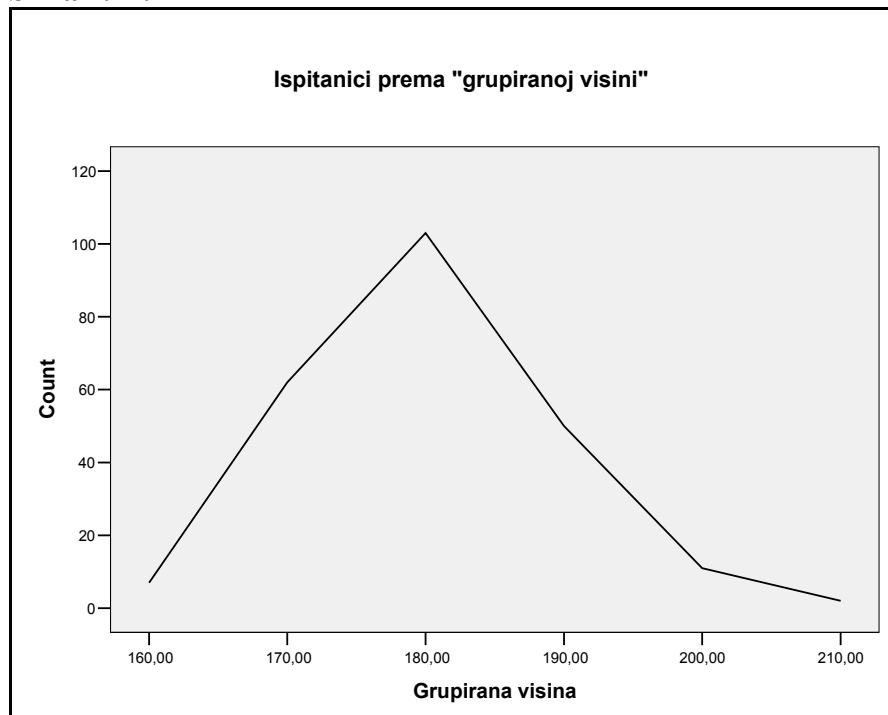


Izvor: Simulirani podaci.

Nakon upisivanja svakog pojedinog **Range: Lowest through**, upisuje se i odgovarajuća **New Value**, koja se svaki put kao nova grupa dodaje klikom na ikonu **Add**. Nakon što su unesene sve grupe ikonom **Continue**, zatim **Change** i **OK**, u novom stupcu **Data View**, kreirat će se nova varijabla "Grupirana visina", odnosno v3.

b) Potrebno je na glavnom izborniku odabrati **Graphs**, a na njegovu padajućem izborniku **Line chart** na kojem se aktivira ikona **Simple**.

Odabirom ikone **Define** otvara se prozor **Define Simple Bar: Summaries for Groups of Cases**, gdje se u prostor **Category Axis** prebaci željena varijabla, što je u ovom slučaju Grupirana visina (v3). Grafikonu se može definirati i naslov na način da se klikne na ikonu **Titles...** Klikom na **OK** u **Outputu** programa **SPSS** dobije se traženi grafikon koji je prikazan na slici 2.21.

Slika 2.21.

Izvor: Simulirani podaci.

c) Na grafikonu se može vidjeti da najveći broj ispitanika ima visinu 180 cm. Kako se visina smanjuje ili povećava, smanjuje se broj ispitanika koji ima toliku visinu.

2.5. Relativni brojevi koordinacije i indeksi

Relativni brojevi koordinacije pokazuju odnos dviju pojava ili frekvencija u različitim statističkim nizovima, koje mogu imati različitu jedinicu mjere, a koje ima smisla uspoređivati.

Računaju se:

$$R_i = \frac{A_i}{B_i}, \quad i = 1, 2, \dots, k \quad (2.16)$$

gdje je:

A_i - veličina koja se uspoređuje

B_i - veličina s kojom se A_i uspoređuje.

Grafički prikaz relativnih brojeva koordinacije je tzv. Varzarov znak. To su stupci, odnosno pravokutnici, čija baza na apscisi odgovara nazivniku (B_i), a visina relativnom broju koordinacije (R_i). Ako se zna da je površina stupca (pravokutnika) baza \times visina, može se zaključiti da površina odgovara brojniku (A_i), tj. veličini koja se uspoređuje.

Analitički primjer 2.16.

Statistički nizovi "površine i stanovništvo u odabranim zemljama 2001. godine" prema nominalnom prostornom obilježju "zemlje", prikazani su u tablici 2.13. Zadatak je izračunati broj stanovnika po km^2 .

Tablica 2.13.

Površine i stanovništvo u odabranim zemljama 2001. godine

| Zemlje | Stanovništvo u 000 | Površina u km^2 | Br.stan./ km^2 |
|---------------------|-----------------------|-----------------------------|-------------------------|
| | A_i | B_i | R_i |
| Hrvatska | 4437 | 56594 | 78,40054 |
| Slovenija | 1930 | 20273 | 95,20051 |
| Austrija | 8151 | 83871 | 97,18496 |
| Mađarska | 10106 | 93032 | 108,6293 |
| Bosna i Hercegovina | 3922 | 51197 | 76,60605 |

Izvor: Statistički ljetopis Republike Hrvatske, 2003. godina, str. 743.

Rješenje 2.16.

U posljednjem stupcu tablice 2.13 izračunat je broj stanovnika po km^2 za svaku zemlju, odnosno izračunati su relativni brojevi koordinacije.

Na primjer, za Republiku Hrvatsku:

$$R_1 = \frac{A_1}{B_1} = \frac{4437000}{56594} = 78,40 \text{ stan/km}^2.$$

Rezultat se može komentirati: u Republici Hrvatskoj je po km^2 bilo prosječno 78,40 stanovnika u 2001. godini. Najveću gustoću naseljenosti

od promatranih zemalja imala je Mađarska (108,63 stan/km²), a najnižu Bosna i Hercegovina (76,61 stan/km²).

Potrebno je napomenuti da je stanovništvo u tablici dano u tisućama, pa su kod računanja za vrijednosti A_i dodane tisuće.

Indeksima niza kvalitativnih podataka uspoređuje se smjer i intenzitet varijacija frekvencija nekog statističkog niza s takvim varijacijama drugoga statističkog niza.

Računaju se tako da se svaki član promatranog niza stavi u odnos prema odabranoj bazi, koja može biti jedan od članova niza ili neka druga zadana veličina.

$$I_i = \frac{f_i}{B} \cdot 100, \quad i = 1, 2, \dots, k. \quad (2.17)$$

Dakle, prema izrazu (2.17) indeksi se računaju kao omjer frekvencije promatranog niza i odabrane baze (B).

Grafički prikaz indeksa kvalitativnih nizova su jednostavni stupci koji imaju jednake baze, jer se ovdje članovi niza uspoređuju uvijek s istom veličinom. Specifičnost ovog grafikona jest da se stupci naslanjaju na bazu koja je jednaka 100.

Analitički primjer 2.17.

Statistički niz "prosječna godišnja osobna potrošnja po kućanstvu u kunama, RH, 2002. godine" **prema nominalnom obilježju** "vrsta potrošnje" prikazani su u tablici 2.14.

Zadatak je izračunati **indekse** potrošnje s **bazom** "potrošnja na rekreaciju i kulturu".

Rješenje 2.17.

U tablici 2.14. izračunati su indeksi za svaki pojedini modalitet vrste potrošnje, a baza tih indeksa je potrošnja na rekreaciju i kulturu. Indeksi su izračunati tako da je svaka apsolutna frekvencija niza stavljena u odnos prema zadanoj bazi. **Indeks uz bazu je uvijek = 100.** Na primjer, indeks za potrošnju na hranu i bezalkoholna pića je:

$$I_1 = \frac{f_1}{B} \cdot 100 = \frac{20816}{4155} \cdot 100 = 500,99.$$

Tablica 2.14.

Prosječna godišnja osobna potrošnja po kućanstvu u kunama, RH, 2002.
god.

| Vrsta potrošnje | Potrošnja f_i | Indeksi (Rekr. i kult.=100) |
|------------------------|--------------------|--------------------------------|
| Hrana i bezalk. pića | 20816 | 500,9868 |
| Alkohol i duhan | 2575 | 61,97353 |
| Odjeća i obuća | 5736 | 138,0505 |
| Stanovanje | 8875 | 213,5981 |
| Pokućstvo i održavanje | 3573 | 85,99278 |
| Zdravstvo | 1442 | 34,70517 |
| Rekreacija i kultura | 4155 | 100 |
| Hoteli i restorani | 2049 | 49,31408 |

Izvor: Statistički ljetopis Republike Hrvatske, 2003. godina, str. 185. i 186.

Indeksi se tumače u %, tako da se od njih oduzme 100. Može se zaključiti da je potrošnja na hranu i bezalkoholna pića bila ($I_1 - 100 = 500,99 - 100 = 400,99\%$) za 400,99% veća u odnosu na bazu, odnosno na potrošnju za rekreaciju i kulturu.

Isto tako indeks za potrošnju na odjeću i obuću je:

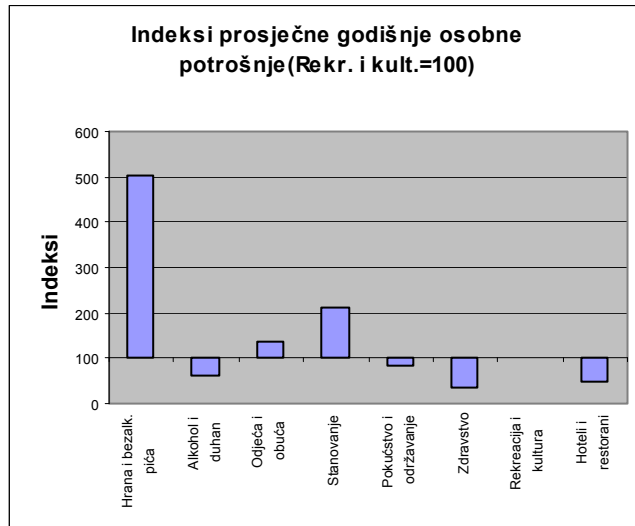
$$I_3 = \frac{f_3}{B} \cdot 100 = \frac{5736}{4155} \cdot 100 = 138,05.$$

Može se zaključiti da je potrošnja na odjeću i obuću bila ($I_3 - 100 = 138,05 - 100 = 38,05\%$) za 38,05% veća u odnosu na bazu, odnosno na potrošnju za rekreaciju i kulturu.

Na isti način indeks za potrošnju na zdravstvo je:

$$I_6 = \frac{f_6}{B} \cdot 100 = \frac{1442}{4155} \cdot 100 = 34,71.$$

Može se zaključiti da je potrošnja na zdravstvo bila ($I_6 - 100 = 34,71 - 100 = -65,29\%$) za 65,29% manja u odnosu na potrošnju za rekreaciju i kulturu.

Slika 2.21.

Izvor: Statistički ljetopis Republike Hrvatske, 2003. godina, str. 185. i 186.

Na slici 2.21 prikazani su izračunati indeksi. Na slici se vidi da su to stupci naslonjeni na bazu 100. Ako je indeks veći od 100, stupac se nalazi iznad baze, a ako je manji od 100, stupac je ispod baze. Tako se iz grafikona jasno vidi za koliki postotak su pojedine kategorije prosječne godišnje osobne potrošnje veće ili manje od baze 100, tj. potrošnje na rekreaciju i kulturu.



Analitički primjer 2.18.

Ako je prosječni džeparac 372 kn, potrebno je za sve ispitanike:

- a) Izračunati indekse varijable džeparac (v3) uzimajući za bazu indeksa prosječni džeparac po ispitaniku.



Rješenje 2.18:

Za sve ispitanike potrebno je izračunati indekse varijable džeparac (v3) uzimajući za bazu indeksa prosječni džeparac po ispitaniku koji

iznosi 372. Dakle, da bi se dobili ovi indeksi, potrebno je džeparac ispitanika (v2) podijeliti s bazom koja iznosi 372 i sve to prema jednadžbi (2.17) pomnožiti sa 100.

Da bi se pomoću postojećih numeričkih varijabli izračunala nova numerička varijabla, potrebno je na glavnom izborniku odabrati **Transform**, a u njegovu padajućem izborniku **Compute**. Na ekranu se pojavi prozor **Compute Variable**.

U polje **Target Variable** potrebno unijeti ime nove varijable v3, dok se u **Numeric Expression** unosi izraz po kojem se iz postojećih varijabli kreira nova varijabla na sljedeći način:

$$v3 = (v2 / 372) \times 100 .$$

Nakon što se klikne ikona **OK**, nova varijabla v3 automatski se formira u dodatnom stupcu u **Data View** programa **SPSS**.

U **Variable View** potrebno je na odgovarajući način definirati tu novu varijablu te u stupcu **Label** naznačiti da su to "Indeksi džeparca po ispitaniku". Ta nova varijabla je numerička i ima skalu mjerenja **Scale**.



Analitički primjer 2.19.

Zadatak je formirati dvostruku statističku tablicu, gdje se varijabla (obilježje) spol (v1) mijenja po redcima (row), a varijabla (obilježje) grupirani džeparac (v2) po stupcima (column):

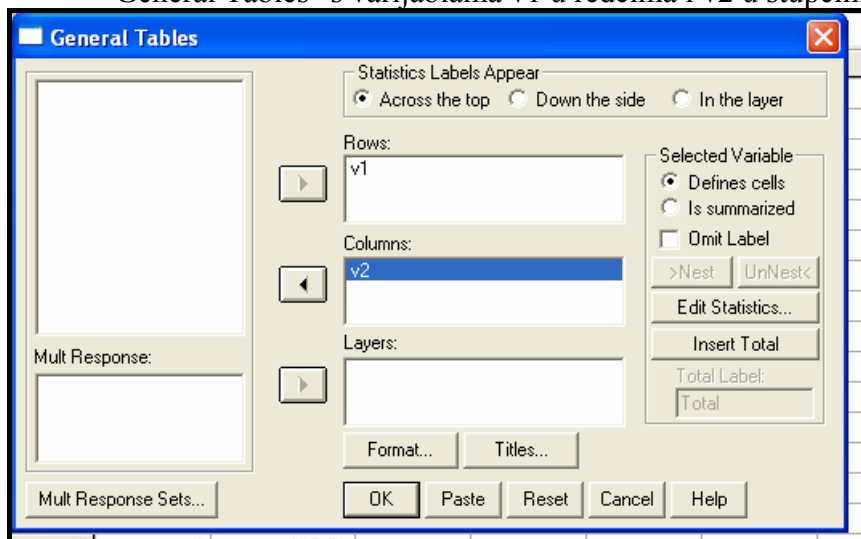
- a) Formirati dvostruku statističku tablicu bez zbirnih redaka i stupaca.
- b) Formirati dvostruku statističku tablicu sa zbirnim redcima i stupcima.
- c) Nacrtati grafikon višestrukih stupaca prema odabranim varijablama i komentirati sliku!

**Rješenje 2.19:**

a) Da bi se formirala dvostruka statistička tablica bez zbirnih redaka i stupaca, potrebno je na glavnom izborniku izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Tables**, i **General Tables**. Prema zahtjevu zadatka pomoću odgovarajućih strelica varijabla v1 - Spol se prebacuje u polje **Rows**, a druga varijabla v2 - Grupirani džeparac u polje **Columns**. To je prikazano na slici 2.23.

Slika 2.23.

"General Tables" s varijablama v1 u redcima i v2 u stupcima



Izvor: Simulirani podaci.

Klikom na **OK**, u **Outputu** se za rezultat dobije tražena tablica koja je prikazana u sljedećoj tablici 2.15.

Tablica 2.15.

Ispitanici prema spolu i grupiranom džeparcu

| | | Grupirani džeparac | | | | | |
|------|--------|--------------------|--------|--------|--------|---------|---------|
| | | 200,00 | 400,00 | 600,00 | 800,00 | 1000,00 | 3500,00 |
| Spol | Muško | 28 | 18 | 10 | 4 | 4 | 4 |
| | Žensko | 45 | 34 | 20 | 2 | 3 | 1 |

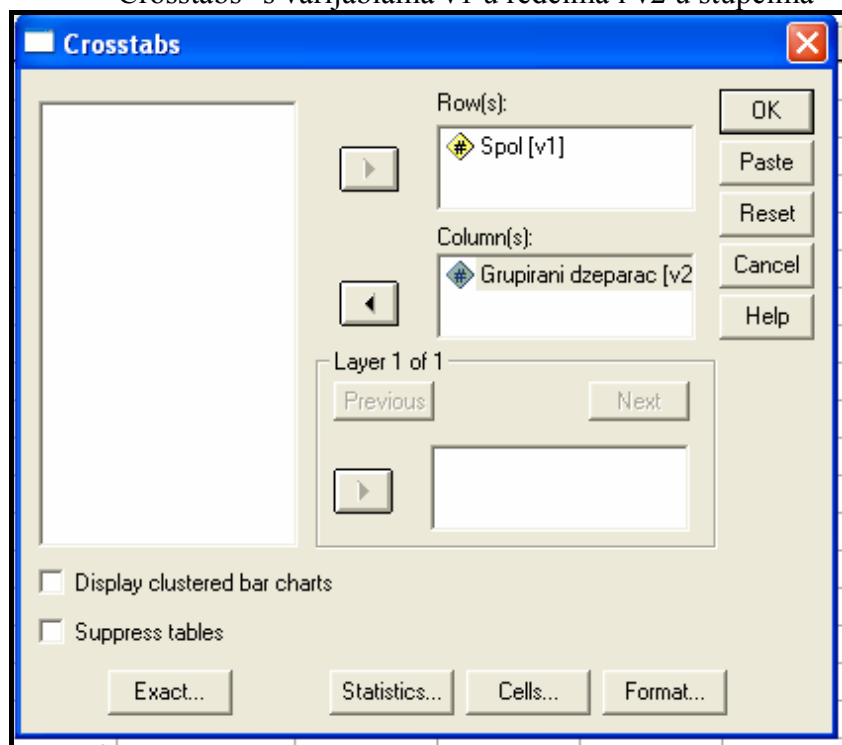
Izvor: Simulirani podaci.

U tablici 2.15. vidi se da najviše ispitanika ima džeparac do 200 kn. Iako je u analizi više ispitanika ženskog spola, pokazalo se da veći džeparac imaju oni muškog spola.

b) Da bi se formirala dvostruka statistička tablica sa zbirnim redcima i stupcima, potrebno je na glavnom izborniku izabrati ikonu *Analyze*, a na njezinu padajućem izborniku *Descriptive Statistics* i *Crosstabs*. Prema zahtjevu zadatka pomoću odgovarajućih strelica varijabla v1 - Spol se prebacuje u polje *Row(s)*, a druga varijabla v2 - Grupirani džeparac u polje *Column(s)*. To je prikazano na slici 2.24.

Slika 2.24.

"Crosstabs" s varijablama v1 u redcima i v2 u stupcima



Izvor: Simulirani podaci.

Klikom na *OK* u *Outputu* se za rezultat dobije tražena tablica sa zbirnim stupcima i redcima koja je prikazana u sljedećoj tablici 2.16.

Tablica 2.16.

Ispitanici prema spolu i grupiranom džeparcu

Spol * Grupirani džeparac Crosstabulation

Count

| | | Grupirani džeparac | | | | | | Total |
|-------|--------|--------------------|--------|--------|--------|---------|---------|-------|
| | | 200,00 | 400,00 | 600,00 | 800,00 | 1000,00 | 3500,00 | |
| Spol | Musko | 28 | 18 | 10 | 4 | 4 | 4 | 68 |
| | Zensko | 45 | 34 | 20 | 2 | 3 | 1 | 105 |
| Total | | 73 | 52 | 30 | 6 | 7 | 5 | 173 |

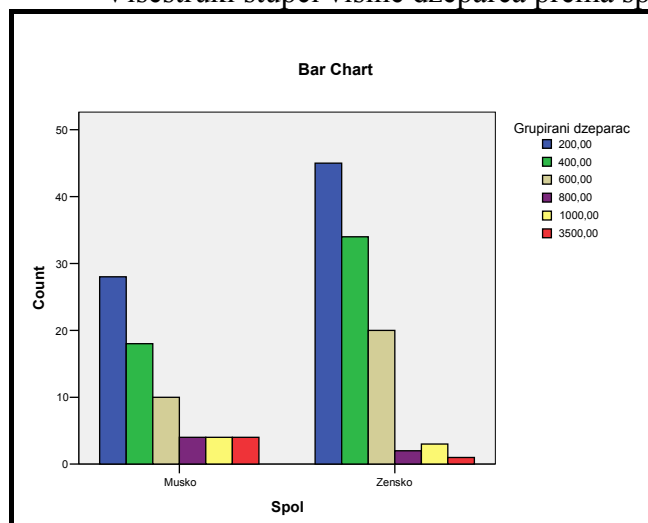
Izvor: Simulirani podaci.

Za razliku od tablice 2.15 u tablici 2.16 postoje zbirni stupci i zbirni redci. U njima se nalaze marginalne frekvencije koje kažu koliko ispitanika ima po jednoj grupi obilježja bez obzira na drugo obilježje. Na primjer, prvi podatak u zbirnom stupcu kaže da je 68 ispitanika muškog spola bez obzira na džeparac. Treći podatak u zbirnom retku kaže da je 30 ispitanika imalo džeparac između 400 i 600 kn bez obzira na spol.

c) Da bi se u **Outputu** dobio i grafikon višestrukih stupaca pri kreiranju dvostruke statističke tablice u prozoru Crosstabs (slika 2.24), potrebno je aktivirati **Display clustered bar charts**.

Slika 2.25.

Višestruki stupci visine džeparca prema spolu

*Izvor: Simulirani podaci.*

Traženi grafikon prikazan je na slici 2.25. Može se vidjeti da se prvi višestruki stupac odnosi na ispitanike muškog spola, a drugi na ispitanike ženskog spola. I u jednom i u drugom dijelu grafikona najveći stupci su oni koji prikazuju niži džeparac, dok oni manji stupci pokazuju viši iznos džeparca.



Analitički primjer 2.20.

Zadatak je formirati dvostruku statističku tablicu gdje se varijabla (obilježje) spol (v1) mijenja po stupcima (column), a varijabla (obilježje) grupirani džeparac (v2) po redcima (row):

- a) Formirati dvostruku statističku tablicu bez zbirnih redaka i stupaca.
- b) Formirati dvostruku statističku tablicu sa zbirnim redcima i stupcima.
- c) Nacrtati grafikon dvostrukih stupaca prema odabranim varijablama i komentirati sliku!



Rješenje 2.20:

a) Da bi se formirala dvostruka statistička tablica bez zbirnih redaka i stupaca, potrebno je na glavnom izborniku izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Tables** i **General Tables**. Prema zahtjevu zadatka pomoću odgovarajućih strelica varijabla v1 - Spol se prebacuje u polje **Columns**, a druga varijabla v2 - Grupirani džeparac u polje **Rows**. Princip je kao što je prikazano na slici 2.20, samo što odabrane varijable po stupcima i po redcima mijenjaju svoje mjesto.

Klikom na **OK**, u **Outputu** se za rezultat dobije tražena tablica koja je prikazana u sljedećoj tablici 2.17.

Tablica 2.17.

Ispitanici prema spolu i grupiranom džeparcu

| | | Spol | |
|--------------------|---------|-------|--------|
| | | Musko | Zensko |
| Grupirani džeparac | 200,00 | 28 | 45 |
| | 400,00 | 18 | 34 |
| | 600,00 | 10 | 20 |
| | 800,00 | 4 | 2 |
| | 1000,00 | 4 | 3 |
| | 3500,00 | 4 | 1 |

Izvor: Simulirani podaci.

Ako se napravi usporedba tablice 2.17 s tablicom 2.15, može se vidjeti da se sada varijabla spol mijenja po stupcima, a varijabla grupirani džeparac po redcima. I iz ove tablice jasno je da najveći broj ispitanika i jednog i drugog spola imaju najmanju kategoriju džeparca, tj. do 200 kn na mjesec.

b) Da bi se formirala dvostruka statistička tablica s zbirnim redcima i stupcima, potrebno je na glavnom izborniku izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Crosstabs**. Prema zahtjevu zadatka pomoću odgovarajućih strelica varijabla v1 - Spol se prebacuje u polje **Column(s)**, a druga varijabla v2 - Grupirani džeparac u polje **Row(s)**.

Tablica 2.18.Ispitanici prema spolu i grupiranom džeparcu
Grupirani džeparac * Spol Crosstabulation

| Count | | Spol | | Total |
|--------------------|---------|-------|--------|-------|
| | | Musko | Zensko | |
| Grupirani džeparac | 200,00 | 28 | 45 | 73 |
| | 400,00 | 18 | 34 | 52 |
| | 600,00 | 10 | 20 | 30 |
| | 800,00 | 4 | 2 | 6 |
| | 1000,00 | 4 | 3 | 7 |
| | 3500,00 | 4 | 1 | 5 |
| Total | | 68 | 105 | 173 |

Izvor: Simulirani podaci.

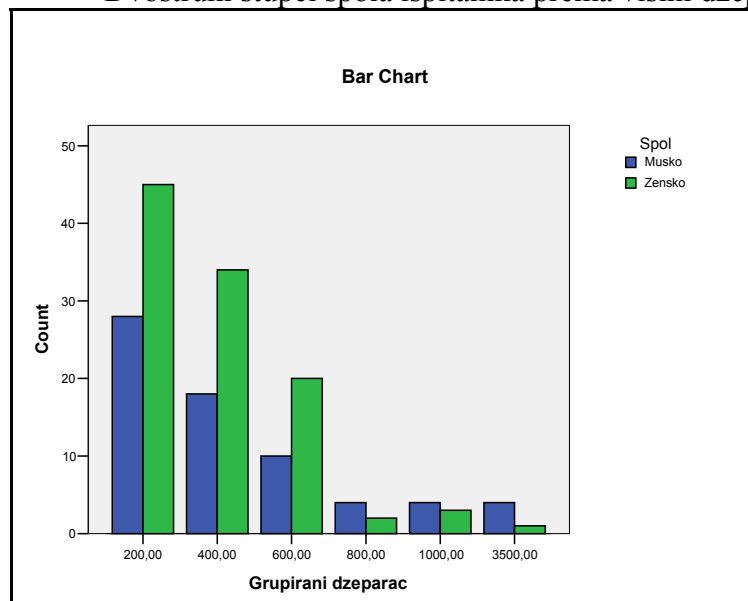
Sličan je princip kao što je prikazano na slici 2.24, samo što odabrane varijable po stupcima i po redcima mijenjaju svoje mjesto. Klikom na **OK** u **Outputu** se za rezultat dobije tražena tablica sa zbirnim stupcima i redcima koja je prikazana u tablici 2.18.

Za razliku od tablice 2.17 u tablici 2.18 postoje zbirni stupci i zbirni redci. U njima se nalaze marginalne frekvencije koje kažu koliko ispitanika ima po jednoj grupi obilježja bez obzira na drugo obilježje. Na primjer, prvi podatak u zbirnom retku kaže da je 68 ispitanika muškog spola bez obzira na džeparac. Treći podatak u zbirnom stupcu kaže da je 30 ispitanika imalo džeparac između 400 i 600 kn bez obzira na spol.

c) Da bi se u **Outputu** dobio i grafikon višestrukih stupaca pri kreiranju dvostruke statističke tablice u prozoru Crosstabs (slika 2.24), potrebno je aktivirati **Display clustered bar charts**.

Slika 2.26.

Dvostruki stupci spola ispitanika prema visini džeparca



Izvor: Simulirani podaci.

Traženi grafikon prikazan je na slici 2.26. Može se vidjeti da se svaki dvostruki stupac odnosi na različitu kategoriju visine džeparca. U prva tri dvostruka stupca veću zastupljenost imaju ispitanici ženskog

spola, a prema posljednja tri stupca može se zaključiti da je muški spol u većini. Dakle, općenito ispitanici ženskog spola imaju niži iznos džeparca, dok ispitanici muškog spola imaju veći iznos džeparca.



Analitički primjer 2.21.

Zadatak je formirati dvostruku statističku tablicu gdje se varijabla (obilježje) školska sprema oca (v1) mijenja po redcima (row), a varijabla (obilježje) priključenje EU (v2) po stupcima (column):

- Tablica sa zbirnim redcima i stupcima te nacrtati grafikon višestrukih stupaca!
- Komentirati sliku i izračunate frekvencije!



Rješenje 2.21:

- Na glavnom izborniku potrebno je izabrati ikonu *Analyze*, a na njezinu padajućem izborniku *Descriptive Statistics* i *Crosstabs*.

Tablica 2.19.

Ispitanici prema školskoj spremi oca i mišljenju o priključenju EU
Školska sprema oca * Jeste li za priključenje RH u EU Crosstabulation

| Count | | Jeste li za priključenje RH EU | | Total |
|--------------------------|-------------------------|-----------------------------------|----|-------|
| | | Da | Ne | |
| Školska sprema oca | Osnovna skola | 22 | 0 | 22 |
| | Srednja četverogodišnja | 67 | 6 | 73 |
| | Srednja trogodišnja | 27 | 2 | 29 |
| | Visa | 39 | 5 | 44 |
| | Visoka | 53 | 4 | 57 |
| | Magisterij | 4 | 1 | 5 |
| | Doktorat | 2 | 1 | 3 |
| Total | | 214 | 19 | 233 |

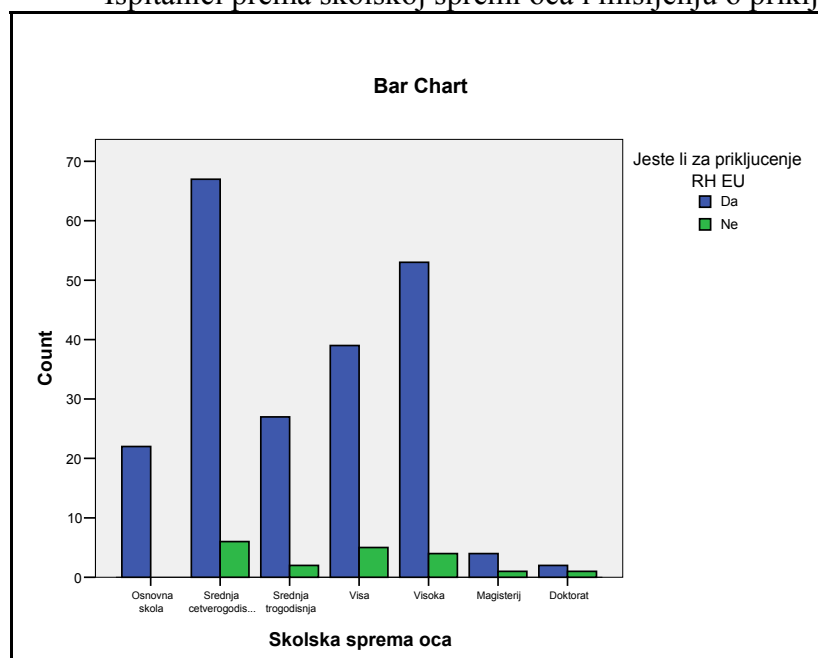
Izvor: Simulirani podaci.

Pomoću odgovarajućih strelica varijabla v1 - Školska sprema oca se prebacuje u polje **Row(s)**, a druga varijabla v2 - Priključenje EU u polje **Column(s)**. Da bi se u **Outputu** dobio i grafikon višestrukih stupaca pri kreiranju dvostruke statističke tablice u prozoru Crosstabs, potrebno je aktivirati **Display clustered bar charts**.

Klikom na **OK** u **Outputu** se za rezultat dobije tražena dvostruka tablica, koja je prikazana tablicom 2.19 i grafikon, prikazan na slici 2.27.

Slika 2.27.

Ispitanici prema školskoj spremi oca i mišljenju o priključenju EU



Izvor: Simulirani podaci.

b) Prema podacima u tablici 2.19 i slici 2.27 može se zaključiti da je većina ispitanika iz odabranog uzorka za priključenje Republike Hrvatske Europskoj uniji, bez obzira na školsku spremu njihova oca. Isto tako može se primijetiti da je u uzorku najveći broj ispitanika kojima otac ima srednju stručnu spremu (što je zaista i općenita situacija u Republici Hrvatskoj - prema podacima u SLJH, razne godine - pa bi se u tom smislu moglo komentirati o slučajnosti izabranog uzorka).



Analitički primjer 2.22.

Zadatak je formirati dvostruku statističku tablicu gdje se varijabla (obilježje) priključenje EU (v1) mijenja po redcima (row), a varijabla (obilježje) vjenčanje sa osobom druge vjere (v2) po stupcima (column):

- Formirati tablicu sa zbirnim redcima i stupcima te nacrtati grafikon višestrukih stupaca!
- Komentirati sliku i izračunate frekvencije!



Rješenje 2.22:

- Na glavnom izborniku potrebno je izabrati ikonu *Analyze*, a na njezinu padajućem izborniku *Descriptive Statistics* i *Crosstabs*.

Pomoću odgovarajućih strelica varijabla v1 - Priključenje EU u polje **Row(s)**, a druga varijabla v2 - Vjenčanje sa osobom druge vjere u polje **Column(s)**. Da bi se u **Outputu** dobio i grafikon višestrukih stupaca pri kreiranju dvostruke statističke tablice u prozoru Crosstabs, potrebno je aktivirati **Display clustered bar charts**.

Klikom na **OK** u **Outputu** se za rezultat dobije tražena dvostruka tablica koja je prikazana tablicom 2.20 i grafikon prikazan na slici 2.28.

Tablica 2.20.

Ispitanici prema mišljenju o priključenju EU i o vjenčanju sa osobom druge vjere

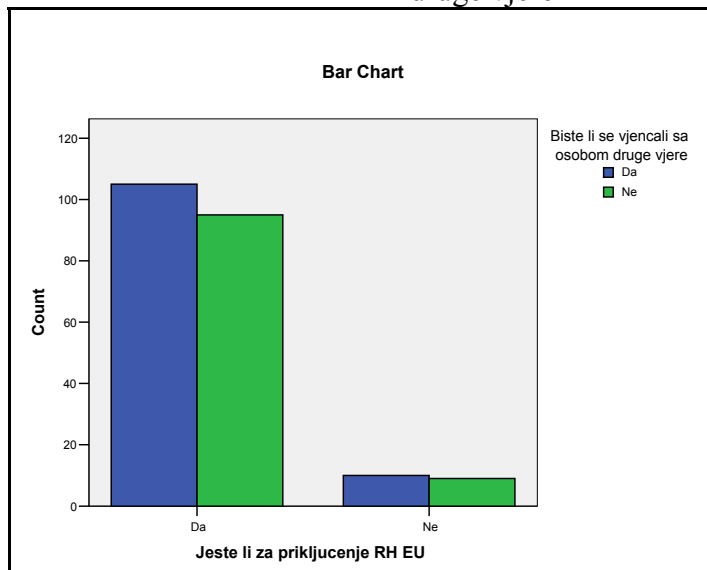
Jeste li za priključenje RH EU * Biste li se vjenčali sa osobom druge vjere
Crosstabulation

| Count | | Biste li se vjenčali sa osobom druge vjere | | Total |
|--------------------------------|----|--|-----|-------|
| | | Da | Ne | |
| Jeste li za priključenje RH EU | Da | 105 | 95 | 200 |
| | Ne | 10 | 9 | 19 |
| Total | | 115 | 104 | 219 |

Izvor: Simulirani podaci.

Slika 2.28

Ispitanici prema mišljenju o priključenju EU i o vjenčanju sa osobom druge vjere



Izvor: Simulirani podaci.

b) Prema podacima u tablici 2.20 i slici 2.28 može se zaključiti da je većina ispitanika iz odabranog uzorka za priključenje Republike Hrvatske Europskoj uniji, što se vidi prema marginalnim frekvencijama u zbirnom stupcu tablice i prema prvom dvostrukom stupcu grafikona. Isto tako može se primijetiti da je u uzorku broj ispitanika koji bi se vjenčao sa osobom druge vjere malo veći od onih koji to ne bi napravili, te da su otprilike ravnomjerno raspoređeni bez obzira na mišljenje o tome jesu li za priključenje Republike Hrvatske Europskoj uniji.

**Analitički primjer 2.23.**

Zadatak je formirati dvostruku statističku tablicu, gdje se varijabla (obilježje) odnosi u obitelji (v1) mijenja po redcima (row), a varijabla (obilježje) uživajte li drogu (v2) po stupcima (column). Potrebno je:

- a) Formirati tablicu sa zbirnim redcima i stupcima te nacrtati grafikon višestrukih stupaca!
- b) Komentirati sliku i izračunate frekvencije!



Rješenje 2.23:

- a) Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Crosstabs**.

Pomoću odgovarajućih strelica varijabla v2 - Odnosi u obitelji prebacuje se u polje **Row(s)**, a druga varijabla v1 - Uživete li drogu u polje **Column(s)**. Da bi se u **Outputu** dobio i grafikon višestrukih stupaca pri kreiranju dvostruke statističke tablice u prozoru Crosstabs, potrebno je aktivirati **Display clustered bar charts**.

Klikom na **OK** u **Outputu** se za rezultat dobije tražena dvostruka tablica, koja je prikazana tablicom 2.21 i grafikon prikazan na slici 2.29.

Tablica 2.21.

Ispitanici prema odnosima u njihovoj obitelji i konzumiranju droge
Odnosi u obitelji * Uživete li drogu Crosstabulation

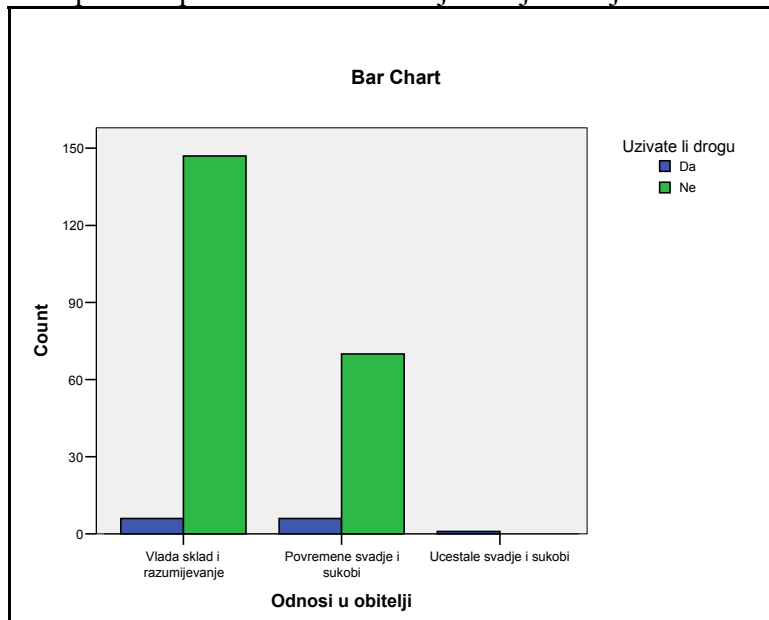
| Count | | Uživete li drogu | | Total |
|-------------------|-----------------------------|------------------|-----|-------|
| | | Da | Ne | |
| Odnosi u obitelji | Vlada sklad i razumijevanje | 6 | 147 | 153 |
| | Povremene svadje i sukobi | 6 | 70 | 76 |
| | Učestale svadje i sukobi | 1 | 0 | 1 |
| Total | | 13 | 217 | 230 |

Izvor: Simulirani podaci.

- b) Prema podacima u tablici 2.21 i slici 2.29 može se zaključiti da većina ispitanika iz odabranog uzorka ne konzumira drogu. Prema prvoj marginalnoj frekvenciji u zbirnom stupcu, te prvom dvostrukom stupcu grafikona, može se vidjeti da u obiteljima kod većine ispitanika vlada sklad i razumijevanje. Samo 1 ispitanik izjasnio se da u obitelji ima učestale svađe i sukobe i on konzumira drogu (treći dvostruki stupac grafikona).

Slika 2.29.

Ispitanici prema odnosima u njihovoj obitelji i konzumiranju droge



Izvor: Simulirani podaci.



Analitički primjer 2.24.

Zadatak je formirati dvostruku statističku tablicu gdje se varijabla (obilježje) Školska sprema oca (v1) mijenja po redcima (row), a varijabla (obilježje) Uživete li drogu (v2) po stupcima (column). Potrebno je:

- Formirati tablicu sa zbirnim redcima i stupcima te nacrtati grafikon višestrukih stupaca!
- Komentirati sliku i izračunate frekvencije!



Rješenje 2.24:

- Na glavnom izborniku potrebno je izabrati ikonu *Analyze*, a na njezinu padajućem izborniku *Descriptive Statistics* i *Crosstabs*.

Pomoću odgovarajućih strelica varijabla v1 - Školska sprema oca **Row(s)**, a druga varijabla v2 - Uživete li drogu u polje **Column(s)**. Da bi se u **Outputu** dobio i grafikon višestrukih stupaca pri kreiranju dvostruke statističke tablice u prozoru Crosstabs, potrebno je aktivirati **Display clustered bar charts**.

Klikom na **OK** u **Outputu** se za rezultat dobije tražena dvostruka tablica, koja je prikazana tablicom 2.22 i grafikon prikazan na slici 2.30.

Tablica 2.22.

Ispitanici prema školskoj spremi oca i konzumiranju droge
Školska sprema oca * Uživete li drogu Crosstabulation

Count

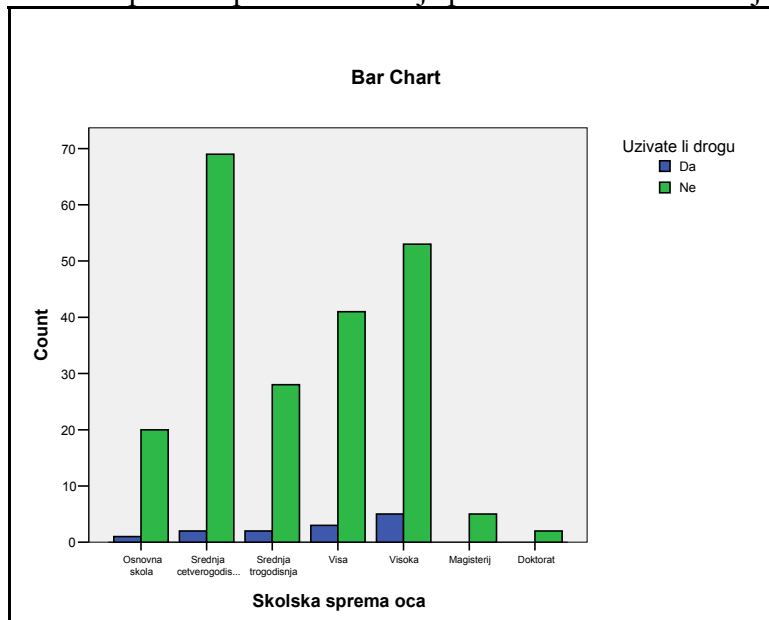
| | | Uživete li drogu | | Total |
|--------------------|-------------------------|------------------|-----|-------|
| | | Da | Ne | |
| Školska sprema oca | Osnovna skola | 1 | 20 | 21 |
| | Srednja četverogodišnja | 2 | 69 | 71 |
| | Srednja trogodišnja | 2 | 28 | 30 |
| | Visa | 3 | 41 | 44 |
| | Visoka | 5 | 53 | 58 |
| | Magisterij | 0 | 5 | 5 |
| | Doktorat | 0 | 2 | 2 |
| Total | | 13 | 218 | 231 |

Izvor: Simulirani podaci.

b) Prema podacima u tablici 2.22 i slici 2.30 može se zaključiti da većina ispitanika iz odabranog uzorka ne konzumira drogu. Prema zadnje dvije marginalne frekvencije u zbirnom stupcu, te zadnja dva dvostruka stupca grafikona, može se vidjeti da samo 5 ispitanika ima oca s titulom magistra, a 2 ispitanika imaju oca s doktoratom. Među njima nitko ne konzumira drogu.

Slika 2.30.

Ispitanici prema školskoj spremi oca i konzumiranju droge



Izvor: Simulirani podaci.



Analitički primjer 2.25.

Zadatak je formirati dvostruku statističku tablicu, gdje se varijabla (obilježje) Broj djece u obitelji (v1) mijenja po redcima (row), a varijabla (obilježje) Uživete li drogu (v2) po stupcima (column). Potrebno je:

- Formirati tablicu sa zbirnim redcima i stupcima te nacrtati grafikon višestrukih stupaca!
- Komentirati sliku i izračunate frekvencije!



Rješenje 2.25:

- Na glavnom izborniku potrebno je izabrati ikonu *Analyze*, a na njezinu padajućem izborniku *Descriptive Statistics* i *Crosstabs*.

Pomoću odgovarajućih strelica varijabla v1 - Broj djece u obitelji prebacuje se u polje **Row(s)**, a druga varijabla v2 - Uživete li drogu u polje **Column(s)**. Da bi se u **Outputu** dobio i grafikon višestrukih stupaca pri kreiranju dvostruke statističke tablice u prozoru Crosstabs, potrebno je aktivirati **Display clustered bar charts**.

Klikom na **OK** u **Outputu** se za rezultat dobije tražena dvostruka tablica koja je prikazana tablicom 2.23 i grafikon prikazan na slici 2.31.

Tablica 2.23.

Ispitanici prema broju djece u obitelji i konzumiranju droge
Broj djece u obitelji * Uživete li drogu Crosstabulation

Count

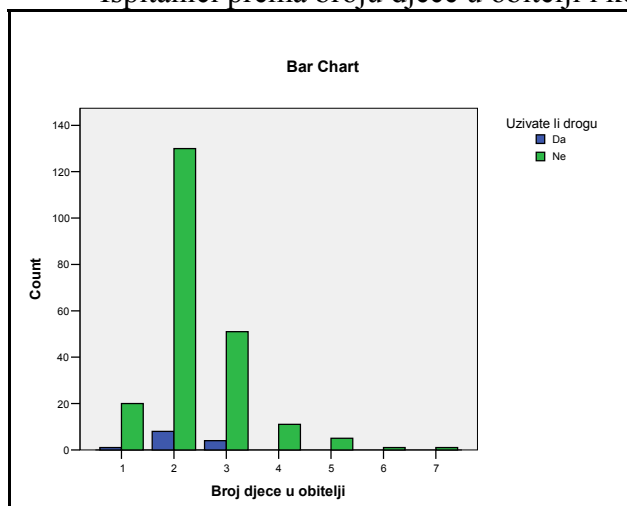
| | | Uživete li drogu | | Total |
|-----------------------|---|------------------|-----|-------|
| | | Da | Ne | |
| Broj djece u obitelji | 1 | 1 | 20 | 21 |
| | 2 | 8 | 130 | 138 |
| | 3 | 4 | 51 | 55 |
| | 4 | 0 | 11 | 11 |
| | 5 | 0 | 5 | 5 |
| | 6 | 0 | 1 | 1 |
| | 7 | 0 | 1 | 1 |
| Total | | 13 | 219 | 232 |

Izvor: Simulirani podaci.

b) Prema podacima u tablici 2.23 i slici 2.31 može se zaključiti da većina ispitanika iz odabranog uzorka ne konzumira drogu. Prema zadnje dvije marginalne frekvencije u zbirnom stupcu, te zadnja dva dvostruka stupca grafikona, može se vidjeti da samo po 1 ispitanik u obitelji ima 6-ero i 7-ero djece. Među njima nitko ne konzumira drogu.

Slika 2.31.

Ispitanici prema broju djece u obitelji i konzumiranju droge



Izvor: Simulirani podaci.



Analitički primjer 2.26.

Zadatak je formirati dvostruku statističku tablicu gdje se varijabla (obilježje) Financijske prilike u obitelji (v1) mijenja po redcima (row), a varijabla (obilježje) Broj djece u obitelji (v2) po stupcima (column). Potrebno je:

- Formirati tablicu sa zbirnim redcima i stupcima te nacrtati grafikon višestrukih stupaca!
- Komentirati sliku i izračunate frekvencije!



Rješenje 2.26:

- Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Crosstabs**.

Pomoću odgovarajućih strelica varijabla v1 - Financijske prilike u obitelji prebacuje se u polje **Row(s)**, a druga varijabla v2 - Broj djece u

obitelji u polje **Column(s)**. Da bi se u **Outputu** dobio i grafikon višestrukih stupaca pri kreiranju dvostruke statističke tablice u prozoru Crosstabs, potrebno je aktivirati **Display clustered bar charts**.

Klikom na **OK** u **Outputu** se za rezultat dobije tražena dvostruka tablica koja je prikazana tablicom 2.24 i grafikon prikazan na slici 2.32.

Tablica 2.24.

Ispitanici prema financijskim prilikama i broju djece u obitelji
Financijske prilike u obitelji * Broj djece u obitelji Crosstabulation

Count

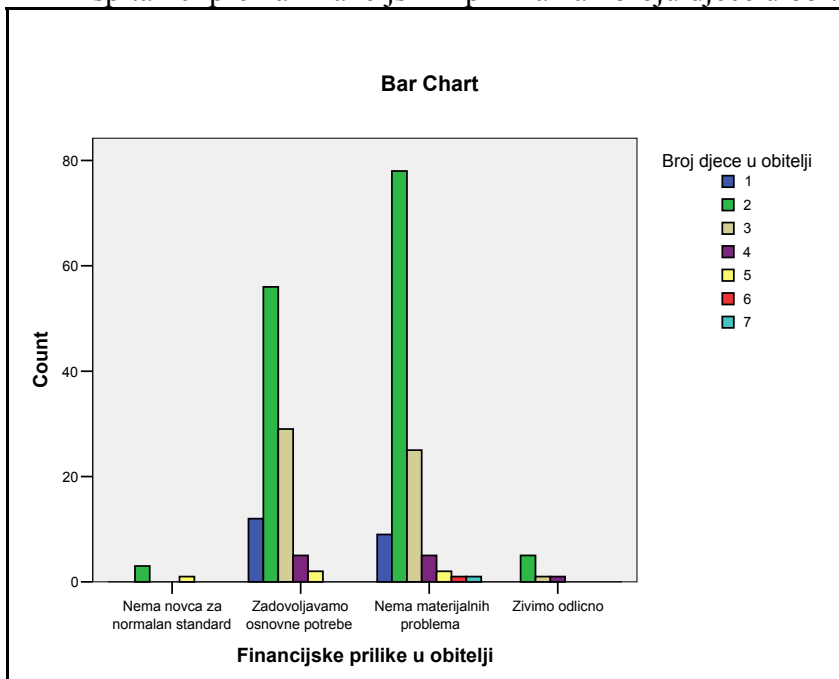
| | | Broj djece u obitelji | | | | | | | Total |
|--------------------------------|---------------------------------|-----------------------|-----|----|----|---|---|---|-------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
| Financijske prilike u obitelji | Nema novca za normalan standard | 0 | 3 | 0 | 0 | 1 | 0 | 0 | 4 |
| | Zadovoljavamo osnovne potrebe | 12 | 56 | 29 | 5 | 2 | 0 | 0 | 104 |
| | Nema materijalnih problema | 9 | 78 | 25 | 5 | 2 | 1 | 1 | 121 |
| | Zivimo odlično | 0 | 5 | 1 | 1 | 0 | 0 | 0 | 7 |
| Total | | 21 | 142 | 55 | 11 | 5 | 1 | 1 | 236 |

Izvor: Simulirani podaci.

b) Prema podacima u tablici 2.24 i slici 2.32 može se zaključiti da kod većine ispitanika iz odabranog uzorka ima 2 djece u obitelji. Ispitanika koji su se izjasnili da nemaju novca za normalan standard te onih koji žive odlično ima znatno manje u odnosu na ostale kategorije (prvi i zadnji višestruki stupac grafikona na slici 2.32).

Slika 2.32.

Ispitanici prema financijskim prilikama i broju djece u obitelji



Izvor: Simulirani podaci.



Analitički primjer 2.27.

Zadatak je formirati dvostruku statističku tablicu gdje se varijabla (obilježje) Financijske prilike u obitelji (v1) mijenja po redcima (row), a varijabla (obilježje) Uživete li drogu (v2) po stupcima (column). Potrebno je:

- Formirati tablicu sa zbirnim redcima i stupcima te nacrtati grafikon višestrukih stupaca!
- Komentirati sliku i izračunate frekvencije!

**Rješenje 2.27:**

a) Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Crosstabs**.

Pomoću odgovarajućih strelica varijabla v1 - Financijske prilike u obitelji prebacuje se u polje **Row(s)**, a druga varijabla v2 - Uživete li drogu u polje **Column(s)**. Da bi se u **Outputu** dobio i grafikon višestrukih stupaca pri kreiranju dvostruke statističke tablice u prozoru Crosstabs, potrebno je aktivirati **Display clustered bar charts**.

Klikom na **OK** u **Outputu** se za rezultat dobije tražena dvostruka tablica koja je prikazana tablicom 2.25 i grafikon prikazan na slici 2.33.

Tablica 2.25.

Ispitanici prema financijskim prilikama u obitelji i konzumiranju droge
Financijske prilike u obitelji * Uživete li drogu Crosstabulation

Count

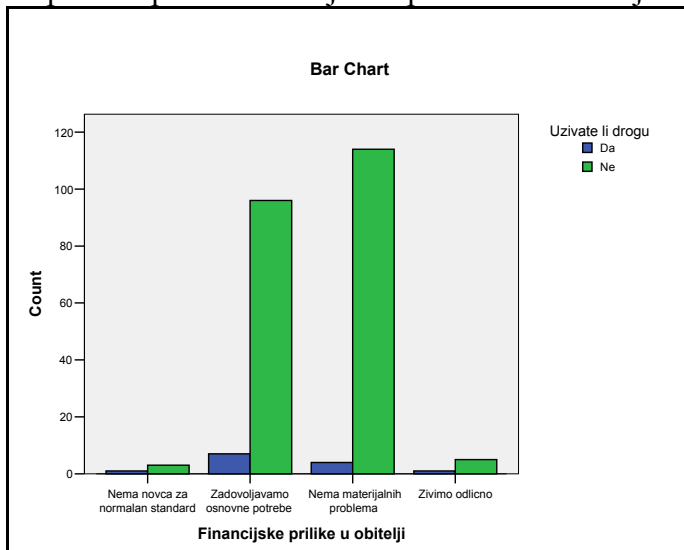
| | | Uživete li drogu | | Total |
|--------------------------------|---------------------------------|------------------|-----|-------|
| | | Da | Ne | |
| Financijske prilike u obitelji | Nema novca za normalan standard | 1 | 3 | 4 |
| | Zadovoljavamo osnovne potrebe | 7 | 96 | 103 |
| | Nema materijalnih problema | 4 | 114 | 118 |
| | Zivimo odlično | 1 | 5 | 6 |
| Total | | 13 | 218 | 231 |

Izvor: Simulirani podaci.

b) Prema podacima u tablici 2.25 i slici 2.33 može se zaključiti da većina ispitanika iz odabranog uzorka ipak ne konzumira drogu. Ispitanika koji su se izjasnili da nemaju novca za normalan standard te onih koji žive odlično ima znatno manje u odnosu na ostale kategorije (prvi i zadnji višestruki stupac grafikona na slici 2.33).

Slika 2.33.

Ispitanici prema financijskim prilikama u obitelji i konzumiranju droge



Izvor: Simulirani podaci.



Analitički primjer 2.28.

Zadatak je formirati dvostruku statističku tablicu gdje se varijabla (obilježje) Grupirana ocjena u srednjoj školi (v1) mijenja po redcima (row), a varijabla (obilježje) Uživete li drogu (v2) po stupcima (column). Potrebno je:

- Formirati tablicu sa zbirnim redcima i stupcima te nacrtati grafikon višestrukih stupaca!
- Komentirati sliku i izračunate frekvencije!



Rješenje 2.28:

- Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Crosstabs**.

Tablica 2.26.

Ispitanici prema ocjeni iz srednje škole i konzumiranju droge
Grupirana ocjena u srednjoj školi * Uživete li drogu Crosstabulation

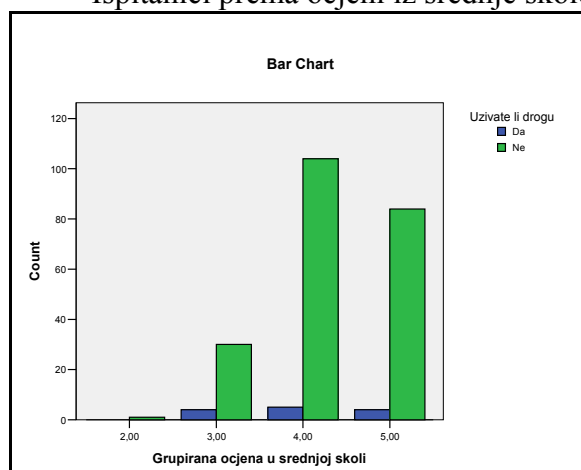
Count

| | | Uživete li drogu | | Total |
|-----------------------------------|------|------------------|-----|-------|
| | | Da | Ne | |
| Grupirana ocjena u srednjoj školi | 2,00 | 0 | 1 | 1 |
| | 3,00 | 4 | 30 | 34 |
| | 4,00 | 5 | 104 | 109 |
| | 5,00 | 4 | 84 | 88 |
| Total | | 13 | 219 | 232 |

Izvor: Simulirani podaci.

Slika 2.34.

Ispitanici prema ocjeni iz srednje škole i konzumiranju droge



Izvor: Simulirani podaci.

Pomoću odgovarajućih strelica varijabla v1 - Grupirana ocjena u srednjoj školi prebacuje se u polje **Row(s)**, a druga varijabla v2 - Uživete li drogu u polje **Column(s)**. Da bi se u **Outputu** dobio i grafikon višestrukih stupaca pri kreiranju dvostruke statističke tablice u prozoru Crosstabs, potrebno je aktivirati **Display clustered bar charts**.

Klikom na **OK** u **Outputu** se za rezultat dobije tražena dvostruka tablica koja je prikazana tablicom 2.26. i grafikon prikazan na slici 2.34.

b) Prema podacima u tablici 2.26 i slici 2.34 može se zaključiti da većina ispitanika u srednjoj školi ima prosječnu ocjenu 4. S obzirom na ukupan broj ispitanika koji su u srednjoj školi imali prosječnu ocjenu 3, broj onih koji između njih konzumiraju drogu je 4, a to je relativno najveći udio (drugi dvostruki stupac grafikona na slici 2.34).



Analitički primjer 2.29.

Zadatak je formirati dvostruku statističku tablicu gdje se varijabla (obilježje) Uživete li drogu (v2) mijenja po redcima (row), a varijabla (obilježje) Spol (v1) po stupcima (column). Potrebno je:

- Formirati tablicu sa zbirnim redcima i stupcima te nacrtati grafikon višestrukih stupaca!
- Komentirati sliku i izračunate frekvencije!



Rješenje 2.29:

- Na glavnom izborniku potrebno je izabrati ikonu *Analyze*, a na njezinu padajućem izborniku *Descriptive Statistics* i *Crosstabs*.

Pomoću odgovarajućih strelica varijabla v2 - Uživete li drogu premjestiti u polje *Row(s)*, a drugu varijablu v1 - Spol u polje *Column(s)*. Da bi se u *Outputu* dobio i grafikon višestrukih stupaca pri kreiranju dvostruke statističke tablice u prozoru *Crosstabs*, potrebno je aktivirati *Display clustered bar charts*.

Tablica 2.27.

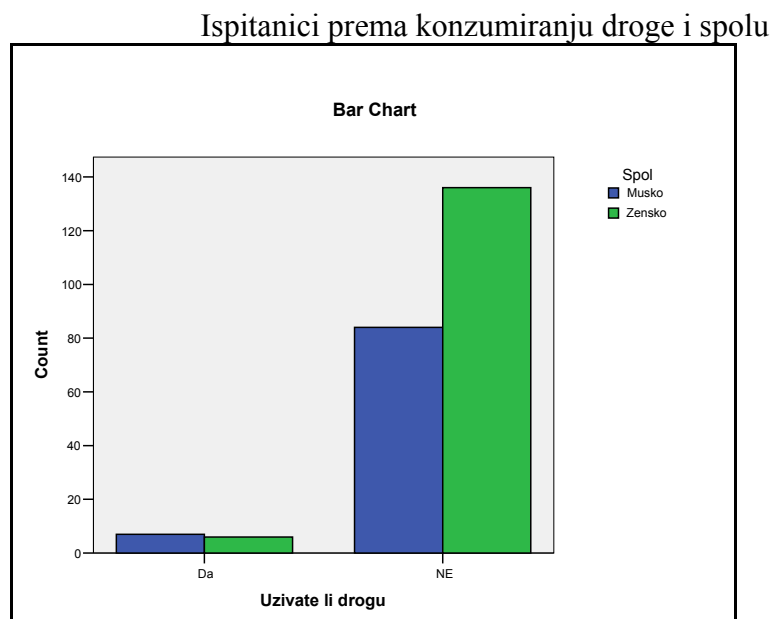
Ispitanici prema konzumiranju droge i spolu
Uživete li drogu * Spol Crosstabulation

| Count | | Spol | | Total |
|----------|----|-------|--------|-------|
| | | Musko | Zensko | |
| Uživete | Da | 7 | 6 | 13 |
| li drogu | NE | 84 | 136 | 220 |
| Total | | 91 | 142 | 233 |

Izvor: Simulirani podaci.

Klikom na **OK** u **Outputu** se za rezultat dobije tražena dvostruka tablica koja je prikazana tablicom 2.27 i grafikom prikazan na slici 2.35.

Slika 2.35.



Izvor: Simulirani podaci.

b) Prema podacima u tablici 2.27 i slici 2.35 može se zaključiti da je većina ispitanika u odabranom uzorku ženskog spola. Prvi dvostruki stupac grafikona odnosi se na one ispitanike koji konzumiraju drogu. Može se vidjeti da među njima prevladavaju oni muškog spola (i prema prvom retku tablice 2.27).



Analitički primjer 2.30.

Zadatak je formirati dvostruku statističku tablicu, gdje se varijabla (obilježje) Pušite li cigarete (v1) mijenja po redcima (row), a varijabla (obilježje) Grupirana ocjena u srednjoj školi (v2) po stupcima (column). Potrebno je:

- a) Formirati tablicu sa zbirnim redcima i stupcima te nacrtati grafikon višestrukih stupaca!
- b) Komentirati sliku i izračunate frekvencije!



Rješenje 2.30:

- a) Na glavnom izborniku potrebno je izabrati ikonu *Analyze*, a na njezinu padajućem izborniku *Descriptive Statistics* i *Crosstabs*.

Pomoću odgovarajućih strelica varijabla v1 - Pušite li cigarete prebacuje se u polje *Row(s)*, a druga varijabla v2 - Grupirana ocjena u srednjoj školi u polje *Column(s)*. Da bi se u *Outputu* dobio i grafikon višestrukih stupaca pri kreiranju dvostruke statističke tablice u prozoru Crosstabs, potrebno je aktivirati *Display clustered bar charts*.

Tablica 2.28.

Ispitanici prema pušenju cigareta i ocjeni iz srednje škole
Pušte li cigarete * Grupirana ocjena u srednjoj školi Crosstabulation

Count

| | | Grupirana ocjena u srednjoj školi | | | | Total |
|-------------------|-------------------------|-----------------------------------|------|------|------|-------|
| | | 2,00 | 3,00 | 4,00 | 5,00 | |
| Pušte li cigarete | Ne | 0 | 18 | 83 | 72 | 173 |
| | Do 10 cigareta na dan | 0 | 4 | 14 | 7 | 25 |
| | Do 1 kutije na dan | 1 | 11 | 8 | 9 | 29 |
| | Vise od 1 kutije na dan | 0 | 1 | 7 | 2 | 10 |
| Total | | 1 | 34 | 112 | 90 | 237 |

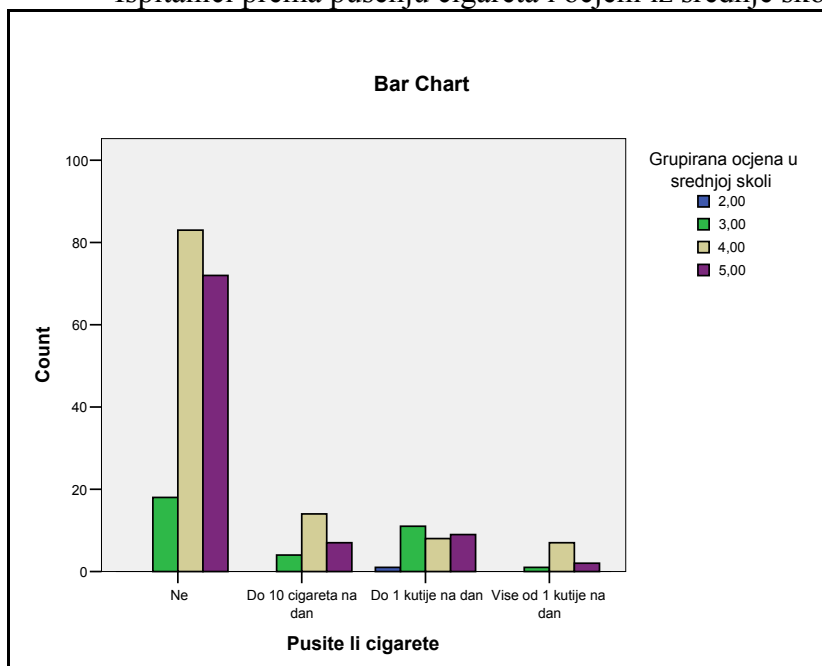
Izvor: Simulirani podaci.

Klikom na **OK** u *Outputu* se za rezultat dobije tražena dvostruka tablica koja je prikazana tablicom 2.28 i grafikon prikazan na slici 2.36.

- b) Prema podacima u tablici 2.28 i slici 2.36 može se zaključiti da većina ispitanika ne puši cigarete. U svim kategorijama pušača cigareta najveći je broj onih s prosječnom ocjenom iz srednje škole vrlo dobar, tj. 4, osim kod onih koji puše do 1 kutije na dan, gdje prevladavaju ispitanici s prosječnom ocjenom iz srednje škole dobar, tj. 3 (treći višestruki stupac grafikona na slici 2.36).

Slika 2.36.

Ispitanici prema pušenju cigareta i ocjeni iz srednje škole

*Izvor: Simulirani podaci.*

3. SREDNJE VRIJEDNOSTI

3.1. Srednje vrijednosti

Prikupljeni statistički podaci u svom izvornom obliku često zbog svog obujma nemaju razumljivu formu. Zbog toga se i vrši njihovo grupiranje, odnosno formiranje statističkih nizova. Na taj se način dobiva detaljniji uvid u svojstva promatranoga statističkog niza.

Računanjem srednjih vrijednosti dolazi se do informacija o vrijednostima statističkog obilježja oko kojih se raspoređuju elementi statističkog niza.

Srednja vrijednost je vrijednost statističkog obilježja oko koje se grupiraju podaci statističkog niza. Još se zove i "mjera centralne tendencije".

Srednje vrijednosti mogu se podijeliti: na

- 1. Položajne srednje vrijednosti** određuju se položajem podataka u nizu. Najvažnije položajne srednje vrijednosti su :
 - a) mod**
 - b) medijan.**
- 2. Potpune srednje vrijednosti** računaju se upotrebom svih podataka u statističkom nizu. Potpune srednje vrijednost su:
 - a) aritmetička sredina**
 - b) harmonijska sredina**
 - c) geometrijska sredina** (koja nema svoju primjenu u sociološkim istraživanjima).

3.2. Aritmetička sredina

Aritmetička sredina spada u potpune srednje vrijednosti i računa se upotrebom svih podataka u statističkom nizu. To je najvažnija i najčešće korištena srednja vrijednost (ali ne i uvijek najbolja srednja vrijednost).

Aritmetička sredina je omjer zbroja svih vrijednosti numeričkog obilježja jednog niza i broja elemenata tog niza.

Ako se radi o negrupiranom statističkom numeričkom nizu, računa se jednostavna aritmetička sredina:

$$\bar{X} = \frac{\sum_{i=1}^N X_i}{N} = \frac{X_1 + X_2 + \dots + X_N}{N} \quad (3.1)$$

Ako se radi o grupiranom statističkom numeričkom nizu, računa se složena, vagana ili ponderirana aritmetička sredina.

Poznato je da apsolutne frekvencije označavaju broj elemenata skupa koji pripadaju određenom modalitetu promatranoga statističkog obilježja. Iz toga se može zaključiti da je važnost svake vrijednosti obilježja ovisna o frekvenciji jer veća frekvencija znači veći udio tog obilježja u promatranom skupu. Zato se frekvencija još naziva težinski faktor ili ponder kod grupiranih nizova.

Vagana ili ponderirana aritmetička sredina računa se prema izrazu:

$$\bar{X} = \frac{\sum_{i=1}^k f_i x_i}{\sum_{i=1}^k f_i} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_k x_k}{f_1 + f_2 + \dots + f_k} \quad (3.2)$$

U jednadžbi (3.2) k predstavlja broj grupa, odnosno modaliteta promatranog numeričkog obilježja. Nazivnik ovog izraza, kao suma svih apsolutnih frekvencija, odgovara opsegu statističkog skupa.

Ako su veličine razreda grupiranog numeričkog niza različite od 1, za računanje vagane ili ponderirane aritmetičke sredine

potrebno je izvršiti aproksimaciju pomoću sredine razreda. Sredina razreda, kako je već rečeno, dobije se kao jednostavan prosjek donje i gornje granice razreda. Na taj način ona predstavlja sve vrijednosti obilježja koje se javljaju u jednoj grupi. Stoga je, u ovom slučaju, aritmetička sredina procjena stvarne aritmetičke sredine numeričkog niza.

Analitički primjer 3.1.

Statistički skup "Zaposlene žene u građevinarstvu u RH, 31. 03. 2002. godine" promatran prema **neprekidnom numeričkom obilježju** "dob" dan je u tablici 3.1..

Zadatak je **izračunati aritmetičku sredinu.**

Tablica 3.1.

Zaposlene žene u građevinarstvu u RH, 31. 03. 2004. godine

| Dob x_i | Broj žena f_i | Sredina razreda x_i | $f_i x_i$ |
|----------------|--------------------|-----------------------------|-----------|
| (18)-25 | 712 | 21,5 | 15308 |
| 25-30 | 1367 | 27,5 | 37592,5 |
| 30-35 | 1277 | 32,5 | 41502,5 |
| 35-40 | 1417 | 37,5 | 53137,5 |
| 40-45 | 1580 | 42,5 | 67150 |
| 45-50 | 1557 | 47,5 | 73957,5 |
| 50-65 | 1322 | 57,5 | 76015 |
| Ukupno: | 9232 | - | 364663 |

Izvor: Statistički ljetopis Republike Hrvatske, 2005. godine, str. 145.

Rješenje 3.1:

Zadani niz je grupiran prema modalitetima obilježja. Razredi nisu jednake veličine, a ima ih 7.

Kod grupiranih nizova s veličinom razreda različitom od 1, u izrazu za aritmetičku sredinu, x_i predstavlja sredinu i -tog razreda. Stoga je aproksimacija aritmetičke sredine sljedeća:

$$\bar{X} = \frac{\sum_{i=1}^7 f_i x_i}{\sum_{i=1}^7 f_i} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_7 x_7}{f_1 + f_2 + \dots + f_7} = \frac{364663}{9232} = 39,5 \text{ godina}$$

Dakle približno, prosječna je dob žena zaposlenih u građevinarstvu RH (stanje 31. 03. 2004. god.) 39,5 godina.

Aritmetička sredina izražava se u originalnim jedinicama mjere promatranoga numeričkog obilježja, obuhvaća sve elemente nekog skupa, te se pomoću nje mogu uspoređivati nizovi koji su grupirani po jednakom obilježju. Uz ove, aritmetička sredina zadovoljava i sljedeće kriterije:

- a) Aritmetička sredina nalazi su između najveće i najmanje vrijednosti promatranoga numeričkog obilježja:

$$x_{\min} \leq \bar{X} \leq x_{\max} \quad (3.3)$$

- b) Zbroj odstupanja vrijednosti numeričkog obilježja od aritmetičke sredine u jednoj distribuciji je uvijek nula.

$$\sum_{i=1}^N (x_i - \bar{X}) = 0, \text{ za negrupirani niz,} \quad (3.4)$$

$$\sum_{i=1}^k f_i (x_i - \bar{X}) = 0, \text{ za grupirani niz.} \quad (3.5)$$

- c) Zbroj kvadrata odstupanja vrijednosti numeričkog obilježja od aritmetičke sredine u jednoj distribuciji je manji ili jednak zbroju kvadrata odstupanja vrijednosti numeričkog obilježja u istoj distribuciji od bilo koje druge vrijednosti (a).

$$\sum_{i=1}^N (x_i - \bar{X})^2 \leq \sum_{i=1}^N (x_i - a)^2, \text{ za negrupirani niz,} \quad (3.6)$$

$$\sum_{i=1}^k f_i (x_i - \bar{X})^2 \leq \sum_{i=1}^k f_i (x_i - a)^2, \text{ za grupirani niz.} \quad (3.7)$$

Ako neki numerički niz sadrži ekstremno male ili velike vrijednosti obilježja, aritmetička sredina kao prosječna vrijednost gubi na svojoj reprezentativnosti. Taj problem je dodatno izražen kada u distribuciji postoje razredi s otvorenom donjom, odnosno gornjom granicom razreda, i kada nije moguće te granice objektivno procijeniti.

3.3. Mod

Mod je vrijednost statističkog obilježja koja se najčešće javlja u nekom nizu, tj. vrijednost obilježja kojoj pripada najveća frekvencija.

Mod se može primijeniti na kvalitativne i kvantitativne statističke nizove, a spada u položajne srednje vrijednosti.

Kod nominalnih obilježja mod se određuje brojenjem na način da se traži koja se vrijednost obilježja u nizu najčešće javlja. Ako je niz grupiran, traži se najveća apsolutna frekvencija. Vrijednost obilježja kojoj pripada ta najveća apsolutna frekvencija je mod.

Analitički primjer 3.2.

Statistički skup "prodane cipele u prodavaonici "K" na dan 05. 05. 2005. god." grupiran je prema **nominalnom obilježju** "boja". Modaliteti toga obilježja su: crna (C), bijela (B), ružičasta (R), smeđa (S).

Uređeni podaci dobiveni evidencijom na određeni dan su:

C C C C C B R R R S

Zadatak je **odrediti mod**.

Rješenje 3.2:

Dakle, na dan 05. 05. 2005. god. u prodavaonici "K" je prodano 10 cipela. Od toga 5 crnih, 1 bijele, 3 ružičaste i 1 smeđe. Vrijednost obilježja "boja" koja se u ovom nizu najčešće javlja je crna, pa se može zaključiti da je mod, $Mo = crna\ boja$.

Statistički niz grupiran u razrede prikazan je u tablici 3.2.

Kod grupiranih nizova mod se određuje prema najvećoj apsolutnoj frekvenciji: $f_{\max} = 5$. Modalitet obilježja koji odgovara toj frekvenciji je *crno*, što znači da je: $Mo = crna\ boja$.

Tablica 3.2.

Prodane cipele u prodavaonici "K" na dan 05. 05. 2005. god. prema boji

| Boja cipela | Broj prodanih cipela |
|----------------|----------------------|
| X_i | f_i |
| crna | 5 |
| bijela | 1 |
| ružičasta | 3 |
| smeđa | 1 |
| Ukupno: | 10 |

Izvor: Evidencija prodavaonice "K", svibanj 2005. godine

Ovo je primjer tzv. **unimodalnih nizova** koji imaju samo jedan mod. Postoje statistički nizovi u kojima se dvije ili više vrijednosti obilježja mogu pojavljivati češće u odnosu na ostale modalitete obilježja. U tom slučaju kaže se su to **bimodalni ili multimodalni nizovi**. Kod bimodalne distribucije, koja ima dva vrha, postoji glavni mod i lokalni mod. Kada je u nizu prisutno više od jednog moda, potrebno je statistički skup podijeliti na više podskupova, od kojih će svaki imati svoja karakteristična svojstva, te izvršiti analizu svakoga podskupa posebno.

**Analitički primjer 3.3.**

Za varijablu (obilježje) visina u cm (v_2) potrebno je odrediti:

- a) Najmanju i najveću vrijednost obilježja;
- b) Aritmetičku sredinu;
- c) Mod;
- d) Komentirati rezultate!

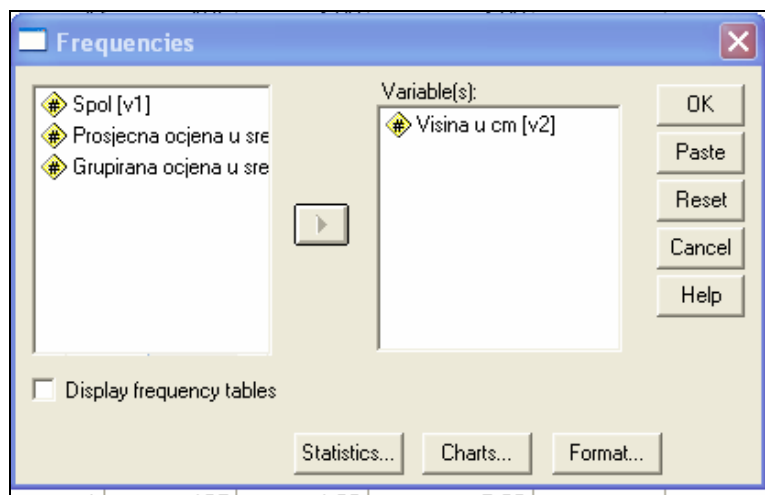
**Rješenje 3.3:**

Na glavnom izborniku potrebno je izabrati ikonu *Analyze*, a na njezinu padajućem izborniku *Descriptive Statistics* i *Frequencies*.

U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v2 - visina u cm se prebacuje u polje **Variable(s)**, kako je prikazano na slici 3.1.

Slika 3.1.

Prozor "Frequencies" sa izabranom varijablom (obilježjem) visina u cm (v2)



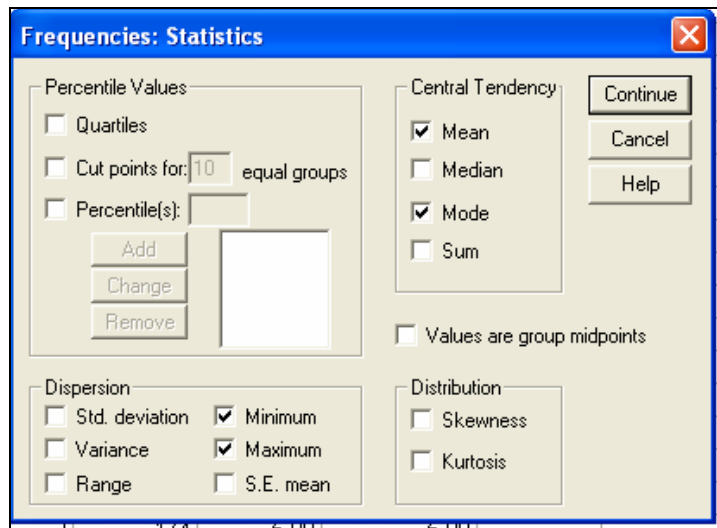
Izvor: Simulirani podaci.

Klikom na ikonu **Statistics** otvara se novi prozor **Frequencies Statistics**, gdje je potrebno aktivirati određene statističke veličine prema zahtjevu zadatka, a koje se odnose na odabranu varijablu visina u cm (v2).

Da bi se u izlaznim rezultatima dobila najmanja i najveća vrijednost niza, u spomenutom prozoru se aktivira **Minimum** i **Maximum**. Aritmetička sredina dobit će se aktiviranjem veličine **Mean**, a za mod potrebno je označiti **Mode**. Prozor **Frequencies Statistics** s potrebnim aktiviranim veličinama prikazan je na slici 3.2.

Slika 3.2.

Prozor "Frequencies Statistics" sa izabranim odgovarajućim statističkim veličinama



Izvor: Simulirani podaci.

Da bi se u **Outputu** programa SPSS dobili željeni podaci o visini ispitanika, potrebno je kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 3.3.

Tablica 3.3.

Statistički podaci o visini ispitanika
Statistics

Visina u cm

| | | |
|---------|---------|--------|
| N | Valid | 235 |
| | Missing | 4 |
| Mean | | 175,77 |
| Mode | | 172 |
| Minimum | | 154 |
| Maximum | | 206 |

Izvor: Simulirani podaci.

Prema podacima iz tablice 3.3. može se vidjeti da je 325 ispitanika dalo odgovor o svojoj visini, dok ih 4 nije odgovorilo na to pitanje.

Prosječna visina ispitanika u odabranom uzorku (od onih koji su dali odgovor), odnosno aritmetička sredina je 175,77 cm.

Mod, odnosno najčešća visina ispitanika je 172 cm.

Najmanja visina ispitanika u uzorku je 154 cm.

Najveća visina ispitanika u uzorku je 206 cm.



Analitički primjer 3.4.

Za varijablu (obilježje) prosječna ocjena u srednjoj školi (v3)³ potrebno je odrediti:

- a) Najmanju i najveću vrijednost obilježja;
- b) Aritmetičku sredinu;
- c) Mod;
- d) Komentirati rezultate!



Rješenje 3.4:

Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v3 - Prosječna ocjena u srednjoj školi se prebacuje u polje **Variable(s)**.

Klikom na ikonu **Statistics** otvara se novi prozor **Frequencies Statistics**, gdje je potrebno aktivirati određene statističke veličine prema zahtjevu zadatka, a koje se odnose na odabranu varijablu prosječna ocjena u srednjoj školi (v3).

Da bi se u izlaznim rezultatima dobila najmanja i najveća vrijednost niza, u spomenutom prozoru se aktivira **Minimum** i

³ Aritmetička sredina se po definiciji izračunava samo za numerička obilježja, ali se za cjelovitu analizu često primjenjuje i za redoslijedno obilježje.

Maximum. Aritmetička sredina dobit će se aktiviranjem veličine **Mean**, a za mod potrebno je označiti **Mode**.

Da bi se u **Outputu** programa SPSS dobili željeni podaci o prosječnoj ocjeni u srednjoj školi ispitanika, potrebno je kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 3.4.

Tablica 3.4.

Statistički podaci o prosječnoj ocjeni ispitanika u srednjoj školi

| Statistics | | |
|-----------------------------------|---------|--------|
| Prosječna ocjena u srednjoj školi | | |
| N | Valid | 238 |
| | Missing | 1 |
| Mean | | 4,1773 |
| Mode | | 4,00 |
| Minimum | | 2,00 |
| Maximum | | 5,00 |

Izvor: Simulirani podaci.

Prema podacima iz tablice 3.4. može se vidjeti da je 238 ispitanika dalo odgovor o svojoj prosječnoj ocjeni u srednjoj školi, dok 1 ispitanik nije odgovorio na to pitanje. Prosječna ocjena u srednjoj školi za sve ispitanike u odabranom uzorku (od onih koji su dali odgovor), odnosno aritmetička sredina je 4,177.

Mod, odnosno najčešća prosječna ocjena ispitanika u srednjoj školi je 4.

Najmanja prosječna ocjena ispitanika u srednjoj školi u uzorku je 2.

Najveća prosječna ocjena ispitanika u srednjoj školi u uzorku je 5.



Analitički primjer 3.5.

Za varijablu (obilježje) grupirana ocjena u srednjoj školi (v4) potrebno je odrediti:

- Tablicu frekvencija;
- Najmanju i najveću vrijednost obilježja;

- c) Aritmetičku sredinu;
- d) Mod;
- e) Komentirati rezultate!



Rješenje 3.5:

Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v4 - Grupirana ocjena u srednjoj školi se prebacuje u polje **Variable(s)**.

Klikom na ikonu **Statistics** otvara se novi prozor **Frequencies Statistics**, gdje je potrebno aktivirati određene statističke veličine prema zahtjevu zadatka, a koje se odnose na odabranu varijablu grupirana ocjena u srednjoj školi (v4). S obzirom na to da se u ovom primjeru traži i tablica frekvencija, potrebno je aktivirati **Display Frequency Tables**.

Da bi se u izlaznim rezultatima dobila najmanja i najveća vrijednost niza u spomenutom prozoru se aktivira **Minimum** i **Maximum**. Aritmetička sredina dobit će se aktiviranjem veličine **Mean**, a za mod potrebno je označiti **Mode**.

Da bi se u **Outputu** programa SPSS dobili željeni podaci o grupiranoj ocjeni ispitanika u srednjoj školi, potrebno je kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 3.5.

Tablica 3.5.

Statistički podaci o grupiranoj ocjeni ispitanika u srednjoj školi

| Statistics | | |
|-----------------------------------|---------|--------|
| Grupirana ocjena u srednjoj školi | | |
| N | Valid | 238 |
| | Missing | 1 |
| Mean | | 4,2269 |
| Mode | | 4,00 |
| Minimum | | 2,00 |
| Maximum | | 5,00 |

Izvor: Simulirani podaci.

Prema podacima iz tablice 3.5. može se vidjeti da je 238 ispitanika dalo odgovor o svojoj prosječnoj ocjeni u srednjoj školi, dok 1 ispitanik nije odgovorio na to pitanje. Prosječna ocjena u srednjoj školi za sve ispitanike u odabranom uzorku (od onih koji su dali odgovor), odnosno aritmetička sredina je 4,23.

Mod, odnosno najčešća prosječna ocjena u srednjoj školi ispitanika je 4.

Najmanja prosječna ocjena ispitanika u srednjoj školi u uzorku je 2.

Najveća prosječna ocjena ispitanika u srednjoj školi u uzorku je 5.

U Outputu je dana i tablica frekvencija prema grupiranoj ocjeni u srednjoj školi. Prikazana je u tablici 3.6.

Tablica 3.6.

| Grupirana ocjena u srednjoj školi | | | | | |
|-----------------------------------|--------|-----------|---------|---------------|--------------------|
| | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 2,00 | 1 | ,4 | ,4 | ,4 |
| | 3,00 | 34 | 14,2 | 14,3 | 14,7 |
| | 4,00 | 113 | 47,3 | 47,5 | 62,2 |
| | 5,00 | 90 | 37,7 | 37,8 | 100,0 |
| | Total | 238 | 99,6 | 100,0 | |
| Missing | System | 1 | ,4 | | |
| Total | | 239 | 100,0 | | |

Izvor: Simulirani podaci.

U stupcu Frequency može se vidjeti da je od 238 ispitanika koji su dali odgovor na postavljeno pitanje ocjenu 2 imao samo 1, ocjenu 3 njih 34, ocjenu 4 njih 113, dok je ocjenu 5 imalo 90 ispitanika.



Analitički primjer 3.6.

Potrebno je dokazati 3 svojstva aritmetičke sredine za anketirane ispitanike prema varijabli (obilježju) prosječna ocjena u srednjoj školi (v2) .

- a) Provjeriti 1. svojstvo da je: $x_{\min} \leq \bar{X} \leq x_{\max}$.
- b) Provjeriti 2. svojstvo da je: $\sum_{i=1}^N (x_i - \bar{X}) = 0$, (za negrupirani niz).
- c) Provjeriti 3. svojstvo da je: $\sum_{i=1}^N (x_i - \bar{X})^2 \leq \sum_{i=1}^N (x_i - a)^2$, (za negrupirani niz) ako je npr. $a = 3$.



Rješenje 3.6:

- a) Da bi se dokazalo 1. svojstvo da se aritmetička sredina nalazi između najveće i najmanje vrijednosti obilježja, treba izračunati \bar{X}, x_{\min} i x_{\max} .

Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v2 - Prosječna ocjena u srednjoj školi se prebacuje u polje **Variable(s)**.

Klikom na ikonu **Statistics** otvara se novi prozor **Frequencies Statistics**, gdje je potrebno aktivirati **Minimum**, **Maximum** i **Mean**.

Dalje je potrebno kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 3.7.

Tablica 3.7.

| Statistics | | |
|-----------------------------------|---------|--------|
| Prosječna ocjena u srednjoj školi | | |
| N | Valid | 238 |
| | Missing | 1 |
| Mean | | 4,1773 |
| Minimum | | 2,00 |
| Maximum | | 5,00 |

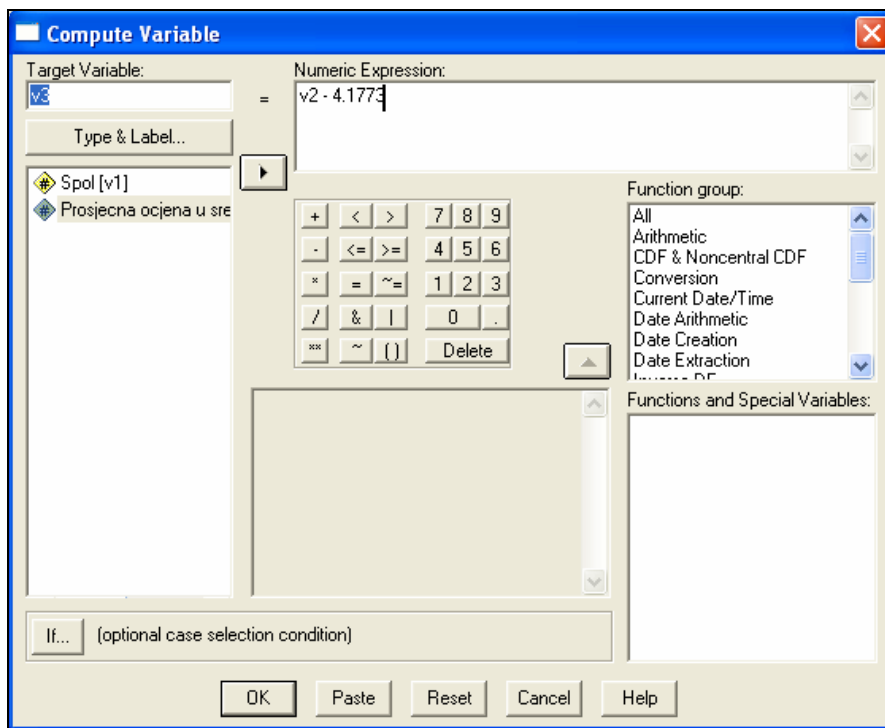
Izvor: Simulirani podaci.

Iz podataka u tablici 3.7. vidi se da je aritmetička sredina 4,1773 i vrijedi da je između najmanje (2) i najveće (5) vrijednosti obilježja: $2 \leq 4,1773 \leq 5$.

b) Da bi se dokazalo 2. svojstvo aritmetičke sredine da je zbroj odstupanja vrijednosti obilježja od prosjeka jednak 0, potrebno je formirati novi stupac (varijablu) gdje će se izračunati ta odstupanja. Zbroj svih podataka iz tog novog stupca mora biti 0. Na glavnom izborniku potrebno je odabrati **Transform**, a na njegovu padajućem izborniku se bira **Compute**. Na taj način se otvori prozor **Compute Variable** koji je prikazan na slici 3.3.

Slika 3.3.

Prozor "Compute Variable" gdje se računaju odstupanja od prosjeka kao nova varijabla v3



Izvor: Simulirani podaci.

Pod **Target Variable** upisuje se novo ime v3, a u **Numeric Expression** se upisuje: $v2 - 4,773$, odnosno da se od svih vrijednosti varijable v2 oduzme njihov prosjek 4,773. Potrebno je napomenuti da se **decimalni zarez** u programu **SPSS** označava **točkom**. Klikom na **OK** u prostoru **Data View** formira se nova varijabla v3.

Da bi se vidio zbroj podataka u stupcu na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Descriptives**. Varijabla v3 prebacuje se u **Variable(s)**. U **Options** potrebno je aktivirati **Sum** (dok se druge veličine mogu deaktivirati jer u ovom slučaju nisu potrebne).

Tablica 3.8.

Zbroj odstupanja vrijednosti obilježja od prosjeka

| Descriptive Statistics | | |
|------------------------|-----|-----|
| | N | Sum |
| v3 | 238 | ,00 |
| Valid N (listwise) | 238 | |

Izvor: Simulirani podaci.

Suma vrijednosti novoformirane varijable v3 u zadnjem stupcu tablice 3.8 je 0, čime je dokazano 2. svojstvo aritmetičke sredine.

c) Da bi se dokazalo 3. svojstvo aritmetičke sredine da je zbroj kvadrata odstupanja vrijednosti obilježja od prosjeka manje ili jednako zbroju kvadrata odstupanja vrijednosti obilježja od bilo kojeg drugog broja, npr. 3, potrebno je formirati nove stupce gdje će se izračunati ti kvadrati odstupanja. Usporedbom zbroja tih stupaca dokazat će se svojstvo 3. Na glavnom izborniku potrebno je odabrati **Transform**, a na njegovu padajućem izborniku bira se **Compute**. Na taj način se otvori prozor **Compute Variable**.

Pod **Target Variable** upisuje se novo ime v4, a u **Numeric Expression** se upisuje: $(v2 - 4,773)^2$, odnosno da se od svih vrijednosti varijable v2 oduzme njihov prosjek 4,773 i sve se kvadrira. Potrebno je napomenuti da se **potencija** u programu SPSS **označava dvjema zvjezdicama**. Klikom na **OK** u prostoru Data View formira se nova varijabla v4.

Zatim se ponovo u **Target Variable** upisuje novo ime v5, a u **Numeric Expression** se upisuje: $(v2 - 3)^2$, odnosno od svih se vrijednosti varijable v2 oduzme 3 i sve se kvadrira. Klikom na **OK** u prostoru **Data View** formira se nova varijabla v5.

Da bi se vidio zbroj podataka u ova dva nova stupca v4 i v5 na glavnom izborniku, potrebno je izabrati ikonu **Analyze**, a na njezinu

padajućem izborniku *Descriptive Statistics* i *Descriptives*. Varijable v4 i v5 prebacuju se u *Variable(s)*. U *Options* potrebno je aktivirati *Sum* (dok se druge veličine mogu deaktivirati jer u ovom slučaju nisu potrebne).

Tablica 3.9.

Zbroj kvadrata odstupanja vrijednosti obilježja od prosjeka i od vrijednosti 3

| Descriptive Statistics | | |
|------------------------|-----|--------|
| | N | Sum |
| v4 | 238 | 103,84 |
| v5 | 238 | 433,72 |
| Valid N (listwise) | 238 | |

Izvor: Simulirani podaci.

Sume vrijednosti novoformiranih varijabli v4 i v5 prikazane su u zadnjem stupcu tablice 3.9, čime je dokazano 3. svojstvo aritmetičke sredine, tj. da je: $\sum_{i=1}^N (x_i - \bar{X})^2 \leq \sum_{i=1}^N (x_i - a)^2$, odnosno da je: $103,84 \leq 433,72$.

3.6. Harmonijska sredina

Harmonijska sredina spada u potpune srednje vrijednosti i računa se upotrebom svih podataka u statističkom nizu.

Harmonijska sredina je recipročna vrijednost aritmetičke sredine recipročnih vrijednosti numeričkog obilježja u jednom nizu.

Ako se radi o **negrupiranom statističkom numeričkom nizu**, računa se **jednostavna harmonijska sredina**:

$$H = \frac{N}{\sum_{i=1}^N \frac{1}{x_i}} \quad (3.7)$$

Ako se radi o **grupiranom statističkom numeričkom nizu**, računa se **složena ili vagana harmonijska sredina**:

$$H = \frac{\sum_{i=1}^k f_i}{\sum_{i=1}^k \frac{f_i}{x_i}} \quad (3.8)$$

Ponderi su i ovdje apsolutne frekvencije. Ako su veličine razreda različite od 1, i za harmonijsku sredinu se radi aproksimacija na način da se računaju sredine razreda x_i .

Harmonijska sredina rijetko se primjenjuje u praksi. Specifično područje njezine primjene je izračunavanje srednjih vrijednosti relativnih brojeva koordinacije.

Relativni brojevi koordinacije (R_i), kako je već rečeno, pokazuju odnos dviju pojava koje ima smisla uspoređivati (A_i i B_i).

$$R_i = \frac{A_i}{B_i} \quad (3.9)$$

Općenito srednja vrijednost relativnih brojeva koordinacije je omjer zbroja svih vrijednosti veličine koja se uspoređuje ($\sum_{i=1}^k A_i$) i zbroja svih vrijednosti veličine s kojom se uspoređuje veličina iz brojnika ($\sum_{i=1}^k B_i$):

$$\bar{R} = \frac{\sum_{i=1}^k A_i}{\sum_{i=1}^k B_i} \quad (3.10)$$

Katkada u konkretnom statističkom istraživanju **nisu dostupni podaci o veličinama sadržanim u relativnom broju koordinacije, pa su pri izračunavanju njegove srednje vrijednosti moguće 3 situacije:**

- a) Ako su dostupni podaci o relativnom broju koordinacije (R_i) i o veličini s kojom se uspoređuje veličina iz brojnika (B_i), srednja će se vrijednost (\bar{R}) izračunati pomoću izraza za **složenu ili vaganu aritmetičku sredinu:**

$$\bar{R} = \frac{\sum_{i=1}^k R_i B_i}{\sum_{i=1}^k B_i} \quad (3.11)$$

Na temelju osnovne relacije (3.10) jasno je da vrijedi da je veličina iz brojnika: $\sum_{i=1}^k A_i = \sum_{i=1}^k R_i B_i$. Ponderi su ovdje vrijednosti veličine s kojom se uspoređuje veličina iz brojnika: B_i .

- b) Ako su dostupni podaci o relativnom broju koordinacije (R_i), bez informacija o veličinama u brojniku (A_i) i nazivniku (B_i), srednju vrijednost \bar{R} nije moguće izračunati. Pokušaj aproksimacije ove srednje vrijednosti ne daje vjerodostojan rezultat.
- c) Ako su dostupni podaci o relativnom broju koordinacije (R_i) i o veličini koja se uspoređuje (A_i), srednja će se vrijednost (\bar{R}) izračunati pomoću izraza za **složenu ili vaganu harmonijsku sredinu**:

$$\bar{R} = \frac{\sum_{i=1}^k A_i}{\sum_{i=1}^k \frac{A_i}{R_i}} \quad (3.12)$$

Na temelju osnovne relacije (3.10) jasno je da vrijedi da je veličina iz nazivnika: $\sum_{i=1}^k B_i = \sum_{i=1}^k A_i / R_i$. Ponderi su ovdje vrijednosti veličine koja se uspoređuje iz brojnika: A_i .

Analitički primjer 3.7.

Za nekoliko odabranih fakulteta u RH dani su podaci o broju studenata i broju studenata po 1 nastavniku u akademskoj godini 2004/05.

Zadatak je izračunati prosječan broj studenata po 1 nastavniku za sve promatrane fakultete zajedno, odnosno **prosjeak relativnih brojeva koordinacije** (\bar{R}).

Tablica 3.10.

Broj studenata i broj studenata po 1 nastavniku u nekim odabranim fakultetima u RH u akademskoj godini 2004/05.

| Fakulteti | Broj studenata (A_i) | Broj studenata po 1 nastavniku (R_i) | Broj nastavnika ($A_i/R_i=B_i$) |
|----------------------------|--------------------------|--|-----------------------------------|
| PMF | 3562 | 8,62469734 | 413 |
| Građevinski fakulteti | 2761 | 9,17275748 | 301 |
| Fakult. el. stroj. i brod. | 7704 | 10,1102362 | 762 |
| Medicinski fakulteti | 3315 | 3,64285714 | 910 |
| Ekonomski fakulteti | 14247 | 33,0556845 | 431 |
| Ukupno: | 31589 | 64,6062326 | 2817 |

Izvor: Statistički ljetopis Republike Hrvatske, 2005. godina, str. 492, 495, 496, 497.

Rješenje 3.7:

S obzirom na to da je u navedenom primjeru poznata (A_i) veličina koja se uspoređuje i relativni brojevi koordinacije (R_i), njihov se prosjek računa pomoću izraza za složenu ili vaganu harmonijsku sredinu gdje su ponderi (A_i):

$$H = \frac{\sum_{i=1}^k f_i}{\sum_{i=1}^k \frac{f_i}{x_i}} \Rightarrow \bar{R} = \frac{\sum_{i=1}^k A_i}{\sum_{i=1}^k \frac{A_i}{R_i}} = \frac{31589}{2817} = 11,214.$$

Prosječan broj studenata po jednom nastavniku za sve promatrane fakultete u akademskoj godini 2004/05. zajedno iznosi 11,214 studenata po nastavniku.

3.5. Medijan

Medijan je vrijednost statističkog obilježja koja uređeni statistički niz dijeli na dva jednakobrojna dijela.

Medijan se može primijeniti na redoslijedne i kvantitativne statističke nizove, a spada u položajne srednje vrijednosti. Medijan se ne primjenjuje kod nominalnih nizova jer poredak oblika ovog obilježja može biti proizvoljan.

Kod **negrupiranog, a uređenog niza (po veličini vrijednosti obilježja)**, medijan je vrijednost obilježja koja pripada elementu statističkog niza koji se nalazi u sredini niza. Ako je broj elemenata statističkog niza paran, onda se za medijan uzima jednostavan prosjek vrijednosti obilježja dvaju članova koji se nalaze na sredini statističkog niza.

Ako je **statistički niz grupiran u razrede**, prije računanja medijana potrebno je izračunati frekvencije kumulativnog niza "manje od" ili "više od", te se u takvom nizu traži središnji član. **Ako se računaju frekvencije kumulativnog niza "manje od", traži se prva frekvencija niza koja sadrži $N/2$. Razred koji odgovara ovoj frekvenciji kumulativnog niza "manje od" je medijalni razred.**

Ako su veličine razreda obilježja jednake 1, vrijednost obilježja koja odgovara odabranoj frekvenciji kumulativnog niza je medijan.

Ako su veličine razreda obilježja različite od 1, medijan se računa po izrazu:

$$Me = L_1 + \frac{\frac{N}{2} - \sum_{i=1}^m f_i}{f_{med}} \cdot i, \quad (3.13)$$

gdje je:

L_1 - donja prava ili precizna granica medijalnog razreda

$N/2$ - polovica elemenata statističkog niza

$\sum_{i=1}^m f_i$ - zbroj svih apsolutnih frekvencija do medijalnog razreda, ne uključujući medijalni razred (tj. kumulativna frekvencija ispred kumulativne frekvencije medijalnog razreda)

f_{med} - apsolutna frekvencija medijalnog razreda

i - veličina medijalnog razreda.

Analitički primjer 3.8.

Statistički skup "Sklopljeni brakovi u RH u 2002. godini" prema **kontinuiranom numeričkom obilježju** "dob nevjeste", prikazan je u tablici 3.3.

Zadatak je **izračunati medijan**, odnosno dob nevjeste koja promatrani statistički niz dijeli na dva jednaka dijela.

Tablica 3.11.

Sklopljeni brakovi u RH u 2004. godini prema nevjestinoj dobi

| Nevjestina dob | Sklopljeni brakovi | Kumulativni niz "manje od" | Veličine razreda |
|----------------|--------------------|----------------------------|------------------|
| X_i | f_i | | i |
| 15-20 | 1903 | 1903 | 5 |
| 20-25 | 8104 | 10007 | 5 |
| 25-30 | 7335 | 17342 | 5 |
| 30-35 | 2820 | 20162 | 5 |
| 35-45 | 1516 | 21678 | 10 |
| 45-(80) | 894 | 22572 | 35 |
| Ukupno: | 22572 | - | - |

Izvor: Statistički ljetopis Republike Hrvatske, 2005. godina, str. 124.

Rješenje 3.8.

U tablici 3.11 vidi se da grupirani niz ima veličine razreda različite od 1. Stoga je za izračunavanje medijana potrebno odrediti frekvencije kumulativnog niza "manje od" tako da se postupno zbrajaju apsolutne frekvencije odozgo prema dolje. Traži se prva frekvencija kumulativnog niza koja u sebi sadrži: $N/2 = 22572/2 = 11286$ jedinicu. To je frekvencija 17342, a razred koji odgovara toj frekvenciji je medijalni razred. Njegova donja granica je $L_1 = 25$.

$$Me = L_1 + \frac{\frac{N}{2} - \sum_{i=1}^m f_i}{f_{med}} \cdot i = 25 + \frac{11286 - 10007}{7335} \cdot 5 = 25,87 \text{ godina}$$

Dakle, kod polovice sklopljenih brakova u Republici Hrvatskoj u 2002. godini nevjestica je bila mlađa od 25,87 godina, dok je kod preostale polovice brakova nevjestica bila starija od 25,87 godina.

U statističkim distribucijama postoji samo jedan medijan i on se nalazi između najveće i najmanje vrijednosti obilježja. Prednost medijana je da na njega ne utječu ekstremno male ili velike vrijednosti obilježja, pa je primjerena srednja vrijednost i kod izrazito asimetričnih distribucija.

3.5.1. Kvantili

Kvantili su vrijednosti statističkog obilježja koje uređeni statistički niz dijele na q jednakobrojnih dijelova.

Kod analize statističkih nizova vrlo su često u upotrebi kvantili.

Kvantili su vrijednosti statističkog obilježja koje statistički niz dijele na 4 jednakobrojna dijela. Kvantili se mogu podijeliti: na

- a) donji kvartil (Q_1) i na
- b) gornji kvartil (Q_3)⁴.

Donji kvartil dijeli uređeni statistički niz na četiri jednaka dijela u omjeru 1:3, odnosno 25% elemenata statističkog skupa ima vrijednost obilježja manju ili jednaku donjem kvartilu, a 75% elemenata statističkog skupa ima vrijednost obilježja veću od donjeg kvartila.

Gornji kvartil dijeli uređeni statistički niz na četiri jednaka dijela u omjeru 3:1, odnosno 75% elemenata statističkog skupa ima vrijednost obilježja manju ili jednaku donjem kvartilu, a 25% elemenata statističkog skupa ima vrijednost obilježja veću od gornjeg kvartila.

Kvantili se, slično kao i medijan, mogu primijeniti na redoslijedne i kvantitativne statističke nizove, a određuju se položajem u nizu. Ni oni se ne primjenjuju kod nominalnih nizova jer, kako je već naglašeno, poredak oblika ovog obilježja može biti proizvoljan.

⁴ Vrijedi da je medijan: $Me = Q_2$.

Kod **negrupiranog, a uređenog niza (po veličini vrijednosti obilježja)**, **donji kvartil** je vrijednost obilježja koja pripada elementu statističkog niza čiji rang (r), tj. mjesto u nizu odgovara:

$Q_1 = x_r$, gdje je $r = INT(N/4) + 1$, ako N nije djeljiv s 4;⁵

$Q_1 = \frac{x_r + x_{r+1}}{2}$, gdje je $r = N/4$, $(r+1) = N/4 + 1$, ako je N djeljivo

s 4, računa se prosjek obilježja koji se nalaze na mjestu u nizu: r i $(r+1)$.

Kod **negrupiranog, a uređenog niza (po veličini vrijednosti obilježja)**, **gornji kvartil** je vrijednost obilježja koja pripada elementu statističkog niza čiji rang, tj. mjesto u nizu odgovara:

$Q_3 = x_r$, gdje je $r = INT(3N/4) + 1$, ako N nije djeljiv s 4;

$Q_3 = \frac{x_r + x_{r+1}}{2}$, gdje je $r = 3N/4$, $(r+1) = 3N/4 + 1$, ako je N

djeljivo s 4, računa se prosjek obilježja koji se nalaze na mjestu u nizu: r i $(r+1)$.

Ako je **statistički niz grupiran u razrede**, prije računanja kvartila potrebno je izračunati frekvencije kumulativnog niza "manje od" ili "više od", te se u takvom nizu traži odgovarajući član. Ako se računaju frekvencije kumulativnog niza "manje od", za **donji kvartil** se traži prva frekvencija niza koja sadrži $N/4$. Za **gornji kvartil** se traži prva frekvencija niza koja sadrži $3N/4$. Razred koji odgovara odabranoj frekvenciji kumulativnog niza "manje od" je razred donjeg, odnosno gornjeg kvartila.

Ako su veličine razreda obilježja jednake 1, vrijednost obilježja koja odgovara odabranoj frekvenciji kumulativnog niza je donji, odnosno gornji kvartil.

Ako su veličine razreda obilježja različite od 1, donji kvartil se računa po izrazu:

⁵ INT predstavlja cijeli dio decimalnog broja dobivenog dijeljenjem (eng. integer). npr.

$INT(6,3) = 6$,

$INT(6,7) = 6$.

$$Q_1 = L_1 + \frac{\frac{N}{4} - \sum_{i=1}^q f_i}{f_{k \text{ var } t}} \cdot i, \quad (3.14)$$

gdje je:

L_1 - donja prava ili precizna granica donje kvartilnog razreda

$N/4$ - četvrtina elemenata statističkog niza

$\sum_{i=1}^q f_i$ - zbroj svih apsolutnih frekvencija do donje kvartilnog razreda, ne uključujući taj razred (tj. kumulativna frekvencija ispred kumulativne frekvencije donje kvartilnog razreda)

$f_{k \text{ var } t}$ - apsolutna frekvencija donje kvartilnog razreda

i - veličina kvartilnog razreda.

Ako su veličine razreda obilježja različite od 1, gornji kvartil se računa po izrazu:

$$Q_3 = L_1 + \frac{\frac{3N}{4} - \sum_{i=1}^q f_i}{f_{k \text{ var } t}} \cdot i, \quad (3.15)$$

gdje je:

L_1 - donja prava ili precizna granica gornje kvartilnog razreda

$3N/4$ - tri četvrtine elemenata statističkog niza

$\sum_{i=1}^q f_i$ - zbroj svih apsolutnih frekvencija do gornje kvartilnog razreda, ne uključujući taj razred (tj. kumulativna frekvencija ispred kumulativne frekvencije gornje kvartilnog razreda)

$f_{k \text{ var } t}$ - apsolutna frekvencija gornje kvartilnog razreda

i - veličina kvartilnog razreda.

Sve srednje vrijednosti se, uz prikazane izračune pomoću apsolutnih frekvencija, na isti način mogu računati korištenjem relativnih frekvencija (proporcija i postotaka).

Analitički primjer 3.9.

Statistički skup "sklopljeni brakovi u RH u 2002. godini" prema **kontinuiranom numeričkom obilježju** "nevjestina dob", prikazan je u tablici 3.12 (već je prezentiran u analitičkom primjeru 3.8).

Zadatak je **izračunati donji i gornji kvartil**.

Tablica 3.12.

Sklopljeni brakovi u RH u 2004. godini prema nevjestinoj dobi

| Nevjestina dob | Sklopljeni brakovi | Kumulativni niz "manje od" | Veličine razreda |
|----------------|--------------------|----------------------------|------------------|
| X_i | f_i | | i |
| 15-20 | 1903 | 1903 | 5 |
| 20-25 | 8104 | 10007 | 5 |
| 25-30 | 7335 | 17342 | 5 |
| 30-35 | 2820 | 20162 | 5 |
| 35-45 | 1516 | 21678 | 10 |
| 45-(80) | 894 | 22572 | 35 |
| Ukupno: | 22572 | - | - |

Izvor: Statistički ljetopis Republike Hrvatske, 2005. godina, str. 124.

Rješenje 3.9.

U tablici 3.12 vidi se da grupirani niz ima veličine razreda različite od 1. Stoga je za izračunavanje kvartila potrebno odrediti frekvencije kumulativnog niza "manje od" tako da se postupno zbrajaju apsolutne frekvencije odozgo prema dolje.

Za **donji kvartil** se traži prva frekvencija kumulativnog niza koja u sebi sadrži: $\frac{N}{4} = \frac{22572}{4} = 5643$ jedinice. To je frekvencija 10007, a razred koji odgovara toj frekvenciji je donje kvarilni razred. Njegova donja granica je $L_1 = 20$.

$$Q_1 = L_1 + \frac{\frac{N}{4} - \sum_{i=1}^q f_i}{f_{k \text{ var } t}} \cdot i = 20 + \frac{5643 - 1903}{8104} \cdot 5 = 22,31 \text{ godina}$$

Dakle, kod jedne četvrtine (tj. 25%) sklopljenih brakova u R. Hrvatskoj u 2002. godini nevjestica je bila mlađa od 22,31 godine, dok je

kod preostale tri četvrtine (tj. 75%) brakova nevjesta bila starija od 22,31 godine.

Za **gornji kvartil** se traži prva frekvencija kumulativnog niza koja u sebi sadrži: $\frac{3N}{4} = \frac{3 \cdot 22572}{4} = 16929$ jedinica. To je frekvencija 17342, a razred koji odgovara toj frekvenciji je gornje kvarilni razred. Njegova donja granica je $L_1 = 25$.

$$Q_3 = L_1 + \frac{\frac{3N}{4} - \sum_{i=1}^q f_i}{f_{k \text{ var } t}} \cdot i = 25 + \frac{16929 - 10007}{7335} \cdot 5 = 27,72 \text{ godina}$$

Dakle, kod tri četvrtine (tj. 75%) sklopljenih brakova u R. Hrvatskoj u 2002. godini nevjesta je bila mlađa od 29,48 godina, dok je kod preostale četvrtine (tj. 25%) brakova nevjesta bila starija od 29,48 godina.



Analitički primjer 3.10.

Za varijablu (obilježje) težina u kg (v2) potrebno je odrediti:

- a) Najmanju i najveću vrijednost obilježja;
- b) Aritmetičku sredinu;
- c) Mod;
- d) Medijan;
- e) Donji i gornji kvartil;
- f) Komentirati rezultate!



Rješenje 3.10:

Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

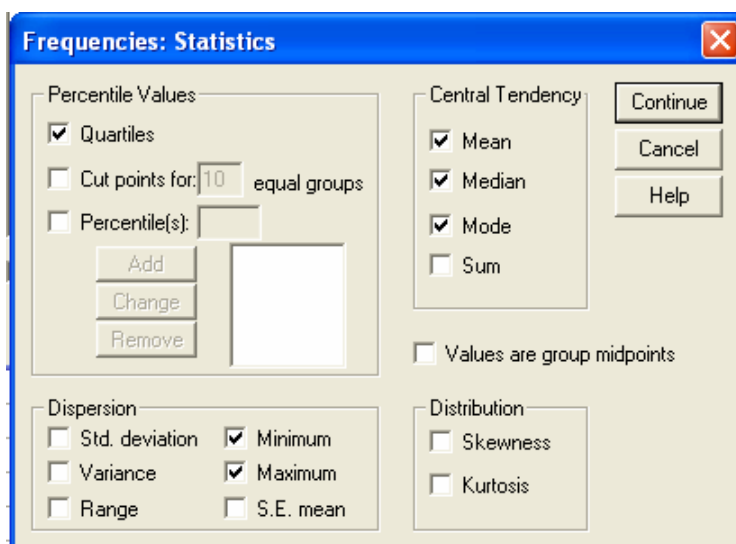
U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v2 - težina u kg se prebacuje u polje **Variable(s)**.

Klikom na ikonu **Statistics** otvara se novi prozor **Frequencies Statistics**, gdje je potrebno aktivirati određene statističke veličine prema zahtjevu zadatka, a koje se odnose na odabranu varijablu težina u kg (v2).

Da bi se u izlaznim rezultatima dobile najmanja i najveća vrijednost niza, u spomenutom se prozoru aktivira **Minimum** i **Maximum**. Aritmetička sredina dobit će se aktiviranjem veličine **Mean**, a za mod potrebno je označiti **Mode**. Medijan se dobije aktiviranjem **Median**, a donji i gornji kvartil označavanjem **Quartiles**. Prozor **Frequencies Statistics** s potrebnim aktiviranim veličinama prikazan je na slici 3.4.

Slika 3.4.

Prozor "Frequencies Statistics" sa izabranim odgovarajućim statističkim veličinama



Izvor: Simulirani podaci.

Da bi se u **Outputu** programa SPSS dobili željeni podaci o težini ispitanika, potrebno je kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 3.13. Prema tim podacima može se vidjeti da je 324 ispitanika dalo odgovor o svojoj težini, dok ih 5 nije odgovorilo na to pitanje. Prosječna težina ispitanika u odabranom uzorku (od onih koji su dali odgovor), odnosno aritmetička sredina je 67,18 kg.

Tablica 3.13.

Statistički podaci o težini ispitanika

Statistics

Težina u kg

| | | |
|-------------|---------|---------|
| N | Valid | 234 |
| | Missing | 5 |
| Mean | | 67,1795 |
| Median | | 64,0000 |
| Mode | | 52,00 |
| Minimum | | 46,00 |
| Maximum | | 110,00 |
| Percentiles | 25 | 56,0000 |
| | 50 | 64,0000 |
| | 75 | 75,0000 |

Izvor: Simulirani podaci.

Medijan, odnosno težina koja odabrani uzorak dijeli na 2 jednakobrojna dijela je $Me = 64\text{ kg}$, što znači da je polovica ispitanika imala težinu manju ili jednaku 64 kg, dok je preostala polovica ispitanika imala težinu veću od 64 kg.

Mod, odnosno najčešća težina ispitanika je 52 kg.

Najmanja težina ispitanika u uzorku je 46 kg.

Najveća težina ispitanika u uzorku je 110 kg.

Donji i gornji kvartil dani su među **Percentiles**. Donji kvaril $Q_1 = 56\text{ kg}$, što znači da je četvrtina ispitanika imala težinu manju ili jednaku 56 kg, dok je preostalih tri četvrtine ispitanika imalo težinu veću od 56 kg. Gornji kvaril $Q_3 = 75\text{ kg}$, što znači da je tri četvrtine ispitanika imalo težinu manju ili jednaku od 75 kg, dok je preostala četvrtina ispitanika imala težinu veću od 75 kg.

**Analitički primjer 3.11.**

Za varijablu (obilježje) dobiveni džeparac (v3) potrebno je odrediti:

- a) Najmanju i najveću vrijednost obilježja;
- b) Aritmetičku sredinu;
- c) Mod;
- d) Medijan;
- e) Donji i gornji kvartil;
- f) Komentirati rezultate!



Rješenje 3.11:

Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v3 - Koliki džeparac dobivate prebacuje se u polje **Variable(s)**. Klikom na ikonu **Statistics** otvara se novi prozor **Frequencies Statistics**, gdje je potrebno aktivirati određene statističke veličine prema zahtjevu zadatka, a koje se odnose na odabranu varijablu džeparca (v3).

Da bi se u izlaznim rezultatima dobile najmanja i najveća vrijednost niza, u spomenutom prozoru aktivira se **Minimum** i **Maximum**. Aritmetička sredina dobit će se aktiviranjem veličine **Mean**, a za mod potrebno je označiti **Mode**. Medijan se dobije aktiviranjem **Median**, a donji i gornji kvartil označavanjem **Quartiles**. Da bi se u **Outputu** programa SPSS dobili željeni podaci o džeparcu ispitanika potrebno je kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 3.14.

Prema podacima iz tablice 3.14 može se vidjeti da je 173 ispitanika dalo odgovor o svom džeparcu, dok ih 66 nije odgovorilo na to pitanje. Prosječni džeparac ispitanika u odabranom uzorku (od onih koji su dali odgovor), odnosno aritmetička sredina je 371,55 kn.

Medijan, odnosno džeparac koji odabrani uzorak dijeli na 2 jednakobrojna dijela je $Me = 300 \text{ kn}$, što znači da je polovica ispitanika imala džeparac manji ili jednak 300 kn, dok je preostala polovica ispitanika imala džeparac veći od 300 kn.

Tablica 3.14.

Statistički podaci o džeparcu ispitanika

Statistics

Koliki džeparac dobivate

| | | |
|-------------|---------|-----------|
| N | Valid | 173 |
| | Missing | 66 |
| Mean | | 371,5491 |
| Median | | 300,0000 |
| Mode | | 200,00(a) |
| Minimum | | ,00 |
| Maximum | | 3500,00 |
| Percentiles | 25 | 150,0000 |
| | 50 | 300,0000 |
| | 75 | 500,0000 |

a Multiple modes exist. The smallest value is shown

Izvor: Simulirani podaci.

Mod, odnosno najčešći džeparac ispitanika je 200 kn. Može se vidjeti napomena ispod tablice koja se odnosi na mod. Zadana distribucija uzorka ima više od 1 moda, a u tablici Outputa dana je najmanja vrijednost. Ako se u Outputu još zatraži i prikaz tablice frekvencija (aktivira se **Display frequency tables**), može se vidjeti da je 26 ispitanika imalo džeparac od 200 kn i a da ih je isto toliko imalo džeparac od 300 kn. Dakle ovo je jedna bimodalna distribucija. Jedan mod je 200 kn, a drugi mod je 300 kn.

Najmanji džeparac ispitanika u uzorku je 0,00 kn.

Najveći džeparac ispitanika u uzorku je 3500,00 kn.

Donji i gornji kvartil dani su među **Percentiles**. Donji kvaril $Q_1 = 150\text{ kn}$, što znači da je četvrtina ispitanika imala džeparac manji od 150 kn ili jednak tom iznosu, dok je preostalih tri četvrtine ispitanika imalo džeparac veći od 150 kn. Gornji kvaril $Q_3 = 500\text{ kn}$, što znači da je tri četvrtine ispitanika imalo džeparac manji od 500 kn ili jednak tom iznosu, dok je preostala četvrtina ispitanika imala džeparac veći od 500 kn.

4. MJERE DISPERZIJE, ASIMETRIJA I ZAOBLJENOST

4.1. Mjere disperzije

Može se dogoditi da neki statistički skupovi imaju jednake npr. aritmetičke sredine, a da su njihovi elementi potpuno različiti. To znači da je raspored elemenata u tim skupovima različit. **Informaciju o rasporedu elemenata daju mjere raspršenosti ili disperzije elemenata numeričkog statističkog niza.**

Postoje apsolutne i relativne mjere raspršenosti. Apsolutni pokazatelji izraženi su u originalnim jedinicama mjere i omogućavaju usporedbu nizova prema istom obilježju. **Apsolutni pokazatelji raspršenosti su: raspon varijacije, interkvartil, varijanca i standardna devijacija.**

Usporedbu raspršenosti elemenata nizova s različitom mjernom jedinicom omogućuju relativni pokazatelji, koji su najčešće izraženi u postocima. **Relativni pokazatelji raspršenost su: koeficijent kvartilne devijacije i koeficijent varijacije.**

4.1.1. Apsolutne mjere disperzije

Raspon varijacije je najjednostavnija mjera disperzije, a predstavlja razliku između najveće i najmanje vrijednosti numeričkog obilježja promatranog niza.

$$R = x_{\max} - x_{\min} \quad (4.1)$$

Ovaj apsolutni pokazatelj raspršenosti izražen je u originalnim jedinicama mjere numeričkog obilježja. Može poprimiti vrijednost 0. To se događa kada svi elementi niza imaju jednaku vrijednost obilježja.

Najveća vrijednost toga pokazatelja nije ograničena jer ona ovisi o konkretnoj raspršenosti promatranih vrijednosti obilježja.

Raspon varijacije je nepotpuna mjera disperzije jer se računa samo na temelju dvije vrijednosti obilježja, odnosno na temelju najveće i najmanje vrijednosti. Može se reći da to nije precizna mjera raspršenosti elemenata niza, pogotovo u slučaju postojanja ekstremno malih i/ili ekstremno velikih vrijednosti obilježja. Tada se dobije veliki raspon varijacije, a možda je većina elemenata skupa raspršena usko oko srednjih vrijednosti.

Taj problem preciznosti rješava interkvartilni raspon ili interkvartil.

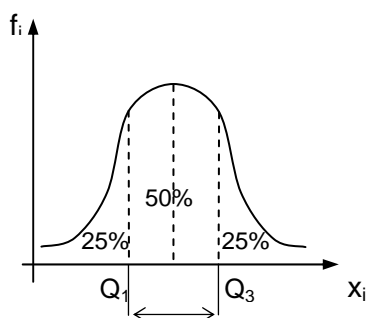
Interkvartil je apsolutna, nepotpuna mjera raspršenosti koja pokazuje disperziju srednjih 50% elemenata uređenoga numeričkog niza.

$$I_q = Q_3 - Q_1 \quad (4.2)$$

Interkvartil predstavlja razliku gornjeg i donjeg kvartila. Na taj način se eliminira 25% ekstremno malih i 25% ekstremno velikih vrijednosti obilježja u nizu.

Slika 4.1.

Simetrična distribucija s označenim gornjim i donjim kvartilom



Izvor: Konstrukcija autora.

Slika 4.1. prikazuje jednu simetričnu distribuciju gdje su elementi skupa ravnomjerno raspoređeni oko srednjih vrijednosti. Poznato je da

donji kvartil (Q_1) dijeli distribuciju u omjeru 1:3, tj. da 25% elemenata skupa ima vrijednost obilježja manju od donjeg kvartila, a 75% elemenata skupa ima vrijednost obilježja veću od donjeg kvartila. Isto tako gornji kvartil (Q_3) dijeli distribuciju u omjeru 3:1, tj. da 75% elemenata skupa ima vrijednost obilježja manju od gornjeg kvartila, a 25% elemenata skupa ima vrijednost obilježja veću od gornjeg kvartila. Na taj način interkvartil pokazuje disperziju srednjih 50% elemenata skupa.

Varijanca spada u potpune mjere raspršenosti jer obuhvaća sve elemente odabranoga numeričkog statističkog niza. Ovaj pokazatelj mjeri odstupanja, tj. raspršenost elemenata skupa od aritmetičke sredine.

S obzirom na poznato svojstvo aritmetičke sredine da je zbroj odstupanja vrijednosti obilježja od aritmetičke sredine u jednoj distribuciji uvijek nula, tj. da je: $\sum_{i=1}^N (x_i - \bar{X}) = 0$ (za negrupirani niz) i $\sum_{i=1}^k f_i (x_i - \bar{X}) = 0$ (za grupirani niz), varijanca se izražava preko kvadrata ovih odstupanja.

Varijanca je prosječno kvadratno odstupanje vrijednosti numeričkog obilježja od aritmetičke sredine.

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N} = \frac{\sum_{i=1}^N x_i^2}{N} - \bar{X}^2, \quad (4.3)$$

za negrupirani niz i

$$\sigma^2 = \frac{\sum_{i=1}^k f_i (x_i - \bar{X})^2}{\sum_{i=1}^k f_i} = \frac{\sum_{i=1}^k f_i x_i^2}{\sum_{i=1}^k f_i} - \bar{X}^2, \quad (4.4)$$

za grupirani niz.

Varijanca je mjera raspršenosti izražena u drugom stupnju, pa kao rezultat daje jedinice mjere numeričkog obilježja na kvadrat. Stoga je otežana njezina interpretacija.

Standardna devijacija je pozitivan korijen iz varijance i izražena je u originalnim jedinicama mjere. Stoga je kao potpuna i

apsolutna mjera disperzije vrlo često u upotrebi. Može se definirati kao **prosječno odstupanje vrijednosti numeričkog obilježja od aritmetičke sredine**.

$$\sigma = +\sqrt{\sigma^2} = +\sqrt{\frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N}} = +\sqrt{\frac{\sum_{i=1}^N x_i^2}{N} - \bar{X}^2}, \quad (4.5)$$

za negrupirani niz.

$$\sigma = +\sqrt{\sigma^2} = +\sqrt{\frac{\sum_{i=1}^k f_i (x_i - \bar{X})^2}{\sum_{i=1}^k f_i}} = +\sqrt{\frac{\sum_{i=1}^k f_i x_i^2}{\sum_{i=1}^k f_i} - \bar{X}^2}, \quad (4.6)$$

za grupirani niz. Pomoću standardne devijacije u originalnim mjernim jedinicama obilježja može se uspoređivati raspršenost oko aritmetičke sredine nizova koji su grupirani po jednakom obilježju.

Analitički primjer 4.1.

Zadan je statistički numerički niz "Zaposleni u sektoru Obrazovanje u RH prema dobi, stanje 31. ožujka 2004. god.". Zadatak je izračunati raspon varijacije obilježja, interkvartil, varijancu i standardnu devijaciju vrijednosti ovog obilježja.

Tablica 4.1.

Zaposleni u sektoru Obrazovanje u RH, stanje 31. ožujka 2004. god.

| Dob x_i | Br. zaposlenih f_i | Kum.niz "<" | Velič. razreda i |
|----------------|-------------------------|--------------|-----------------------|
| (18)-30 | 10172 | 9779 | 12 |
| 30-35 | 11082 | 20861 | 5 |
| 35-40 | 12430 | 33291 | 5 |
| 40-45 | 14130 | 47421 | 5 |
| 45-50 | 12830 | 60251 | 5 |
| 50-55 | 12040 | 72291 | 5 |
| 55-60 | 10039 | 82330 | 5 |
| 60-(70) | 5990 | 88320 | 10 |
| Ukupno: | 88713 | - | - |

Izvor: Statistički ljetopis Republike Hrvatske, 2005. godina, str. 144.

Rješenje 4.1:**Raspon varijacije obilježja:**

$$R = x_{\max} - x_{\min} = 70 - 18 = 52 \text{ godine}$$

Raspon varijacije ovog numeričkog obilježja dobije se kao razlika najveće i najmanje vrijednosti obilježja, a iznosi 52 godine.

Interkvartil:

U tablici 4.1. prikazan je grupirani niz koji ima veličine razreda različite od 1. Stoga je za izračunavanje kvartila potrebno odrediti frekvencije kumulativnog niza "manje od" tako da se postupno zbrajaju apsolutne frekvencije odozgo prema dolje.

Za **donji kvartil** se traži prva frekvencija kumulativnog niza koja u sebi sadrži: $\frac{N}{4} = \frac{88713}{4} = 22178,25$ jedinica. To je frekvencija 33291, a razred koji odgovara toj frekvenciji je donje kvarilni razred. Njegova donja granica je $L_1 = 35$.

$$Q_1 = L_1 + \frac{\frac{N}{4} - \sum_{i=1}^q f_i}{f_{k \text{ var } t}} \cdot i = 35 + \frac{22178,25 - 20861}{12430} \cdot 5 = 35,53$$

Dakle, jedna četvrtina (tj. 25%) zaposlenih u Obrazovanju u Republici Hrvatskoj u 2004. godini je bilo mlađe od 35,53 godina, dok je preostalih tri četvrtine zaposlenih (tj. 75%) bilo starije od 35,53 godina.

Za **gornji kvartil** se traži prva frekvencija kumulativnog niza koja u sebi sadrži: $\frac{3N}{4} = \frac{3 \cdot 88713}{4} = 66534,75$ jedinica. To je frekvencija 72291, a razred koji odgovara toj frekvenciji je gornje kvarilni razred. Njegova donja granica je $L_1 = 50$.

$$Q_3 = L_1 + \frac{\frac{3N}{4} - \sum_{i=1}^q f_i}{f_{k \text{ var } t}} \cdot i = 50 + \frac{66534,75 - 60251}{12040} \cdot 5 = 52,61$$

Dakle, tri četvrtine (tj. 75%) zaposlenih u Obrazovanju u R Hrvatskoj u 2002. godini je bilo mlađe od 52,61 godina, dok je preostala četvrtina zaposlenih (tj. 25%) bilo starije od 52,61 godina.

Interkvartil je razlika kvartila:

$$I_q = Q_3 - Q_1 = 52,61 - 35,53 = 17,08 \text{ godina}$$

Dakle, dob srednjih 50% zaposlenih u Obrazovanju u ovoj distribuciji varira u rasponu od 17,08 godina.

Da bi se izračunala varijanca i standardna devijacija vrijednosti ovog obilježja, prikazani su podaci u tablici 4.2.

Tablica 4.2.

Zaposleni u sektoru Obrazovanje u RH, stanje 31. ožujka 2004. god.

| Dob x_i | Broj zaposlenih f_i | Sredina razreda x_i | $f_i x_i$ | $f_i (x_i - \bar{X})^2$ | $f_i x_i^2$ |
|----------------|-----------------------------|-----------------------------|-----------|-------------------------|-------------|
| (18)-30 | 10172 | 24 | 244128 | 3958060 | 5859072 |
| 30-35 | 11082 | 32,5 | 360165 | 1396576 | 11705363 |
| 35-40 | 12430 | 37,5 | 466125 | 481817,8 | 17479688 |
| 40-45 | 14130 | 42,5 | 600525 | 21236,85 | 25522313 |
| 45-50 | 12830 | 47,5 | 609425 | 182743,2 | 28947688 |
| 50-55 | 12040 | 52,5 | 632100 | 926886,1 | 33185250 |
| 55-60 | 10039 | 57,5 | 577242,5 | 1904643 | 33191444 |
| 60-(70) | 5990 | 65 | 389350 | 2710985 | 25307750 |
| Ukupno: | 88713 | - | 3879061 | 11582947 | 181198566 |

Izvor: Statistički ljetopis Republike Hrvatske, 2005. godina, str. 144.

Zadatak je izračunati varijancu i standardnu devijaciju vrijednosti zadanog obilježja.

Ovo je grupirani niz. Varijanca i standardna devijacija računaju se po izrazima od (4.3) do (4.6). Najprije je potrebno izračunati složenu aritmetičku sredinu za negrupirane nizove.

$$\bar{X} = \frac{\sum_{i=1}^k f_i x_i}{\sum_{i=1}^k f_i} = \frac{3879061}{88713} = 43,726 \text{ godina}$$

Dakle, prosječna dob zaposlenih u sektoru Obrazovanje u RH, stanje 31. ožujka 2004. god., 43,726 je godina. Varijanca je:

$$\sigma^2 = \frac{\sum_{i=1}^k f_i (x_i - \bar{X})^2}{\sum_{i=1}^k f_i} = \frac{11582947}{88713} = 130,57 \text{ godina}^2.$$

Vrijednosti iz brojnika, $\sum_{i=1}^k f_i (x_i - \bar{X})^2$, u izrazu za varijancu su izračunate u 5. stupcu tablice, a njihova suma je dana u zbirnom retku. Isti rezultat se može dobiti i preko izvedene formule za varijancu grupiranih nizova (4.6):

$$\sigma^2 = \frac{\sum_{i=1}^k f_i x_i^2}{\sum_{i=1}^k f_i} - \bar{X}^2 = \frac{181198566}{88713} - 43,726^2 = 130,57 \text{ godina}^2.$$

Zbroj $f_i x_i^2$ dan je u tablici 4.2. u posljednjem retku. Varijanca je mjera raspršenosti vrijednosti obilježja u drugom stupnju pa je izražena u jedinicama mjere na kvadrat. Dalje se računa standardna devijacija koja se izražava u originalnim mjernim jedinicama. Standardna devijacija je pozitivan korijen iz varijance:

$$\sigma = +\sqrt{\sigma^2} = +\sqrt{130,57} = 11,43 \text{ godina}.$$

Dakle, prosječno odstupanje vrijednosti ovog numeričkog obilježja od aritmetičke sredine je 11,43 godina.

4.1.2. Relativne mjere disperzije

Koeficijent varijacije spada u potpune relativne mjere raspršenosti jer obuhvaća sve elemente odabranog numeričkoga statističkog niza, a izražava se u postotcima (%).

Koeficijent varijacije je postotak standardne devijacije od aritmetičke sredine.

$$V = \frac{\sigma}{\bar{X}} \cdot 100. \quad (4.7)$$

Vrijednost koeficijenta varijacije se kreće u intervalu $0 \leq V < +\infty$. Vrijednost od 0% će poprimiti samo u slučaju kada su sve vrijednosti numeričkog obilježja u jednom nizu jednake, odnosno kada nema disperzije. Veća vrijednost ovog pokazatelja upućuje na veću disperziju elemenata promatranog niza.

Taj pokazatelj raspršenosti je izražen u postotcima, pa omogućuje usporedbu disperzije numeričkih nizova s različitim jedinicama mjere.

Koeficijent kvartilne devijacije je relativna nepotpuna mjera raspršenosti. Predstavlja relativnu disperziju srednjih 50% elemenata numeričkog niza. Računa se na temelju samo dviju vrijednosti obilježja, a to su donji i gornji kvartil.

$$V_q = \frac{Q_3 - Q_1}{Q_3 + Q_1} \quad (4.8)$$

Koeficijent kvartilne devijacije se računa kao omjer interkvartila (razlika kvartila) i zbroja donjeg i gornjeg kvartila. Vrijedi da je:

$$0 \leq V_q < 1 \quad (4.9)$$

Ovaj pokazatelj je jednak 0 samo kada nema disperzije. Ako se njegova vrijednost približava 1, to znači da je raspršenost vrijednosti obilježja veća.

S obzirom na to da je to relativni pokazatelj, pomoću njega je moguće uspoređivati raspršenost srednjih 50% elemenata različitih numeričkih distribucija s različitim jedinicama mjere.

Analitički primjer 4.2.

Prema podacima o statističkom numeričkom nizu "Zaposleni u sektoru Obrazovanje u RH prema dobi, stanje 31. ožujka 2004. god.", koji su dani u analitičkom primjeru 4.1., zadatak je izračunati koeficijent varijacije i koeficijent kvartilne devijacije.

Rješenje 4.2:**Koeficijent varijacije:**

$$V = \frac{\sigma}{\bar{X}} \cdot 100 = \frac{11,4266}{43,72595} \cdot 100 = 26,132\%$$

Postotak standardne devijacije od aritmetičke sredine u zadanom nizu je 26,132%.

Koeficijent kvartilne devijacije.

$$V_q = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{I_q}{Q_3 + Q_1} = \frac{17,08}{52,61 + 35,53} = 0,194$$

Dakle, dob srednjih 50% zaposlenih u Obrazovanju u ovoj distribuciji pokazuje slabu raspršenost 0,194 (s obzirom na to da se pokazatelj kreće u intervalu između 0 i 1).

**Analitički primjer 4.3.**

Za varijablu (obilježje) visina u cm (v2) potrebno je odrediti:

- a) Tablicu frekvencija;
- b) Raspon varijacije obilježja;
- c) Aritmetičku sredinu;
- d) Donji i gornji kvartil;
- e) Varijancu;
- f) Standardnu devijaciju;
- g) Na temelju toga izračunati: interkvartil, koeficijent varijacije i koeficijent kvartilne devijacije;
- h) Komentirati rezultate!



Rješenje 4.3:

Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v2 - Visina u cm se prebacuje u polje **Variable(s)**. Klikom na ikonu **Statistics** otvara se novi prozor **Frequencies Statistics** gdje je potrebno aktivirati određene statističke veličine prema zahtjevu zadatka, a koje se odnose na odabranu varijablu visine u cm (v2).

Da bi se u izlaznim rezultatima dobio raspon varijacije obilježja niza, u spomenutom prozoru aktivira se **Range**. Aritmetička sredina dobit će se aktiviranjem veličine **Mean**, a za mod potrebno je označiti **Mode**. Donji i gornji kvartil dobiju se označavanjem **Quartiles**. Stanardna devijacija i varijanca dobiju se aktiviranjem **Std. deviation** i **Variance**. Konačno, da bi se u **Outputu** programa SPSS dobili željeni podaci o visini ispitanika, potrebno je kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 4.3.

Tablica 4.3.

Statistički podaci o visini ispitanika

| Statistics | | |
|----------------|---------|--------|
| Visina u cm | | |
| N | Valid | 235 |
| | Missing | 4 |
| Mean | | 175,77 |
| Std. Deviation | | 8,749 |
| Variance | | 76,539 |
| Range | | 52 |
| Percentiles | 25 | 170,00 |
| | 50 | 174,00 |
| | 75 | 182,00 |

Izvor: Simulirani podaci.

Prema podacima iz tablice 4.3. može se vidjeti da je 235 ispitanika dalo odgovor o svojoj visini, dok ih 4 nije odgovorilo na to pitanje. Prosječna visina ispitanika u odabranom uzorku (od onih koji su dali odgovor), odnosno aritmetička sredina je 175,77 cm.

Razlika između najveće i najmanje visine ispitanika u uzorku, tj. raspon varijacije obilježja je 52 cm.

Donji i gornji kvartil dani su kao **Percentiles**. Donji kvartil $Q_1 = 170 \text{ cm}$, što znači da je četvrtina ispitanika imala visinu manju od 170 cm ili jednaku toj visinu, dok je preostalih tri četvrtine ispitanika bilo s visinom većom od 170 cm. Gornji kvartil $Q_3 = 182 \text{ cm}$, što znači da je tri četvrtine ispitanika bilo s visinom manjom od 182 cm ili jednakom toj visini, dok je preostala četvrtina ispitanika imala visinu veću od 182 cm.

Varijanca zadanog niza, prosječno kvadratno odstupanje vrijednosti zadanog obilježja od prosjeka je $76,539 \text{ cm}^2$. Korijen iz varijance, tj. standardna devijacija zadanog niza je 8,749 cm.

Na temelju ovih podataka iz Outputa programa SPSS može se izračunati interkvartil, koeficijent varijacije i koeficijent kvartilne devijacije.

Interkvartil je razlika kvartila:

$$I_q = Q_3 - Q_1 = 182 - 170 = 12 \text{ cm}.$$

Dakle, visina srednjih 50% ispitanika u ovoj distribuciji varira u rasponu od 12 cm.

Koeficijent varijacije je:

$$V = \frac{\sigma}{\bar{X}} \cdot 100 = \frac{8,749}{175,77} \cdot 100 = 4,98\%.$$

Postotak standardne devijacije od aritmetičke sredine u zadanom nizu je 4,98% i upućuje na vrlo malu disperziju i veliku reprezentativnost aritmetičke sredine.

Koeficijent kvartilne devijacije je:

$$V_q = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{I_q}{Q_3 + Q_1} = \frac{12}{182 + 170} = 0,034.$$

Dakle, visina srednjih 50% ispitanika u ovoj distribuciji pokazuje izrazito slabu raspršenost 0,034 (s obzirom na to da se pokazatelj kreće u intervalu između 0 i 1).



Analitički primjer 4.4.

Za varijablu (obilježje) težina u kg (v2) potrebno je odrediti:

- a) Tablicu frekvencija;
- b) Raspon varijacije obilježja;
- c) Aritmetičku sredinu;
- d) Donji i gornji kvartil;
- e) Varijancu;
- f) Standardnu devijaciju;
- g) Na temelju toga izračunati: interkvartil, koeficijent varijacije i koeficijent kvartilne devijacije;
- h) Komentirati rezultate!



Rješenje 4.4:

Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v2 - Težina u kg se prebacuje u polje **Variable(s)**. Klikom na ikonu **Statistics** otvara se novi prozor **Frequencies Statistics** gdje je potrebno aktivirati određene statističke veličine prema zahtjevu zadatka, a koje se odnose na odabranu varijablu težine u kg (v2).

Da bi se u izlaznim rezultatima dobio raspon varijacije obilježja niza, u spomenutom prozoru aktivira se **Range**. Aritmetička sredina dobit će se aktiviranjem veličine **Mean**, a za mod potrebno je označiti **Mode**. Donji i gornji kvartil dobiju se označavanjem **Quartiles**. Standardna devijacija i varijanca dobiju se aktiviranjem **Std. deviation** i **Variance**. Konačno, da bi se u **Outputu** programa SPSS dobili željeni podaci o težini ispitanika, potrebno je kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 4.4.

Prema podacima iz tablice 4.4. može se vidjeti da je 234 ispitanika dalo odgovor o svojoj težini, dok ih 5 nije odgovorilo na to pitanje.

Prosječna težina ispitanika u odabranom uzorku (od onih koji su dali odgovor), odnosno aritmetička sredina je 67,1795 kg.

Razlika između najveće i najmanje težine ispitanika u uzorku, tj. raspon varijacije obilježja je 64 kg.

Tablica 4.4.

Statistički podaci o težini ispitanika

Statistics

Težina u kg

| | | |
|----------------|---------|----------|
| N | Valid | 234 |
| | Missing | 5 |
| Mean | | 67,1795 |
| Std. Deviation | | 13,52735 |
| Variance | | 182,989 |
| Range | | 64,00 |
| Percentiles | 25 | 56,0000 |
| | 50 | 64,0000 |
| | 75 | 75,0000 |

Izvor: Simulirani podaci.

Donji i gornji kvartil dani su kao **Percentiles**. Donji kvaril $Q_1 = 56\text{ kg}$, što znači da je četvrtina ispitanika imala težinu manju od 56 kg ili jednaku toj težini, dok je preostalih tri četvrtine ispitanika bilo s težinom većom od 56 kg. Gornji kvartil $Q_3 = 75\text{ kg}$, što znači da je tri četvrtine ispitanika bilo s težinom manjom od 75 kg ili jednakom tom težinom, dok je preostala četvrtina ispitanika imala težinu veću od 75 kg.

Varijanca zadanog niza, prosječno kvadratno odstupanje vrijednosti zadanog obilježja od prosjeka je $182,989\text{ kg}^2$. Korijen iz varijance, tj. standardna devijacija zadanog niza je 13,53 kg.

Na temelju ovih podataka iz Outputa programa SPSS može se izračunati interkvartil, koeficijent varijacije i koeficijent kvartilne devijacije.

Interkvartil je razlika kvartila:

$$I_q = Q_3 - Q_1 = 75 - 56 = 19\text{ kg}.$$

Dakle, težina srednjih 50% ispitanika u ovoj distribuciji varira u rasponu od 19 kg.

Koeficijent varijacije je:

$$V = \frac{\sigma}{\bar{X}} \cdot 100 = \frac{13,52735}{67,1795} \cdot 100 = 20,136\%$$

Postotak standardne devijacije od aritmetičke sredine u zadanom nizu je 20,136%. Reprezentativnost je velika, a disperzija mala.

Koeficijent kvartilne devijacije je:

$$V_q = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{I_q}{Q_3 + Q_1} = \frac{19}{75 + 56} = 0,145$$

Dakle, visina srednjih 50% ispitanika u ovoj distribuciji pokazuje slabu raspršenost 0,145.



Analitički primjer 4.5.

Za varijablu (obilježje) dobiveni džeparac (v2) potrebno je odrediti:

- a) Tablicu frekvencija;
- b) Raspon varijacije obilježja;
- c) Aritmetičku sredinu;
- d) Donji i gornji kvartil;
- e) Varijancu;
- f) Standardnu devijaciju;
- g) Na temelju toga izračunati: interkvartil, koeficijent varijacije i koeficijent kvartilne devijacije;
- h) Komentirati rezultate!



Rješenje 4.5:

Na glavnom izborniku potrebno je izabrati ikonu *Analyze*, a na njezinu padajućem izborniku *Descriptive Statistics* i *Frequencies*.

U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v2 - Dobiveni džeparac se prebacuje u polje **Variable(s)**. Klikom na ikonu **Statistics** otvara se novi prozor **Frequencies Statistics** gdje je potrebno aktivirati određene statističke veličine prema zahtjevu zadatka, a koje se odnose na odabranu varijablu dobiveni džeparac (v2).

Da bi se u izlaznim rezultatima dobio raspon varijacije obilježja niza u spomenutom prozoru aktivira se **Range**. Aritmetička sredina dobit će se aktiviranjem veličine **Mean**, a za mod potrebno je označiti **Mode**. Donji i gornji kvartil dobiju se označavanjem **Quartiles**. Stanardna devijacija i varijanca dobiju se aktiviranjem **Std. deviation** i **Variance**. Konačno, da bi se u **Outputu** programa **SPSS** dobili željeni podaci o dobivenom džeparcu ispitanika, potrebno je kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 4.5.

Tablica 4.5.

Statistički podaci o džeparcu ispitanika

| Statistics | | |
|-------------------|---------|------------|
| Dobiveni džeparac | | |
| N | Valid | 173 |
| | Missing | 66 |
| Mean | | 371,5491 |
| Std. Deviation | | 401,30549 |
| Variance | | 161046,098 |
| Range | | 3500,00 |
| Percentiles | 25 | 150,0000 |
| | 50 | 300,0000 |
| | 75 | 500,0000 |

Izvor: Simulirani podaci.

Prema podacima iz tablice 4.5. može se vidjeti da je 173 ispitanika dalo odgovor o svojem džeparcu, dok ih 66 nije odgovorilo na to pitanje. Prosječni džeparac ispitanika u odabranom uzorku (od onih koji su dali odgovor), odnosno aritmetička sredina je 371,55 kn.

Razlika između najveće i najmanje težine ispitanika u uzorku, tj. raspon varijacije obilježja je 3500 kn.

Donji i gornji kvartil dani su kao **Percentiles**. Donji kvaril $Q_1 = 150\text{ kn}$, što znači da je četvrtina ispitanika imala džeparac manji od

150 kn ili jednak tom iznosu, dok je preostalih tri četvrtine ispitanika bilo s džeparcem većim od 150 kn. Gornji kvartil $Q_3 = 500 \text{ kn}$, što znači da je tri četvrtine ispitanika bilo s džeparcem manjim od 500 kn ili jednakim tom iznosu, dok je preostala četvrtina ispitanika imala džeparac veći od 500 kn.

Varijanca zadanog niza, prosječno kvadratno odstupanje vrijednosti zadanog obilježja od prosjeka je $161046,1 \text{ kn}^2$. Korijen iz varijance, tj. standardna devijacija zadanog niza je $401,31 \text{ kn}$.

Na temelju ovih podataka iz Outputa programa SPSS može se izračunati interkvartil, koeficijent varijacije i koeficijent kvartilne devijacije.

Interkvartil je razlika kvartila:

$$I_q = Q_3 - Q_1 = 500 - 150 = 350 \text{ kn}.$$

Dakle, džeparac srednjih 50% ispitanika u ovoj distribuciji varira u rasponu od 350 kn.

Koeficijent varijacije je:

$$V = \frac{\sigma}{\bar{X}} \cdot 100 = \frac{401,30549}{371,5491} \cdot 100 = 108,01\%.$$

Postotak standardne devijacije od aritmetičke sredine u zadanom nizu je $108,01\%$, što ukazuje na jako veliku disperziju džeparca između promatranih ispitanika i malu reprezentativnost \bar{X} .

Koeficijent kvartilne devijacije je:

$$V_q = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{I_q}{Q_3 + Q_1} = \frac{350}{500 + 150} = 0,538.$$

Dakle, visina srednjih 50% ispitanika u ovoj distribuciji pokazuje priličnu raspršenost od $0,538$.



Analitički primjer 4.6.

Za varijablu (obilježje) broj djece u obitelji (v2) potrebno je odrediti:

- a) Tablicu frekvencija;
- b) Raspon varijacije obilježja;
- c) Aritmetičku sredinu;
- d) Donji i gornji kvartil;
- e) Varijancu;
- f) Standardnu devijaciju;
- g) Na temelju toga izračunati: interkvartil, koeficijent varijacije i koeficijent kvartilne devijacije;
- h) Komentirati rezultate!



Rješenje 4.6:

Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v2 - Broj djece u obitelji se prebacuje u polje **Variable(s)**. Klikom na ikonu **Statistics** otvara se novi prozor **Frequencies Statistics** gdje je potrebno aktivirati određene statističke veličine prema zahtjevu zadatka, a koje se odnose na odabranu varijablu broj djece u obitelji (v2).

Da bi se u izlaznim rezultatima dobio raspon varijacije obilježja niza, u spomenutom prozoru aktivira se **Range**. Aritmetička sredina dobit će se aktiviranjem veličine **Mean**, a za mod potrebno je označiti **Mode**. Donji i gornji kvartil dobiju se označavanjem **Quartiles**. Standardna devijacija i varijanca dobiju se aktiviranjem **Std. deviation** i **Variance**.

Konačno, da bi se u **Outputu** programa SPSS dobili željeni podaci o broju djece u obitelji ispitanika, potrebno je kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 4.6.

Prema podacima iz tablice 4.6. može se vidjeti da je 238 ispitanika dalo odgovor o broju djece u obitelji, dok 1 nije odgovorio na to pitanje. Prosječni broj djece u obitelji ispitanika u odabranom uzorku (od onih koji su dali odgovor), odnosno aritmetička sredina je 2,34. S obzirom na

to da broj djece spada u diskontinuirana numerička obilježja, ovaj rezultat se može komentirati kao prosječno 2 djece po obitelji ispitanika.

Tablica 4.6.

Statistički podaci o broju djece u obitelji ispitanika

Statistics

Broj djece u obitelji

| | | |
|----------------|---------|------|
| N | Valid | 238 |
| | Missing | 1 |
| Mean | | 2,34 |
| Std. Deviation | | ,870 |
| Variance | | ,757 |
| Range | | 6 |
| Percentiles | 25 | 2,00 |
| | 50 | 2,00 |
| | 75 | 3,00 |

Izvor: Simulirani podaci.

Razlika između najvećeg i najmanjeg broja djece u obitelji ispitanika u uzorku, tj. raspon varijacije obilježja je 6 djece.

Donji i gornji kvartil dani su kao **Percentiles**. Donji kvaril $Q_1 = 2$ djece, što znači da je četvrtina ispitanika imala broj djece u obitelji manji od 2 ili jednak 2, dok je preostalih tri četvrtine ispitanika bilo s više od 2 djece u obitelji. Gornji kvartil $Q_3 = 3$ djece, što znači da je tri četvrtine ispitanika bilo s brojem djece u obitelji manjim od 3 ili jednakim 3, dok je preostala četvrtina ispitanika imala više od 3 djece u obitelji.

Varijanca zadanog niza, prosječno kvadratno odstupanje vrijednosti zadanog obilježja od prosjeka je $0,757$ djece². Korijen iz varijance, tj. standardna devijacija zadanog niza je $0,87$ djece. Obje vrijednosti se mogu zaokružiti na 1, s obzirom na to da se radi o diskontinuiranom numeričkom obilježju.

Na temelju ovih podataka iz Outputa programa SPSS može se izračunati interkvartil, koeficijent varijacije i koeficijent kvartilne devijacije.

Interkvartil je razlika kvartila:

$$I_q = Q_3 - Q_1 = 3 - 2 = 1 \text{ dijete.}$$

Dakle, broj djece u obitelji srednjih 50% ispitanika u ovoj distribuciji varira u rasponu od 1 djeteta.

Koeficijent varijacije je:

$$V = \frac{\sigma}{\bar{X}} \cdot 100 = \frac{0,870}{2,34} \cdot 100 = 37,18\%.$$

Postotak standardne devijacije od aritmetičke sredine u zadanom nizu je 37,18%.

Koeficijent kvartilne devijacije je:

$$V_q = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{I_q}{Q_3 + Q_1} = \frac{1}{3 + 2} = 0,2.$$

Dakle, broj djece srednjih 50% ispitanika u ovoj distribuciji pokazuje slabu raspršenost od 0,2.

4.1.3. Momenti numeričkih nizova

Momenti su prosječna odstupanja vrijednosti numeričkog obilježja od aritmetičke sredine ili neke druge vrijednosti podignuti na r -tu potenciju.

Momenti numeričkih nizova mogu biti **glavni i pomoćni**.

Glavni ili centralni moment numeričkih nizova su momenti oko aritmetičke sredine:

$$\mu_r = \frac{\sum_{i=1}^N (x_i - \bar{X})^r}{N}, \quad r = 0, 1, 2, \dots \quad - \text{ za negrupirani niz. } \quad (4.9)$$

$$\mu_r = \frac{\sum_{i=1}^k f_i (x_i - \bar{X})^r}{\sum_{i=1}^k f_i}, \quad r = 0, 1, 2, \dots \quad - \text{ za grupirani niz. } \quad (4.10)$$

Centralni moment nultog reda: $\mu_0 = 1$.

Centralni moment prvog reda: $\mu_1 = 0$.

Pomoćni momenti numeričkih nizova su momenti oko nule⁶:

$$m_r = \frac{\sum_{i=1}^N x_i^r}{N}, \quad r = 0, 1, 2, \dots \quad \text{- za negrupirani niz.} \quad (4.11)$$

$$m_r = \frac{\sum_{i=1}^k f_i x_i^r}{\sum_{i=1}^k f_i}, \quad r = 0, 1, 2, \dots \quad \text{- za grupirani niz.} \quad (4.12)$$

Pomoćni moment nultog reda: $m_0 = 1$.

Pomoćni moment prvog reda: $m_1 = \bar{X}$.

Centralni momenti oko aritmetičke sredine mogu se izraziti **preko pomoćnih momenata**.

Centralni moment drugog reda:

$$\mu_2 = m_2 - m_1^2. \quad (4.13)$$

Centralni moment trećeg reda:

$$\mu_3 = m_3 - 3m_1m_2 + 2m_1^3. \quad (4.14)$$

Centralni moment četvrtog reda:

$$\mu_4 = m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4. \quad (4.15)$$

4.2. Mjere asimetrije

Asimetrija distribucije podrazumijeva nagnutost distribucije na lijevu ili desnu stranu.

Pearsonov koeficijent asimetrije je:

$$\alpha_3 = \frac{\mu_3}{\sigma^3}, \quad (4.16)$$

⁶ Postoje pomoćni momenti i oko neke konstante a .

gdje je:

μ_3 - centralni moment trećeg reda,

σ^3 - standardna devijacija na treću potenciju.

Interval u kojem se uobičajeno kreće vrijednost tog koeficijenta je: $-2 \leq \alpha_3 \leq 2$.

U slučaju izrazito asimetričnim distribucijama ovaj koeficijent može poprimiti i vrijednosti izvan intervala: $[-2, 2]$.

Ako je:

$\alpha_3 = 0 \quad \Rightarrow \quad$ simetrična distribucija

$\alpha_3 > 0 \quad \Rightarrow \quad$ pozitivna ili desnostrana asimetrija

$\alpha_3 < 0 \quad \Rightarrow \quad$ negativna ili ljevostrana asimetrija.

U programskom paketu **SPSS** mjera asimetrije ima oznaku **Skewness** i vrijedi da ako je vrijednost ovog pokazatelja:

$= 0 \quad \Rightarrow \quad$ simetrična distribucija

$> 0 \quad \Rightarrow \quad$ pozitivna ili desnostrana asimetrija

$< 0 \quad \Rightarrow \quad$ negativna ili ljevostrana asimetrija.

Pearsonova mjera asimetrije, koja se uobičajeno kreće u intervalu $[-3, 3]$, je:

$$S_k = \frac{(\bar{X} - Mo)}{\sigma}, \quad (4.17)$$

gdje je:

\bar{X} - aritmetička sredina

Mo - mod

σ - standardna devijacija.

Za umjereno asimetrične distribucije vrijedi da je **Pearsonova mjera asimetrije**:

$$S_k = \frac{3 \cdot (\bar{X} - Me)}{\sigma}, \quad (4.18)$$

gdje je:

\bar{X} - aritmetička sredina

Me - medijan

σ - standardna devijacija.

Ako je:

$S_k = 0 \quad \Rightarrow \quad$ simetrična distribucija

$S_k > 0 \quad \Rightarrow \quad$ pozitivna ili desnostrana asimetrija

$S_k < 0 \quad \Rightarrow \quad$ negativna ili ljevostrana asimetrija.

Odnosi među srednjim vrijednostima kod različitih distribucija:

- a) **simetrična distribucija** ima jednake aritmetičku sredinu, medijan i mod:

$$\bar{X} = Me = Mo .$$

- b) **pozitivno ili desnostrano asimetrična distribucija** ima aritmetičku sredinu veću od medijana koji je veći od moda:

$$\bar{X} > Me > Mo .$$

- c) **negativno ili ljevostrano asimetrična distribucija** ima aritmetičku sredinu manju od medijana koji je manji od moda:

$$\bar{X} < Me < Mo .$$

Bowleyeva mjera asimetrije (uobičajeno se kreće u intervalu $[-1,1]$) zasnovana je na odnosu medijana i kvartila:

$$S_{kq} = \frac{Q_3 + Q_1 - 2Me}{Q_3 - Q_1}, \quad (4.19)$$

gdje je:

Me - medijan

Q_1 - donji kvartil

Q_3 - gornji kvartil.

Vrijedi da ako je:

$$S_{kq} = 0 \Rightarrow (Q_3 + Q_1) = 2Me \Rightarrow$$

$$(Q_3 - Me) = (Me - Q_1) \Rightarrow \text{simetrična distribucija}$$

$$S_k > 0 \Rightarrow (Q_3 + Q_1) > 2Me \Rightarrow$$

$(Q_3 - Me) > (Me - Q_1) \Rightarrow$ pozitivna ili desnostrana
asimetrija

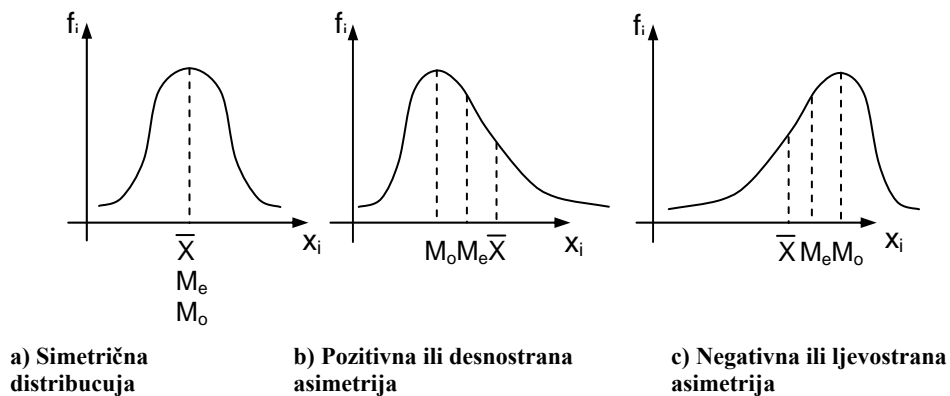
$$S_k < 0 \Rightarrow (Q_3 + Q_1) < 2Me \Rightarrow$$

$(Q_3 - Me) < (Me - Q_1) \Rightarrow$ negativna ili ljevostrana
asimetrija.

Raspored srednjih vrijednosti kod različitih distribucija prikazan je na slici 4.2.

Slika 4.2.

Distribucije s različitom asimetrijom



Izvor: Konstrukcija autora.

Prikazana su 3 slučaja raspršenosti podataka oko srednjih vrijednosti. Na dijelu slike 4.2. pod (a) prikazana je jedna savršeno **simetrična distribucija u kojoj su aritmetička sredina (prosjeck),**

medijan (vrijednost obilježja koja skup dijeli na dva jednaka dijela) i **mod** (najčešća vrijednost obilježja koja odgovara vrhu krivulje) u istoj točki, odnosno **jednaki**: $\bar{X} = Me = Mo$.

Na dijelu slike pod (b) prikazana je jedna **pozitivno ili desnostrano asimetrična distribucija** (vidi se rasipanje vrijednosti obilježja na desnu stranu) **u kojoj je aritmetička sredina veća od medijana koji je veći od moda**: $\bar{X} > Me > Mo$.

Na dijelu slike pod (c) prikazana je jedna **negativno ili ljevostrano asimetrična distribucija** (vidi se rasipanje vrijednosti obilježja na lijevu stranu) **u kojoj je aritmetička sredina manja od medijana koji je manji od moda**: $\bar{X} < Me < Mo$.

4.3. Mjera zaobljenosti

Mjera zaobljenosti predstavlja zaobljenost vrha krivulje distribucije frekvencija:

$$\alpha_4 = \frac{\mu_4}{\sigma^4}, \quad (4.20)$$

gdje je:

μ_4 - centralni moment četvrtog reda

σ^4 - standardna devijacija na četvrtu potenciju.

Vrijedi da ako je:

$\alpha_4 = 3 \Rightarrow$ normalno zaobljena distribucija

$\alpha_4 > 3 \Rightarrow$ šiljatiji vrh od normalno zaobljene distribucije

$\alpha_4 < 3 \Rightarrow$ tupi oblik distribucije.

U programskom paketu **SPSS** mjera zaobljenosti ima oznaku **Kurtosis** ($Kurtosis = k$; $k = \alpha_4 - 3$) i vrijedi da ako je vrijednost ovog pokazatelja:

$= 0 \Rightarrow$ normalno zaobljena distribucija

- $> 0 \Rightarrow$ šiljati vrh od normalno zaobljene distribucije
 $< 0 \Rightarrow$ tupi oblik distribucije.



Analitički primjer 4.7.

Za varijablu (obilježje) visina u cm (v2) potrebno je odrediti:

- Asimetriju distribucije (Skewness);
- Pearsonovu mjeru asimetrije (aritmetička sredina, mod, medijan, standardna devijacija);
- Bowleyevu mjeru asimetrije (kvartili);
- Zaobljenost distribucije (Kurtosis);
- Na histogram distribucije ucrtati i normalnu krivulju;
- Komentirati rezultate!



Rješenje 4.7:

Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

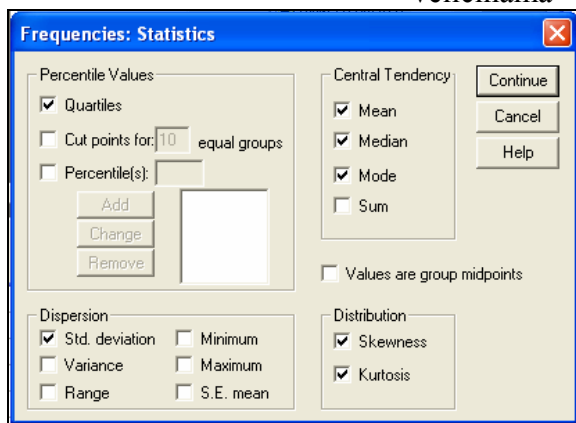
U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v2 - Visina u cm se prebacuje u polje **Variable(s)**. Klikom na ikonu **Statistics** otvara se novi prozor **Frequencies Statistics** gdje je potrebno aktivirati određene statističke veličine prema zahtjevu zadatka, a koje se odnose na odabranu varijablu visina u cm (v2).

Da bi se u izlaznim rezultatima dobio pokazatelj asimetrije zadanog niza, u spomenutom prozoru aktivira se **Skewness**. Ostali pokazatelji asimetrije mogu se izračunati pomoću aritmetičke sredine koja će se dobiti aktiviranjem veličine **Mean**. Za mod je potrebno označiti **Mode**. Donji i gornji kvartil dobiju se označavanjem **Quartiles**. Standardna devijacija dobije se aktiviranjem **Std. deviation**, a mjera zaobljenosti aktiviranjem veličine **Kurtosis**.

Odabrane veličine u programu **SPSS** prikazane su na slici 4.3.

Slika 4.3.

Prozor "Frequencies Statistics" sa izabranim odgovarajućim statističkim veličinama



Izvor: Simulirani podaci.

Tablica 4.7.

Statistički podaci o visini ispitanika

| | | |
|------------------------|---------|--------|
| N | Valid | 235 |
| | Missing | 4 |
| Mean | | 175,77 |
| Median | | 174,00 |
| Mode | | 172 |
| Std. Deviation | | 8,749 |
| Skewness | | ,485 |
| Std. Error of Skewness | | ,159 |
| Kurtosis | | ,184 |
| Std. Error of Kurtosis | | ,316 |
| Percentiles | 25 | 170,00 |
| | 50 | 174,00 |
| | 75 | 182,00 |

Izvor: Simulirani podaci.

Da bi se u izlaznim rezultatima dobio histogram zadane distribucije, klikom na ikonu **Charts** otvara se novi prozor **Frequencies: Charts** u kojem je potrebno aktivirati **Histograms: (With normal curve)**.

Konačno, da bi se u **Outputu** programa SPSS dobili željeni podaci o visini ispitanika, potrebno je kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 4.7. Prema podacima iz tablice 4.7 može se vidjeti da je 235 ispitanika dalo odgovor o svojoj visini, dok ih 4 nije odgovorio na to pitanje. Prema pokazatelju Skewness asimetrija zadane distribucije je 0,485, što upućuje na pozitivu ili desnostranu asimetriju. Pearsonova mjera asimetrije je:

$$S_k = \frac{(\bar{X} - Mo)}{\sigma} = \frac{(175,77 - 172)}{8,749} = 0,431,$$

što opet upućuje da se radi o desnostranoj asimetriji.

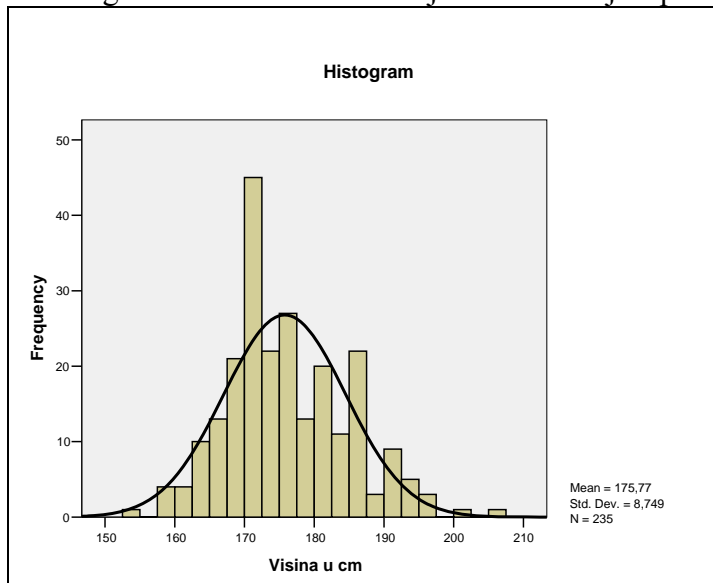
Bowleyeva mjera asimetrije je:

$$S_{kq} = \frac{Q_3 + Q_1 - 2Me}{Q_3 - Q_1} = \frac{182 + 170 - 2 \cdot 174}{182 - 170} = 0,333,$$

što potvrđuje prethodne zaključke da je zadana distribucija visine ispitanika pozitivno ili desnostrano asimetrična.

Slika 4.4.

Histogram s normalnom krivuljom distribucije ispitanika prema visini



Izvor: Simulirani podaci.

Zaobljenost zadane distribucije može se komentirati pomoću pokazatelja Kurtosis koji u ovom primjer iznosi 0,184, što upućuje na šiljatiji vrh distribucije u odnosu na normalno zaobljenu distribuciju.

Histogram zadane distribucije s ucrtanom normalnom krivuljom prikazan je na slici 4.4. te se na osnovi njega može opet zaključiti da zadana distribucija pokazuje blagu asimetriju na desnu stranu i šiljatiji vrh od normalne distribucije.



Analitički primjer 4.8.

Za varijablu (obilježje) mjesečni džeparac (v2) potrebno je odrediti:

- a) Asimetriju distribucije (Skewness);
- b) Pearsonovu mjeru asimetrije (aritmetička sredina, mod, medijan, standardna devijacija);
- c) Bowleyevu mjeru asimetrije (kvartili);
- d) Zaobljenost distribucije (Kurtosis);
- e) Na histogram distribucije ucrtati i normalnu krivulju;
- f) Komentirati rezultate!



Rješenje 4.8:

Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v2 - Mjesečni džeparac se prebacuje u polje **Variable(s)**. Klikom na ikonu **Statistics** otvara se novi prozor **Frequencies Statistics** gdje je potrebno aktivirati određene statističke veličine prema zahtjevu zadatka, a koje se odnose na odabranu varijablu mjesečni džeparac (v2).

Da bi se u izlaznim rezultatima dobio pokazatelj asimetrije zadanog niza, u spomenutom prozoru aktivira se **Skewness**. Ostali pokazatelji asimetrije mogu se izračunati pomoću aritmetičke sredine koja će se dobiti aktiviranjem veličine **Mean**. Za mod je potrebno

označiti **Mode**. Donji i gornji kvartil dobiju se označavanjem **Quartiles**. Stanardna devijacija dobije se aktiviranjem **Std. deviation**, a mjera zaobljenosti aktiviranjem veličine **Kurtosis**.

Da bi se u izlaznim rezultatima dobio histogram zadane distribucije, klikom na ikonu **Charts** otvara se novi prozor **Frequencies: Chatrs** u kojem je potrebno aktivirati **Histograms: (With normal curve)**.

Tablica 4.8.

Statistički podaci o džeparcu ispitanika

Statistics

Mjesečni džeparac

| | | |
|------------------------|---------|-----------|
| N | Valid | 173 |
| | Missing | 66 |
| Mean | | 371,5491 |
| Median | | 300,0000 |
| Mode | | 200,00(a) |
| Std. Deviation | | 401,30549 |
| Skewness | | 4,314 |
| Std. Error of Skewness | | ,185 |
| Kurtosis | | 26,422 |
| Std. Error of Kurtosis | | ,367 |
| Percentiles | 25 | 150,0000 |
| | 50 | 300,0000 |
| | 75 | 500,0000 |

a Multiple modes exist. The smallest value is shown

Izvor: Simulirani podaci.

Konačno, da bi se u **Outputu** programa SPSS dobili željeni podaci o visini ispitanika, potrebno je kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 4.8. Prema podacima iz tablice 4.8. može se vidjeti da je 173 ispitanika dalo odgovor o svojem mjesečnom džeparcu, dok ih 66 nije odgovorio na to pitanje.

Prema pokazatelju Skewness asimetrija zadane distribucije je 4,314, što upućuje na izrazitu pozitivu ili desnostranu asimetriju. Pearsonova mjera asimetrije je:

$$S_k = \frac{(\bar{X} - Mo)}{\sigma} = \frac{(371,5491 - 200)}{401,30549} = 0,4275,$$

što opet upućuje da se radi o desnostranoj asimetriji. Ovaj rezultat treba uzeti s rezervom jer zadana distribucija nije unimodalna, tj. nema samo jedan mod.

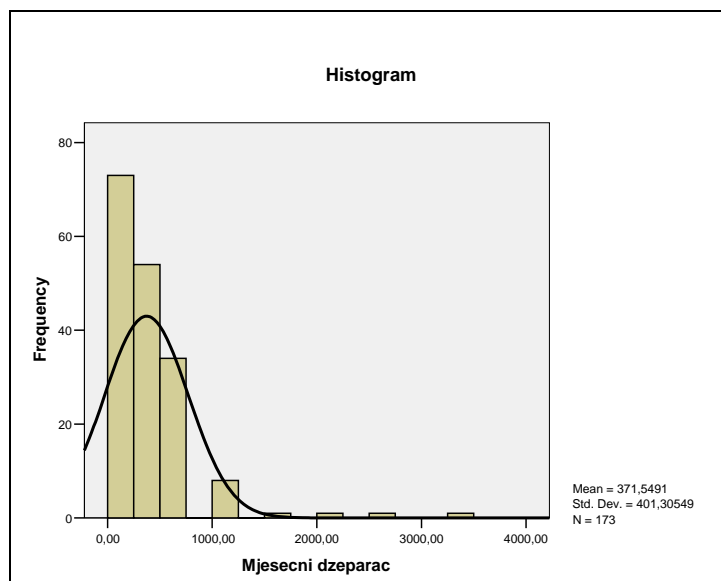
Bowleyeva mjera asimetrije je:

$$S_{kq} = \frac{Q_3 + Q_1 - 2Me}{Q_3 - Q_1} = \frac{500 + 150 - 2 \cdot 300}{500 - 150} = 0,143,$$

što potvrđuje prethodne zaključke da je zadana distribucija džeparca ispitanika pozitivno ili desnostrano asimetrična. Ovaj pokazatelj ne mjeri konkretnu asimetriju ekstremnih 25% najvećih i najmanjih vrijednosti distribucije, koji u ovom slučaju prema Skewness pokazatelju upućuju na izrazito desnu tj. pozitivnu asimetriju. Zaobljenost zadane distribucije može se komentirati pomoću pokazatelja Kurtosis koji u ovom primjer iznosi 26,422, što upućuje na izrazito šiljati vrh distribucije u odnosu na normalno zaobljenu distribuciju.

Slika 4.5.

Histogram s normalnom krivuljom distribucije ispitanika prema visini



Izvor: Simulirani podaci.

Histogram zadane distribucije s ucrtanom normalnom krivuljom prikazan je na slici 4.5. te se na osnovi njega potvrđuju ranije doneseni zaključci da zadana distribucija pokazuje veliku asimetriju na desnu stranu i izrazito šiljatiji vrh od normalne distribucije.



Analitički primjer 4.9.

Za varijablu (obilježje) broj cipela (v2) potrebno je odrediti:

- Asimetriju distribucije (Skewness);
- Pearsonovu mjeru asimetrije (aritmetička sredina, mod, medijan, standardna devijacija);
- Bowleyevu mjeru asimetrije (kvartili);
- Zaobljenost distribucije (Kurtosis);
- Na histogram distribucije ucrtati i normalnu krivulju;
- Komentirati rezultate!



Rješenje 4.9:

Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v2 - Broj cipela se prebacuje u polje **Variable(s)**. Klikom na ikonu **Statistics** otvara se novi prozor **Frequencies Statistics** gdje je potrebno aktivirati određene statističke veličine prema zahtjevu zadatka, a koje se odnose na odabranu varijablu broj cipela (v2).

Da bi se u izlaznim rezultatima dobio pokazatelj asimetrije zadanog niza, u spomenutom prozoru aktivira se **Skewness**. Ostali pokazatelji asimetrije mogu se izračunati pomoću aritmetičke sredine koja će se dobiti aktiviranjem veličine **Mean**. Za mod je potrebno označiti **Mode**. Donji i gornji kvartil dobiju se označavanjem **Quartiles**. Standardna devijacija dobije se aktiviranjem **Std. deviation**, a mjera zaobljenosti aktiviranjem veličine **Kurtosis**.

Da bi se u izlaznim rezultatima dobio histogram zadane distribucije, klikom na ikonu **Charts** otvara se novi prozor **Frequencies: Chatrs** u kojem je potrebno aktivirati **Histograms: (With normal curve)**.

Konačno, da bi se u **Outputu** programa SPSS dobili željeni podaci o broju cipela ispitanika, potrebno je kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 4.9.

Tablica 4.9.

Statistički podaci o broju cipela ispitanika

| Statistics | | |
|------------------------|---------|-------|
| Broj cipela | | |
| N | Valid | 237 |
| | Missing | 2 |
| Mean | | 40,86 |
| Median | | 40,00 |
| Mode | | 39 |
| Std. Deviation | | 3,078 |
| Skewness | | ,511 |
| Std. Error of Skewness | | ,158 |
| Kurtosis | | -,334 |
| Std. Error of Kurtosis | | ,315 |
| Percentiles | 25 | 39,00 |
| | 50 | 40,00 |
| | 75 | 43,00 |

Izvor: Simulirani podaci.

Prema podacima iz tablice 4.9. može se vidjeti da je 237 ispitanika dalo odgovor o svom broju cipela, dok ih 2 nije odgovorio na to pitanje.

Prema pokazatelju Skewness asimetrija zadane distribucije je 0,511, što upućuje na blagu pozitivu ili desnostranu asimetriju. Pearsonova mjera asimetrije je:

$$S_k = \frac{(\bar{X} - Mo)}{\sigma} = \frac{(40,86 - 39)}{3,078} = 0,604,$$

što opet upućuje da se radi o desnostranoj asimetriji.

Bowleyeva mjera asimetrije je:

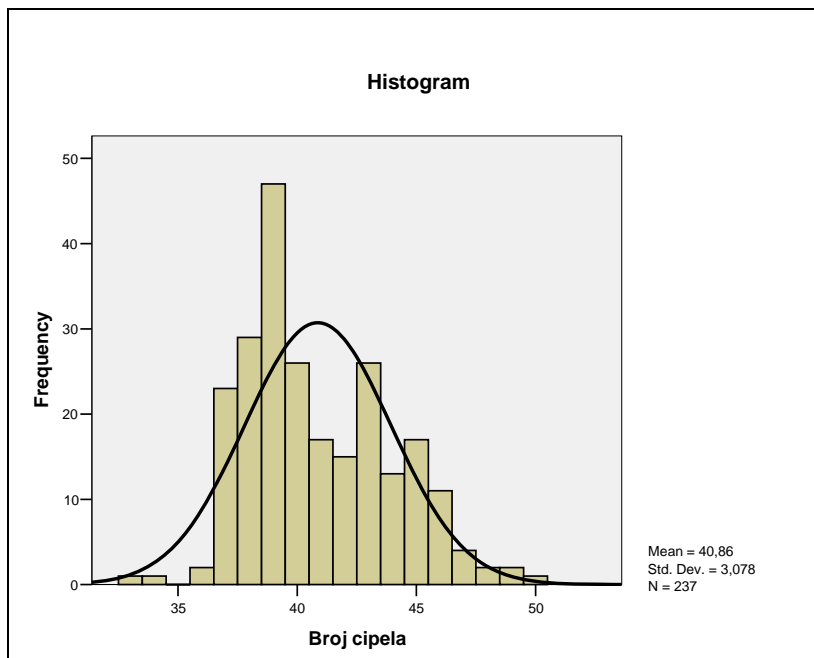
$$S_{kq} = \frac{Q_3 + Q_1 - 2Me}{Q_3 - Q_1} = \frac{43 + 39 - 2 \cdot 40}{43 - 39} = 0,5,$$

što potvrđuje prethodne zaključke da je zadana distribucija o broju cipela ispitanika pozitivno ili desnostrano asimetrična.

Zaobljenost zadane distribucije može se komentirati pomoću pokazatelja Kurtosis koji u ovom primjer iznosi -0,334, što upućuje na blago tupi vrh distribucije u odnosu na normalno zaobljenu distribuciju.

Slika 4.6.

Histogram s normalnom krivuljom distribucije ispitanika prema broju cipela



Izvor: Simulirani podaci.

Histogram zadane distribucije s ucrtanom normalnom krivuljom prikazan je na slici 4.6 te se na osnovi njega može opet zaključiti da zadana distribucija pokazuje laganu asimetriju na desnu stranu i blago tupi vrh u odnosu na normalnu distribuciju.

5. VJEROJATNOST

5.1. Klasična definicija vjerojatnosti

Slučajni događaj je takav događaj koji se može, ali ne mora realizirati, tj. realizira se uz određenu vjerojatnost.

Vjerojatnost *realizacije slučajnog događaja "A"* jednaka je omjeru broja povoljnih ishoda i svih mogućih ishoda:

$$P(A) = \frac{m(A)}{n}, \quad 0 \leq P \leq 1, \quad (5.1)$$

gdje je:

$m(A)$ - broj svih povoljnih ishoda za događaj A

n - broj svih mogućih ishoda.

Za siguran događaj: $\Rightarrow P = 1$ (obrat ne vrijedi).

Za nemoguć događaj: $\Rightarrow P = 0$ (obrat ne vrijedi).

Vjerojatnost da se **slučajni događaj "A"** *ne realizira* jednaka je omjeru broja nepovoljnih ishoda i svih mogućih ishoda:

$$P(\bar{A}) = \frac{n - m(A)}{n}, \quad 0 \leq P \leq 1, \quad (5.2)$$

gdje je:

$n - m(A)$ - broj svih nepovoljnih ishoda za događaj A

n - broj svih mogućih ishoda.

Vrijedi da je:

$$P(A) + P(\bar{A}) = 1. \quad (5.3)$$

Za ovakav slučaj kada je unaprijed poznat broj svih povoljnih/nepovoljnih i ukupnih ishoda izračunata vjerojatnost se naziva **vjerojatnost "a priori"**.

Ako vjerojatnost realizacije slučajnog događaja A nije poznata unaprijed, može se izračunati tzv. **vjerojatnost "a posteriori"**:

$$P(A) = p \lim \frac{m(A)}{n}, \quad (5.4)$$

gdje je:

$p \lim$ - čita se "granična vrijednost po vjerojatnosti" (da se razlikuje od limesa u linearnoj algebri).

5.2. Teorem o zbrajanju vjerojatnosti

Za međusobno isključive događaje (tj. one događaje koji ne mogu nastupiti istodobno, odnosno čiji su pripadni skupovi elementarnih događaja disjunktni), vjerojatnost realizacije jednog **ili** drugog događaja jednaka je zbroju vjerojatnosti realizacije jednog i vjerojatnosti realizacije drugog događaja.

$$P(A \text{ ili } B) = P(A) + P(B). \quad (5.5)$$

Za događaje koji se međusobno ne isključuju, vjerojatnost realizacije jednog **ili** drugog događaja je:

$$P(A \text{ ili } B) = P(A) + P(B) - P(AB), \quad (5.6)$$

gdje je:

$P(AB)$ -vjerojatnost da slučajni događaji A i B nastupe istovremeno.

Analitički primjer 5.1.

U posudi se nalazi 400 kuglica, od toga 100 crvenih, 140 plavih i 160 zelenih.

a) Kolika je vjerojatnost da će slučajnim izvlačenjem 1 kuglice biti izvučena kuglica zelena boje?

b) Kolika je vjerojatnost da će slučajno izvučena kuglica biti crvena ili zelena?

Rješenje 5.1:

U ovom primjeru zadano je:

$n = 400$; što predstavlja broj ukupnih ishoda;

$m(C) = 100$; što predstavlja broj povoljnih ishoda za crvenu kuglicu;

$m(P) = 140$; što predstavlja broj povoljnih ishoda za plavu kuglicu;

$m(Z) = 160$; što predstavlja broj povoljnih ishoda za zelenu kuglicu.

a) Tražena vjerojatnost da će slučajnim izvlačenjem 1 kuglice biti izvučena kuglica zelena boje jednaka je omjeru broja povoljnih ishoda, tj. broja zelenih kuglica i svih mogućih ishoda, tj. svih kuglica:

$$P(Z) = \frac{m(Z)}{n} = \frac{160}{400} = 0,4.$$

b) Za traženu vjerojatnost da će slučajno izvučena kuglica biti crvena ili zelena primjenjuje se teorem o zbrajanju vjerojatnosti za međusobno isključive događaje jer na 1 izvlačenje ne može istovremeno biti izvučena kuglica crvene i zelene boje (događaji ne mogu nastupiti istodobno):

$$P(C \text{ ili } Z) = \frac{m(C)}{n} + \frac{m(Z)}{n} = \frac{100}{400} + \frac{160}{400} = 0,65.$$

5.3. Teorem o množenju vjerojatnosti

Ako se događaji A i B **međusobno ne isključuju**, tj. vjerojatnost realizacije jednog ne ovisi o vjerojatnosti realizacije drugog događaja (nezavisni događaji), tada je vjerojatnost istovremene realizacije događaja A i B jednaka produktu vjerojatnosti realizacije događaja A i događaja B:

$$P(A \text{ i } B) = P(A) \cdot P(B). \quad (5.7)$$

5.4. Uvjetna vjerojatnost

Ako je realizacija događaja A uvjetovana prethodnom realizacijom događaja B, vrijedi da je **uvjetna ili kondicionalna vjerojatnost** događaja A:

$$P(A/B) = \frac{P(AB)}{P(B)}, \quad P(B) > 0, \quad (5.8)$$

odnosno:

$$P(B/A) = \frac{P(AB)}{P(A)}, \quad P(A) > 0. \quad (5.9)$$

Ako su **događaji nezavisni**, vrijedi da je:

$$P(AB) = P(A) \cdot P(B) \Rightarrow P(A/B) = \frac{P(A) \cdot P(B)}{P(A)} = P(B) \quad (5.10)$$

Bayesov teorem:

Ako se događaj B realizira istovremeno kada nastupi jedan od n disjunktih događaja A_1, A_2, \dots, A_n za koje je $\sum_{i=1}^n P(A_i) = 1$ onda se vjerojatnost događaja B dobiva po formuli potpune vjerojatnosti:

$$P(B) = \sum_{i=1}^n P(A_i) \cdot P(B/A_i) \quad (5.11)$$

$$P(A_i/B) = \frac{P(A_i) \cdot P(B/A_i)}{\sum_{i=1}^n P(A_i) \cdot P(B/A_i)}, \quad \sum_{i=1}^n P(A_i) = 1. \quad (5.12)$$



Analitički primjer 5.2.

Potrebno je izračunati vjerojatnost da odabrani ispitanik bude ženskog spola (varijabla: Spol (v1)).



Rješenje 5.2:

Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v1 - Spol se prebacuje u polje **Variable(s)**.

Konačno, da bi se u **Outputu** programa **SPSS** dobili željeni podaci o spolu ispitanika, potrebno je kliknuti na **OK**. Rezultat je prikazan u tablici 5.1.

Tablica 5.1.

Podaci o spolu ispitanika

| | | Spol | | | |
|---------|--------|-----------|---------|---------------|--------------------|
| | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Musko | 94 | 39,3 | 39,5 | 39,5 |
| | Zensko | 144 | 60,3 | 60,5 | 100,0 |
| | Total | 238 | 99,6 | 100,0 | |
| Missing | System | 1 | ,4 | | |
| Total | | 239 | 100,0 | | |

Izvor: Simulirani podaci.

Prema podacima iz tablice 5.1. u stupcu *Valid Percent* može se vidjeti da je 60,5% ispitanika ženskog spola, odnosno vjerojatnost da ispitanik bude ženskog spola je 0,605.



Analitički primjer 5.3.

Za varijablu grupirana visina (v1) na sljedeći način:

Range: Lowest through

- 160 ⇒ 160
- 170 ⇒ 170
- 180 ⇒ 180
- 190 ⇒ 190
- 200 ⇒ 200
- 210 ⇒ 210

potrebno je izračunati vjerojatnost da ispitanik bude visok između 171 i 180 cm.



Rješenje 5.3:

Na glavnom izborniku potrebno je izabrati ikonu *Analyze*, a na njezinu padajućem izborniku *Descriptive Statistics* i *Frequencies*.

U prozoru *Frequencies* pomoću odgovarajuće strelice varijabla v1 - Grupirana visina se prebacuje u polje *Variable(s)*.

Konačno, da bi se u *Outputu* programa SPSS dobili željeni podaci o visini ispitanika, potrebno je kliknuti na **OK**. Rezultat je prikazan u tablici 5.2.

Tablica 5.2.

Podaci o visini ispitanika
Grupirana visina

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------|--------|-----------|---------|---------------|--------------------|
| Valid | 160,00 | 7 | 2,9 | 3,0 | 3,0 |
| | 170,00 | 62 | 25,9 | 26,4 | 29,4 |
| | 180,00 | 103 | 43,1 | 43,8 | 73,2 |
| | 190,00 | 50 | 20,9 | 21,3 | 94,5 |
| | 200,00 | 11 | 4,6 | 4,7 | 99,1 |
| | 210,00 | 2 | ,8 | ,9 | 100,0 |
| | Total | 235 | 98,3 | 100,0 | |
| Missing | System | 4 | 1,7 | | |
| Total | | 239 | 100,0 | | |

Izvor: Simulirani podaci.

Prema podacima iz tablice 5.2. u stupcu *Valid Percent* može se vidjeti da je 43,8% ispitanika s visinom između 171 i 180 cm, odnosno vjerojatnost da ispitanik bude visok između 171 i 180 cm je 0,438.



Analitički primjer 5.4.

Potrebno je izračunati vjerojatnost da se slučajno odabere osoba kojoj su roditelji u braku.



Rješenje 5.4:

Na glavnom izborniku potrebno je izabrati ikonu *Analyze*, a na njezinu padajućem izborniku *Descriptive Statistics* i *Frequencies*.

U prozoru *Frequencies* pomoću odgovarajuće strelice varijabla v1 - Bračno stanje roditelja se prebacuje u polje *Variable(s)*.

Konačno, da bi se u *Outputu* programa SPSS dobili željeni podaci o bračnom stanju roditelja ispitanika, potrebno je kliknuti na *OK*. Rezultat je prikazan u tablici 5.3.

Tablica 5.3.

Podaci o bračnom stanju roditelja ispitanika

Bracno stanje roditelja

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|-------------------------|-----------|---------|---------------|--------------------|
| Valid | U braku | 215 | 90,0 | 90,0 | 90,0 |
| | Rastavljeni | 10 | 4,2 | 4,2 | 94,1 |
| | Umro jedan od roditelja | 11 | 4,6 | 4,6 | 98,7 |
| | Ostalo | 3 | 1,3 | 1,3 | 100,0 |
| | Total | 239 | 100,0 | 100,0 | |

Izvor: Simulirani podaci.

Prema podacima iz tablice 5.3. u stupcu *Valid Percent* može se vidjeti da 90% ispitanika ima roditelje u braku, odnosno vjerojatnost da ispitanik ima roditelje u braku je 0,9.



Analitički primjer 5.5.

Potrebno je izračunati vjerojatnost da se slučajno odabere osoba koja pije alkoholna pića samo u posebnim prigodama.

**Rješenje 5.5:**

Na glavnom izborniku potrebno je izabrati ikonu *Analyze*, a na njezinu padajućem izborniku *Descriptive Statistics* i *Frequencies*.

U prozoru *Frequencies* pomoću odgovarajuće strelice varijabla v1 - o Konzumiranju alkohola se prebacuje u polje *Variable(s)*.

Konačno, da bi se u *Outputu* programa *SPSS* dobili željeni podaci o konzumaciji alkohola ispitanika, potrebno je kliknuti na *OK*. Rezultat je prikazan u tablici 5.4.

Tablica 5.4.

Podaci o konzumiranju alkohola ispitanika

Konzumirate li alkohol

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------|--------------------------|-----------|---------|---------------|--------------------|
| Valid | Nikada | 54 | 22,6 | 22,8 | 22,8 |
| | Samo uz obroke | 9 | 3,8 | 3,8 | 26,6 |
| | U posebnim prilikama | 165 | 69,0 | 69,6 | 96,2 |
| | Cesto i u svakoj prigodi | 9 | 3,8 | 3,8 | 100,0 |
| | Total | 237 | 99,2 | 100,0 | |
| Missing | System | 2 | ,8 | | |
| Total | | 239 | 100,0 | | |

Izvor: Simulirani podaci.

Prema podacima iz tablice 5.4. u stupcu *Valid Percent* može se vidjeti da 69,6% ispitanika pije alkohol samo u posebnim prilikama, odnosno vjerojatnost da ispitanik pije alkohol samo u posebnim prilikama je 0,696.

**Analitički primjer 5.6.**

- Potrebno je izračunati vjerojatnost da ispitanik ne pije nikada alkoholna pića.
- Potrebno je izračunati vjerojatnost da ispitanik ženskog spola ne pije nikada alkoholna pića.



Rješenje 5.6:

a) Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v1 - o Konzumiranju alkohola se prebacuje u polje **Variable(s)**.

Konačno, da bi se u **Outputu** programa **SPSS** dobili željeni podaci o konzumaciji alkohola ispitanika, potrebno je kliknuti na **OK**. Rezultat je prikazan u tablici 5.4.

Prema podacima iz tablice 5.4. u stupcu *Valid Percent* može se vidjeti da 22,8% ispitanika ne pije nikada alkohol, odnosno vjerojatnost da ispitanik ne pije alkohol nikada je 0,228.

b) Da bi se izračunala vjerojatnost da ispitanik ženskog spola ne pije nikada alkoholna pića, potrebno je ostaviti u analizi samo ispitanike ženskog spola, odnosno filtrirati željeni uzorak.

Na glavnom izborniku potrebno je izabrati ikonu **Data**, a na njezinu padajućem izborniku **Select Cases**.

U prozoru **Select Cases** potrebno je aktivirati opciju **If conditions is satisfied**. Klikom na ikonu **if** otvara se prozor **Select Cases:If** gdje je potrebno odabrati varijablu v2 i postaviti uvjet da je $v2 = 2$, odnosno da je spol 2, što označava ženski spol.

Konačno, da bi se filtrirali samo ispitanici ženskog spola, potrebno je kliknuti na **Continue** i **OK**. Daljnja analiza se sada odnosi samo na ženski spol.

Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v1 - o Konzumiranju alkohola se prebacuje u polje **Variable(s)**.

Konačno, da bi se u **Outputu** programa **SPSS** dobili željeni podaci o konzumaciji alkohola ispitanika ženskog spola, potrebno je kliknuti na **OK**. Rezultat je prikazan u tablici 5.5.

Tablica 5.5.

Podaci o konzumiranju alkohola ispitanika ženskog spola
Konzumirate li alkohol

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------|--------------------------|-----------|---------|---------------|--------------------|
| Valid | Nikada | 37 | 25,7 | 26,1 | 26,1 |
| | Samo uz obroke | 3 | 2,1 | 2,1 | 28,2 |
| | U posebnim prilikama | 94 | 65,3 | 66,2 | 94,4 |
| | Cesto i u svakoj prigodi | 8 | 5,6 | 5,6 | 100,0 |
| | Total | 142 | 98,6 | 100,0 | |
| Missing | System | 2 | 1,4 | | |
| Total | | 144 | 100,0 | | |

Izvor: Simulirani podaci.

Prema podacima iz tablice 5.5 u stupcu **Valid Percent** može se vidjeti da 26,1% od ispitanika ženskog spola ne pije nikada alkohol, odnosno vjerojatnost da ispitanik ženskog spola nikada ne pije alkohol je 0,261.

5.5. Diskontinuirana slučajna varijabla

Diskontinuirana ili diskretna slučajna varijabla je takva varijabla koja na slučaj može poprimiti najviše prebrojivo mnogo vrijednosti i svaku od njih s određenom vjerojatnošću.

Zadovoljava uvjete:

- normativnost: $\sum_{i=1}^{\infty} P(x_i) = 1$
- nenegativnost: $P(x_i) \geq 0, \forall i$.

Skup svih uređenih parova vrijednosti slučajne varijable i njoj pripadajućih vjerojatnosti naziva se **distribucija vjerojatnosti slučajne varijable** X :

$$\{[x_i; P(x_i)], i = 1, 2, \dots, \infty\}. \quad (5.13)$$

Zakon po kojem svakoj vrijednosti slučajne varijable X pripada vjerojatnost $P(x_i)$ naziva se **zakon vjerojatnosti slučajne varijable X** .

Funkcija distribucije slučajne varijable X je funkcija koja daje vjerojatnost da će slučajna varijabla X poprimiti vrijednost jednaku ili manju od nekog realnog broja x_k :

$$F(x_k) = P(X \leq x_k) = \sum_{i=1}^k P(x_i) \quad (5.14)$$

Može se reći da je to **kumulativ vjerojatnosti** od $F(X=0)$ do $F(X=x_k)$. Funkcija distribucije slučajne varijable X je monotono neopadajuća funkcija.

Parametri distribucije:

- **Očekivanje slučajne varijable** je zbroj umnožaka vrijednosti varijable X i njoj pripadajućih odgovarajućih vjerojatnosti $P(x_i)$:

$$E(X) = \sum_{i=1}^{\infty} x_i \cdot P(x_i) = \mu. \quad (5.15)$$

- **Varijanca slučajne varijable** je očekivanje kvadratnog odstupanja vrijednosti varijable X od njezina očekivanja:

$$V(X) = E(X^2) - [E(X)]^2 = E(X - \mu)^2 = \sum_{i=1}^{\infty} x_i^2 \cdot P(x_i) - \mu^2. \quad (5.16)$$

5.6. Teorijske distribucije diskontinuirane slučajne varijable

5.6.1. Binomna distribucija

Ako slučajni pokus koji se ponavlja n puta ima dva moguća ishoda: *uspjeh* i *neuspjeh*, i ako je vjerojatnost nastupa ishoda *uspjeh* poznata, unaprijed utvrđena i konstantna tijekom cijelog istraživanja (iznosi p), kaže se da se diskontinuirana slučajna varijabla X ravna prema tzv. **binomnoj distribuciji**.

Vjerojatnost da se neki slučajni događaj X u n pokusa realizira x puta prema **binomnom zakonu vjerojatnosti** je:

$$P(X = x) = \binom{n}{x} \cdot p^x \cdot (1 - p)^{n-x} = \binom{n}{x} \cdot p^x \cdot q^{n-x}, \quad x = 0, 1, 2, \dots, n \quad (5.17)$$

gdje je:

n - broj ponavljanja pokusa

p - vjerojatnost realizacije slučajnog događaja (povoljnog ishoda)

q - vjerojatnost da se slučajni događaj ne realizira

x - broj povoljnih ishoda u n ponavljanja slučajnog pokusa

$n - x$ - broj nepovoljnih ishoda u n ponavljanja slučajnog pokusa.

Očekivanje slučajne varijable X kod binomne distribucije:

$$E(X) = n \cdot p. \quad (5.18)$$

Varijanca slučajne varijable X kod binomne distribucije:

$$\sigma^2 = n \cdot p \cdot q. \quad (5.19)$$

Koeficijent asimetrije slučajne varijable X kod binomne distribucije:

$$\alpha_3 = \frac{q - p}{\sqrt{n \cdot p \cdot q}}. \quad (5.20)$$

Koeficijent zaobljenosti slučajne varijable X kod binomne distribucije:

$$\alpha_4 = 3 + \frac{1 - 6 \cdot p \cdot q}{n \cdot p \cdot q}. \quad (5.21)$$

Općenito se kaže da se **slučajna varijabla X ravna po binomnoj distribuciji** koja je određena parametrima n i p :

$$X \sim B(n, p). \quad (5.22)$$

U programskom paketu SPSS postoje **funkcije gustoće vjerojatnosti i kumulativne funkcije distribucije vjerojatnosti Binomne distribucije.**

Funkcije gustoće vjerojatnosti (Probability Density Functions)

- **Transform; Compute; Function group (izabrati: PDF & Noncentral PDF); izabrati: Pdf.Binom u Numeric Expression:**

PDF.BINOM(quant,n,prob)

Rezultat ove funkcije su vjerojatnosti da broj uspjeha u n pokušaja, uz vjerojatnost realizacije $p = prob$, budu jednake nekom broju $x = quant$.

U Target Variable: potrebno je imenovati novu varijablu izračunatih vjerojatnosti po Binomnom zakonu.

Kumulativne funkcije distribucije vjerojatnosti (Cumulative Distribution Functions)

- **Transform; Compute; Function group (izabrati: CDF & Noncentral CDF); izabrati: Cdf.Binom u Numeric Expression:**

CDF.BINOM(quant,n,prob)

Rezultat ove funkcije je kumulativ vjerojatnosti da broj uspjeha u n pokušaja, uz vjerojatnost realizacije $p = prob$, bude manji ili jednak nekom broju $x = quant$.

U Target Variable: potrebno je imenovati novu varijablu izračunatih kumulativa vjerojatnosti po Binomnom zakonu.



Analitički primjer 5.7.

Vjerojatnost da je student Odsjeka za sociologiju Filozofskoga fakulteta u Splitu muškog spola je: $p = \frac{m(M)}{n} = \frac{7}{31} = 0,225806451$ (gdje je n ukupan broj studenata Odsjeka za sociologiju, a $m(M)$ broj je povoljnih ishoda, odnosno broj studenata muškog spola).

- a) Kolika je vjerojatnost da između 7 slučajno odabranih studenata budu 3 studenta muškog spola?
- b) Kolika je vjerojatnost da između 7 slučajno odabranih studenata budu najviše 2 studenta muškog spola?
- c) Kolika je vjerojatnost da između 7 slučajno odabranih studenata budu najviše 4 ili najmanje 3 studenta muškog spola?
- d) Kolika je vjerojatnost da između 7 slučajno odabranih studenata bude barem 1 student muškog spola?
- e) Kolika je vjerojatnost da između 7 slučajno odabranih studenata budu svi studenti muškog spola?

**Rješenje 5.7:**

Na glavnom izborniku potrebno je odabrati ikonu **Transform**, a na njezinu padajućem izborniku **Compute**. Nakon toga otvara se prozor **Compute Variable**, koji je prikazan na slici 5.1. Da bi se aktivirala naredba za izračun vjerojatnosti po Binomnom zakonu, potrebno je u izborniku **Function group**: odabrati **PDF & Noncentral PDF**. Nakon toga se u izborniku **Function and Special Variables**: bira **Pdf.Binom** i klikom dva puta na tu funkciju ona se pojavljuje u prozoru **Numeric Expression**: na sljedeći način: **PDF.BINOM(?,?,?)**.

U navedenoj funkciji prvi upitnik predstavlja **quant**, tj. **x** vrijednost koju poprima slučajna varijabla u ovom primjeru **X1**. Drugi upitnik funkcije predstavlja **n**, odnosno veličinu odabranog uzorka. Treći upitnik je **prob**, tj. **p** vjerojatnost realizacije slučajnog događaja. U ovom primjeru je zadano:

$$\text{quant} = X1 ; n = 7 ; \text{prob} = 0,225806451,$$

pa vrijedi da je:

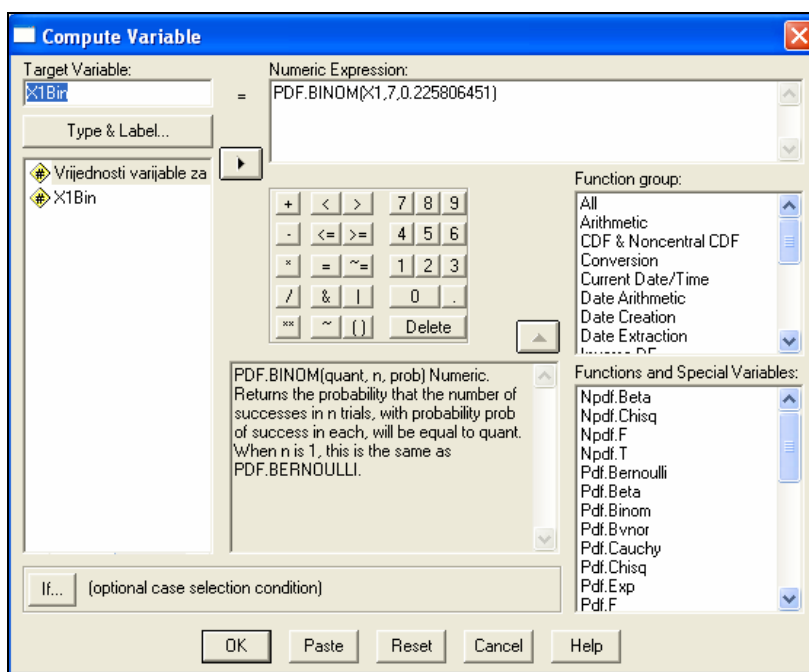
$$\text{PDF.BINOM}(X1, 7, 0.225806451),$$

što je i prikazano u prozoru **Numeric Expression** na slici 5.1.

U **Target Variable**: potrebno je imenovati novu varijablu izračunatih vjerojatnosti po Binomnom zakonu i u ovom primjeru nazvana je **X1Bin**.

Slika 5.1.

Prozor u SPSS-u za izračunavanje vjerojatnosti diskontinuirane slučajne varijable po Binomnom zakonu



Izvor: Simulirani podaci.

Konačno, da bi se u **dodatnom stupcu** programa SPSS dobili željeni podaci **vjerojatnosti slučajne varijable X_1 po Binomnoj distribuciji**, potrebno je kliknuti na **OK**. Rezultat je prikazan u tablici 5.6.

Prema rezultatima iz tablice 5.6. mogu se očitati i/ili izračunati tražene vjerojatnosti.

- a) Vjerojatnost da između 7 slučajno odabranih studenata budu 3 studenta muškog spola je:

$$p(X_1 = 3) = 0,144769.$$

- b) Vjerojatnost da između 7 slučajno odabranih studenata budu najviše 2 studenta muškog spola je:

$$\begin{aligned} p(X1 \leq 2) &= p(X1 = 0) + p(X1 = 1) + p(X1 = 2) = \\ &= 0,166704 + 0,340355 + 0,297810 = 0,805. \end{aligned}$$

Tablica 5.6.

Podaci u SPSS-u vrijednosti i vjerojatnosti diskontinuirane slučajne varijable po Binomnom zakonu

| X1 | X1Bin |
|------|---------|
| ,00 | ,166704 |
| 1,00 | ,340355 |
| 2,00 | ,297810 |
| 3,00 | ,144769 |
| 4,00 | ,042224 |
| 5,00 | ,007389 |
| 6,00 | ,000718 |
| 7,00 | ,000030 |

Izvor: Simulirani podaci.

c) Vjerojatnost da između 7 slučajno odabranih studenata budu najviše 4 ili najmanje 3 studenta muškog spola je:

$$p(X1 \leq 4 \text{ ili } X1 \geq 3) = 1.$$

Ovom vjerojatnosti su obuhvaćene sve vrijednosti slučajne varijable X1, a zbroj svih vjerojatnosti u jednoj distribuciji je 1.

d) Vjerojatnost da između 7 slučajno odabranih studenata bude barem 1 student muškog spola je:

$$\begin{aligned} p(X1 \geq 1) &= p(X1 = 1) + p(X1 = 2) + \dots + p(X1 = 7) = \\ &= 1 - p(X1 = 0) = 1 - 0,166704 = 0,8333. \end{aligned}$$

e) Vjerojatnost da između 7 slučajno odabranih studenata budu svi studenti muškog spola je:

$$p(X1 = 7) = 0,00003.$$

5.6.2. Poissonova distribucija

Ako je vjerojatnost da se dogodi neki događaj poznata, unaprijed utvrđena, konstantna i jako mala tijekom cijelog istraživanja (iznosi p), te broj pokusa teži u beskonačnost ($n \rightarrow \infty$) umjesto binomne distribucije, upotrebljava se **Poissonova distribucija**.

U praksi vrijedi da ako je:

$$n \geq 50 \text{ i } p \leq 0,10 \quad (5.23)$$

vrijedi Poissonova distribucija. Može se reći da je ta distribucija definirana za **rijetke događaje**, odnosno one događaje koji imaju veliki uzorak i malu vjerojatnost.

Zakon vjerojatnosti po Poissonovoj distribuciji je:

$$P(X = x) = \frac{(np)^x \cdot e^{-np}}{x!} = \frac{(\mu)^x \cdot e^{-\mu}}{x!}, \quad x = 0, 1, 2, \dots, \infty, \quad (5.24)$$

jer je: $\mu = n \cdot p$;

gdje je:

n - broj pokusa

p - vjerojatnost realizacije slučajnog događaja

x - broj povoljnih ishoda u n pokusa.

Očekivanje slučajne varijable X kod Poissonove distribucije:

$$E(X) = n \cdot p = \mu. \quad (5.25)$$

Varijanca slučajne varijable X kod Poissonove distribucije:

$$\sigma^2 = n \cdot p = \mu. \quad (5.26)$$

Koeficijent asimetrije slučajne varijable X kod Poissonove distribucije:

$$\alpha_3 = \frac{1}{\sqrt{\mu}}. \quad (5.27)$$

Koeficijent zaobljenosti slučajne varijable X kod Poissonove distribucije:

$$\alpha_4 = 3 + \frac{1}{\mu}. \quad (5.28)$$

Općenito se kaže da se **slučajna varijabla X ravna po Poissonovoj distribuciji** koja je određena parametrom μ :

$$X \sim P(\mu). \quad (5.29)$$

U programskom paketu SPSS postoje **funkcije gustoće vjerojatnosti i kumulativne funkcije distribucije vjerojatnosti Poissonove distribucije**.

Funkcije gustoće vjerojatnosti (Probability Density Functions)

- **Transform; Compute; Function group (izabrati: PDF & Noncentral PDF); izabrati: Pdf.Poisson u Numeric Expression:**

PDF.POISSON(quant,mean)

Rezultat ove funkcije su vjerojatnosti da broj uspjeha u n pokušaja, uz vjerojatnost realizacije $p = prob$, budu jednake nekom broju $x = quant$. Vrijedi da je očekivanje $mean = \mu = n \cdot p$.

U **Target Variable:** potrebno je imenovati novu varijablu izračunatih vjerojatnosti po Poissonovu zakonu.

Kumulativne funkcije distribucije vjerojatnosti (Cumulative Distribution Functions)

- **Transform; Compute; Function group (izabrati: CDF & Noncentral CDF); izabrati: Cdf.Poisson u Numeric Expression:**

CDF.POISSON(quant,mean)

Rezultat ove funkcije je kumulativ vjerojatnosti da broj uspjeha u n pokušaja, uz vjerojatnost realizacije $p = prob$, bude manji ili jednak nekom broju $x = quant$. Vrijedi da je očekivanje $mean = \mu = n \cdot p$.

U Target Variable: potrebno je imenovati novu varijablu izračunatih kumulativa vjerojatnosti po Poissonovu zakonu.



Analitički primjer 5.8.

Vjerojatnost da se među studentskom populacijom konzumira droga je: $p = 0,056$.

- Kolika je vjerojatnost da između 60 slučajno odabranih studenata budu 2 studenta koja konzumiraju drogu?
- Kolika je vjerojatnost da između 60 slučajno odabranih studenata barem 1 student konzumira drogu?
- Kolika je vjerojatnost da između 60 slučajno odabranih studenata najviše 10 studenata konzumira drogu?



Rješenje 5.8:

Zadana vjerojatnost slučajnog događaja da se među studentskom populacijom konzumira droga je manja od 0,10, a uzorak je u svim slučajevima veći od 50, pa se u ovom primjeru pretpostavlja da se slučajna varijabla X_2 ravna po Poissonovu zakonu.

Na glavnom izborniku potrebno je odabrati ikonu **Transform**, a na njezinu padajućem izborniku **Compute**. Nakon toga otvara se prozor **Compute Variable**. Da bi se aktivirala naredba za izračun vjerojatnosti po Poissonovu zakonu, potrebno je u izborniku **Function group**: odabrati **PDF & Noncentral PDF**. Nakon toga se u izborniku **Function and Special Variables**: bira **Pdf.Poisson** i klikom dva puta na tu funkciju ona se pojavljuje u prozoru **Numeric Expression**: na slijedeći način: **PDF.POISSON(?,?)**.

U navedenoj funkciji prvi upitnik predstavlja **quant**, tj. x vrijednost koju poprima slučajna varijabla u ovom primjeru X_2 . Drugi upitnik funkcije predstavlja **mean**, odnosno očekivanje slučajne varijable koja se po Poissonovu zakonu računa na sljedeći način:

$$n = 60 ; \text{prob} = 0,056 : \Rightarrow \mu = n \cdot p = 60 \cdot 0,056 = 3,36.$$

U ovom primjeru vrijedi da je:

$$quant = X2; \mu = 3,36,$$

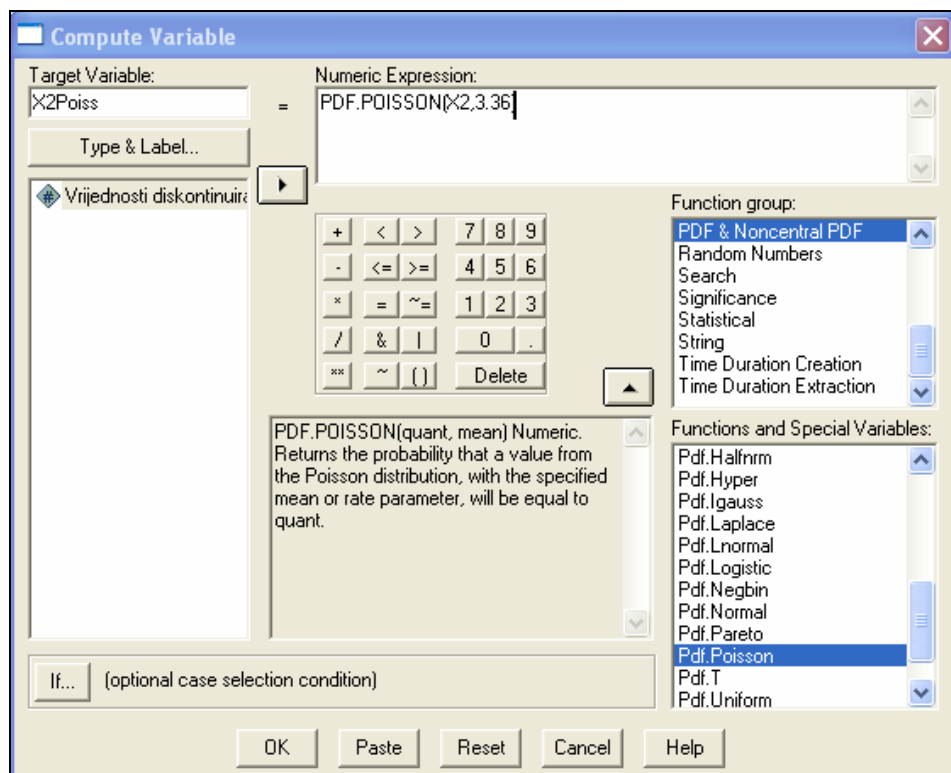
pa se može upisati:

PDF.POISSON(X2,3.36),

što je i prikazano u prozoru *Numeric Expression* na slici 5.2.

Slika 5.2.

Prozor u SPSS-u za izračunavanje vjerojatnosti diskontinuirane slučajne varijable po Poissonovu zakonu



Izvor: Simulirani podaci.

U ***Target Variable:*** potrebno je imenovati novu varijablu izračunatih vjerojatnosti po Poissonovu zakonu i u ovom primjeru nazvana je ***X2Poiss***.

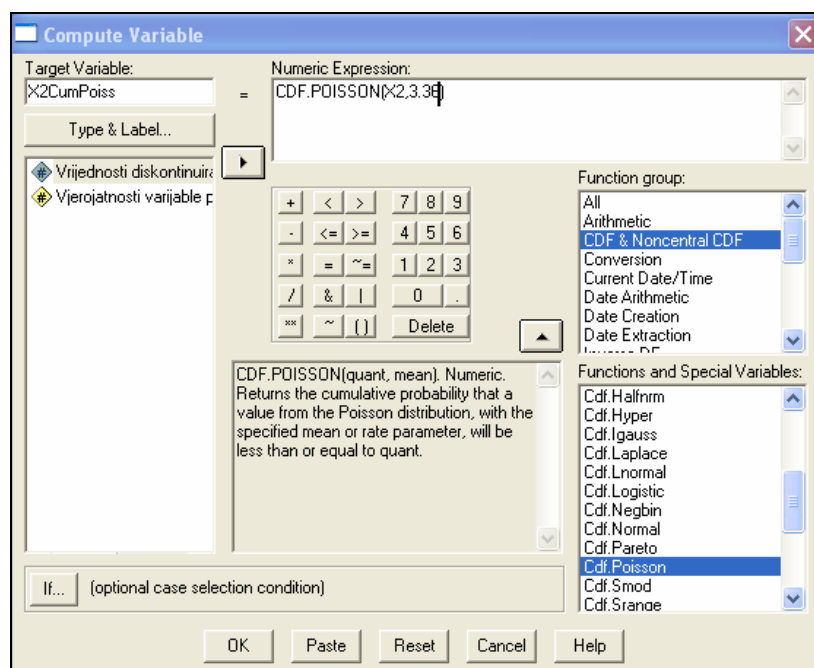
Konačno, da bi se u *dodatnom stupcu* programa SPSS dobili željeni podaci *vjerojatnosti slučajne varijable X_2 po Poissonovoj distribuciji*, potrebno je kliknuti na **OK**. Rezultat je prikazan u tablici 5.7.

Da bi se dobile kumulativne vjerojatnosti, potrebno je ponoviti sličan postupak.

Na glavnom izborniku potrebno je odabrati ikonu **Transform**, a na njezinu padajućem izborniku **Compute**. Nakon toga otvara se prozor **Compute Variable**. Da bi se aktivirala naredba za izračun kumulativnih vjerojatnosti po Poissonovu zakonu, potrebno je u izborniku **Function group**: odabrati **CDF & Noncentral CDF**. Nakon toga se u izborniku **Function and Special Variables**: bira **Cdf.Poisson** i klikom dva puta na tu funkciju ona se pojavljuje u prozoru **Numeric Expression**: na sljedeći način: **CDF.POISSON(?,?)**.

Slika 5.3.

Prozor u SPSS-u za izračunavanje kumulativnih vjerojatnosti diskontinuirane slučajne varijable po Poissonovu zakonu



Izvor: Simulirani podaci.

U navedenoj funkciji prvi upitnik opet predstavlja *quant*, tj. *x* vrijednost koju poprima slučajna varijabla u ovom primjeru *X2*. Drugi upitnik funkcije predstavlja *mean*, odnosno očekivanje slučajne varijable:

$$quant = X2; \mu = 3,36,$$

pa se može upisati:

CDF.POISSON(X2,3.36),

što je i prikazano u prozoru *Numeric Expression* na slici 5.3.

U *Target Variable*: potrebno je imenovati novu varijablu izračunatih vjerojatnosti po Poissonovu zakonu i u ovom primjeru nazvana je *X2CumPoiss*.

Konačno, da bi se u *dodatnom stupcu* programa *SPSS* dobili željeni podaci *kumulativnih vjerojatnosti slučajne varijable X2 po Poissonovoj distribuciji*, potrebno je kliknuti na *OK*. Rezultat je prikazan u tablici 5.7.

Tablica 5.7.

Podaci u SPSS-u vrijednosti, vjerojatnosti i kumulativne vjerojatnosti diskontinuirane slučajne varijable po Poissonovu zakonu

| X2 | X2Poiss | X2CumPoiss |
|-------|---------|------------|
| ,00 | ,034735 | ,034735 |
| 1,00 | ,116710 | ,151446 |
| 2,00 | ,196074 | ,347519 |
| 3,00 | ,219602 | ,567122 |
| 4,00 | ,184466 | ,751588 |
| 5,00 | ,123961 | ,875549 |
| 6,00 | ,069418 | ,944967 |
| 7,00 | ,033321 | ,978288 |
| 8,00 | ,013995 | ,992283 |
| 9,00 | ,005225 | ,997507 |
| 10,00 | ,001755 | ,999263 |
| 11,00 | ,000536 | ,999799 |
| 12,00 | ,000150 | ,999949 |
| 13,00 | ,000039 | ,999988 |

Izvor: Simulirani podaci.

Prema rezultatima iz tablice 5.7. (prikazano je samo 13 vrijednosti od ukupno 60 vrijednosti zadane diskontinuirane slučajne varijable X_2) mogu se očitati i/ili izračunati tražene vjerojatnosti.

- a) Vjerojatnost da između 60 slučajno odabranih studenata budu 2 studenta koja konzumiraju drogu je:

$$p(X_1 = 2) = 0,196074.$$

- b) Vjerojatnost da između 60 slučajno odabranih studenata barem 1 student konzumira drogu je:

$$\begin{aligned} p(X_1 \geq 1) &= p(X_1 = 1) + p(X_1 = 2) + \dots + p(X_1 = 60) = \\ &= 1 - p(X_1 = 0) = 1 - 0,034735 = 0,9653. \end{aligned}$$

Ovdje je iskorišteno svojstvo da je zbroj svih vjerojatnosti u distribuciji 1.

- c) Kolika je vjerojatnost da između 60 slučajno odabranih studenata najviše 10 studenata konzumira drogu?

$$\begin{aligned} p(X_1 \leq 10) &= p(X_1 = 0) + p(X_1 = 1) + \dots + p(X_1 = 10) = \\ &= 0,999263. \end{aligned}$$

Ovdje je tražena vjerojatnost jednostavno očitana iz rezultata stupca kumulativnih vjerojatnosti Poissonove distribucije. Jednak rezultat bi se dobio postupnim zbrajanjem vjerojatnosti iz stupca vjerojatnosti Poissonove distribucije.

5.6.3. Dvodimenzionalna diskontinuirana distribucija

Diskontinuirana slučajna varijabla X može poprimiti vrijednosti $x_1, x_2, x_3, \dots, x_n$, a diskontinuirana slučajna varijabla Y može istovremeno poprimiti vrijednosti $y_1, y_2, y_3, \dots, y_m$.

Vjerojatnost da slučajna varijabla X poprimi vrijednost x_i , a istovremeno slučajna varijabla Y poprimi vrijednost y_j je:

$$P(X = x_i, Y = y_j) = P(x_i, y_j) \quad (5.30)$$

Za svaku distribuciju vjerojatnosti mora biti ispunjen uvjet normativnosti:

$$\sum_{i=1}^n \sum_{j=1}^m P(x_i, y_j) = 1 \quad (5.31)$$

Skup svih uređenih parova $\{(x_i, y_j), P(x_i, y_j)\}$ dvodimenzionalna je distribucija vjerojatnosti slučajnih varijabli (X, Y) .

5.6.4. Marginalne distribucije diskontinuirane slučajne varijable

Marginalna distribucija slučajne varijable X skup je uređenih parova odgovarajućih vrijednosti varijable X i njima pripadajućih vjerojatnosti da slučajna varijabla X poprimi neku vrijednost x_i bez obzira na to koju će vrijednost poprimiti slučajna varijabla Y .

Marginalna distribucija slučajne varijable Y skup je uređenih parova odgovarajućih vrijednosti varijable Y i njima pripadajućih vjerojatnosti da slučajna varijabla Y poprimi neku vrijednost y_j bez obzira na to koju će vrijednost poprimiti slučajna varijabla X .

Marginalna vjerojatnost slučajne varijable X je:

$$P(x_i \cdot) = \sum_{j=1}^m P(x_i, y_j), \quad (5.32)$$

i obično se u dvostrukoj statističkoj tablici nalazi u zbirnom stupcu (ako se slučajna varijabla X mijenja po redcima).

Marginalna vjerojatnost slučajne varijable Y je:

$$P(\cdot y_j) = \sum_{i=1}^n P(x_i, y_j), \quad (5.33)$$

i obično se u dvostrukoj statističkoj tablici nalazi u zbirnom retku (ako se slučajna varijabla Y mijenja po stupcima).

5.6.5. Uvjetne distribucije diskontinuirane slučajne varijable

Vjerojatnost da slučajna varijabla X poprimi neku vrijednost x_i pod uvjetom da je varijabla Y poprimila neku vrijednost y_j uvjetna je ili kondicionalna vjerojatnost:

$$P(x_i / y_j) = \frac{P(x_i, y_j)}{P(y_j)}, \quad P(y_j) > 0. \quad (5.34)$$

Slično, vjerojatnost da slučajna varijabla Y poprimi neku vrijednost y_j pod uvjetom da je varijabla X poprimila neku vrijednost x_i uvjetna je ili kondicionalna vjerojatnost:

$$P(y_j / x_i) = \frac{P(x_i, y_j)}{P(x_i)}, \quad P(x_i) > 0. \quad (5.35)$$

Ako su **slučajne varijable nezavisne**, primjenjuje se multiplikacijski teorem:

$$P(x_i, y_j) = P(x_i) \cdot P(y_j), \quad (5.36)$$

pa vrijedi da je:

$$P(x_i / y_j) = P(x_i), \quad (5.37)$$

$$P(y_j / x_i) = P(y_j). \quad (5.38)$$



Analitički primjer 5.9.

a) Zadatak je izračunati dvodimenzionalnu distribuciju vjerojatnosti: za

- varijablu (obilježje) *grupirani džeparac (v1)* koja se mijenja po redcima (row) i za

- varijablu (obilježje) *grupirana ocjena u srednjoj školi (v2)* koja se mijenja po stupcima (column).

b) Komentirati izračunate vjerojatnosti dvodimenzionalne distribucije!

- c) Komentirati izračunate vjerojatnosti marginalne distribucije!
- d) Uz pretpostavku da promatrane slučajne varijable nisu nezavisne potrebno je izračunati uvjetnu distribuciju vjerojatnosti: $P(p4 / p3 = 400)$!

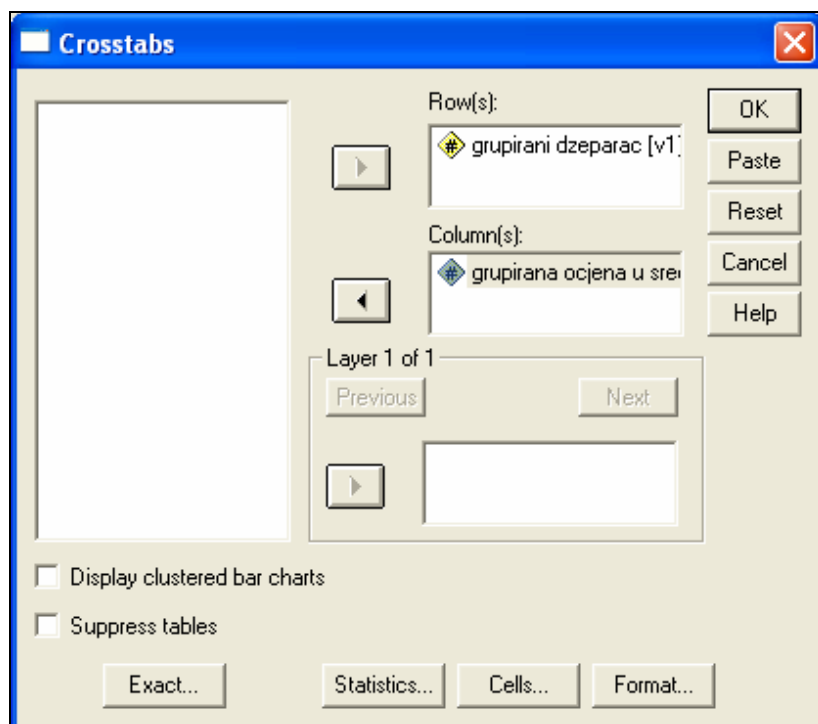


Rješenje 5.9:

- a) Na glavnom izborniku potrebno je odabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Crosstabs**. Nakon toga otvara se prozor **Crosstabs**.

Slika 5.4.

Prozor Crosstabs za izračunavanje dvodimenzionalnih vrijednosti zadanih slučajnih varijabli



Izvor: Simulirani podaci.

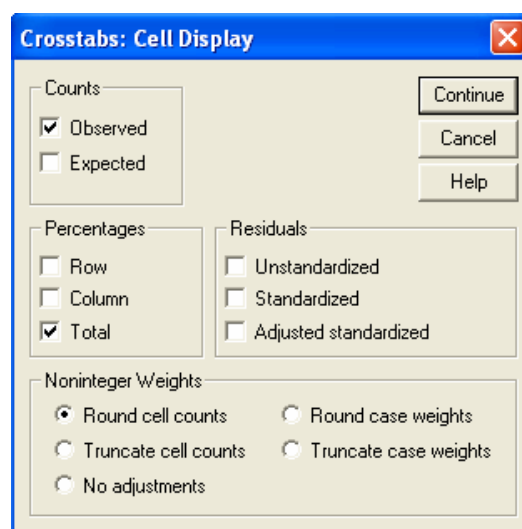
Odgovarajućim strelicama prema zahtjevu zadatka potrebno je varijablu (obilježje) *grupirani džeparac (v1)* prebaciti u prozor za retke

Row(s) i varijablu (obilježje) *grupirana ocjena u srednjoj školi (v2)* prebaciti u prozor za stupce Column(s) kako je prikazano na slici 5.4.

Klikom na ikonu **Cells...** otvara se prozor **Crosstabs: Cell Display** gdje je uz postojeće originalne frekvencije (**Counts**) potrebno aktivirati **Percentages; Total** kako je prikazano na slici 5.5.

Da bi se u **Outputu** programa **SPSS** dobili traženi podaci potrebno je kliknuti na **Continue** i **OK**.

Slika 5.5.



Izvor: Simulirani podaci.

Rezultati su prikazani u tablici 5.8.

b) Izračunate vjerojatnosti dvodimenzionalne distribucije u tablici 5.8. izražene su u postotcima.

Vjerojatnost u prvom retku i prvom stupcu numeričkog dijela tablice iznosi 3,5%, što znači da vjerojatnost da ispitanik ima džeparac do 200 kn i ocjenu iz srednje škole 3 iznosi 0,035.

Vjerojatnost u prvom retku i drugom stupcu numeričkog dijela tablice iznosi 22,1%, što znači da vjerojatnost da ispitanik ima džeparac do 200 kn i ocjenu iz srednje škole 4 iznosi 0,221.

Tablica 5.8.

Dvostruka tablica obilježja džeparca i prosječne ocjene u srednjoj školi

| grupirani džeparac * grupirana ocjena u srednjoj školi Crosstabulation | | | | | | |
|--|---------|------------|-----------------------------------|-------|-------|--------|
| | | | grupirana ocjena u srednjoj školi | | | Total |
| | | | 3,00 | 4,00 | 5,00 | |
| grupirani džeparac | 200,00 | Count | 6 | 38 | 29 | 73 |
| | | % of Total | 3,5% | 22,1% | 16,9% | 42,4% |
| | 400,00 | Count | 14 | 20 | 18 | 52 |
| | | % of Total | 8,1% | 11,6% | 10,5% | 30,2% |
| | 600,00 | Count | 5 | 14 | 10 | 29 |
| | | % of Total | 2,9% | 8,1% | 5,8% | 16,9% |
| | 800,00 | Count | 1 | 3 | 2 | 6 |
| | | % of Total | ,6% | 1,7% | 1,2% | 3,5% |
| | 1000,00 | Count | 0 | 2 | 5 | 7 |
| | | % of Total | ,0% | 1,2% | 2,9% | 4,1% |
| | 3500,00 | Count | 0 | 3 | 2 | 5 |
| | | % of Total | ,0% | 1,7% | 1,2% | 2,9% |
| Total | | Count | 26 | 80 | 66 | 172 |
| | | % of Total | 15,1% | 46,5% | 38,4% | 100,0% |

Izvor: Simulirani podaci.

Vjerojatnost u četvrtom retku i drugom stupcu numeričkog dijela tablice iznosi 1,7%, što znači da vjerojatnost da ispitanik ima džeparac između 600 kn i 800 kn i ocjenu iz srednje škole 4 iznosi 0,071.

Posljednja vjerojatnost numeričkog dijela tablice, tj. vjerojatnost u šestom retku i trećem stupcu iznosi 2,9%, što znači da vjerojatnost da ispitanik ima džeparac između 1000 kn i 3500 kn i ocjenu iz srednje škole 5 iznosi 0,012.

c) Izračunate vjerojatnosti marginalne distribucije nalaze se u zbirnom retku i zbirnom stupcu tablice 5.8.

Na primjer, druga marginalna vjerojatnost u zbirnom stupcu iznosi 30,2%, tj. vjerojatnost da ispitanik ima džeparac između 200 kn i 400 kn bez obzira na prosječnu ocjenu u srednjoj školi iznosi 0,302.

Isto tako treća marginalna vjerojatnost u zbirnom retku iznosi 38,4%, tj. vjerojatnost da ispitanik ima prosječnu ocjenu u srednjoj školi 5 bez obzira na visinu džeparca iznosi 0,384.

Zbroj svih vjerojatnosti u izračunatoj dvodimenzionalnoj distribuciji vjerojatnosti je 1, odnosno 100%.

d) Uz pretpostavku da promatrane slučajne varijable nisu nezavisne, uvjetna distribucija vjerojatnosti: $P(v_2 / v_1 = 400)$ izračunata je na sljedeći način i prikazana je u tablici 5.9:

$$P(v_2 = 3 / v_1 = 400) = \frac{P(v_2 = 3, v_1 = 400)}{P(v_1 = 400)} = \frac{0,081}{0,302} = 0,2682,$$

$$P(v_2 = 4 / v_1 = 400) = \frac{P(v_2 = 4, v_1 = 400)}{P(v_1 = 400)} = \frac{0,116}{0,302} = 0,3841,$$

$$P(v_2 = 5 / v_1 = 400) = \frac{P(v_2 = 5, v_1 = 400)}{P(v_1 = 400)} = \frac{0,105}{0,302} = 0,3477.$$

Tablica 5.9.

Uvjetna distribucija vjerojatnosti

| v_2 | 3 | 4 | 5 | Σ |
|----------------------|--------|--------|--------|----------|
| $P(v_2 / v_1 = 400)$ | 0,2682 | 0,3841 | 0,3477 | 1 |

Izvor: Izračunato prema podacima u tablici 5.8.

Druga izračunata vjerojatnost u tablici 5.9. znači da vjerojatnost da ispitanik ima prosječnu ocjenu iz srednje škole 4 uz uvjet da mu je džeparac između 200 kn i 400 kn, iznosi 0,3841.

5.7. Kontinuirana slučajna varijabla

Kontinuirana ili diskretna slučajna varijabla X takva je varijabla koja može poprimiti neprebrojivo beskonačno mnogo vrijednosti. Zato se za kontinuiranu slučajnu varijablu ne računa vjerojatnost u određenoj točki, nego nad određenim intervalom.

Funkcija vjerojatnosti ili funkcija gustoće vjerojatnosti kontinuirane slučajne varijable X ima svojstva:

- $f(x) \geq 0, \forall x,$

- $\int_{-\infty}^{+\infty} f(x)dx = 1$, (površina ispod krivulje funkcije vjerojatnosti je 1),
- $P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} f(x)dx$, $[x_2 > x_1]$, (**vjerojatnost** da slučajna varijabla X poprimi vrijednost veću od x_1 i manju ili jednaku x_2 , **jednaka je** određenom integralu funkcije vjerojatnosti $f(x)$ na tom intervalu, odnosno **površini koju krivulja funkcije vjerojatnosti na tom intervalu zatvara s pozitivnim smjerom osi x**).

Funkcija distribucije slučajne varijable X je funkcija koja daje vjerojatnost da će slučajna varijabla X poprimiti vrijednost jednaku ili manju od nekog realnog broja x :

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(x)dx \quad (5.39)$$

Parametri distribucije:

- **Očekivanje slučajne varijable** (ako integral konvergira):

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f(x)dx = \mu. \quad (5.40)$$

- **Varijanca slučajne varijable:**

$$V(X) = \int_{-\infty}^{+\infty} (x - \mu)^2 \cdot f(x)dx. \quad (5.41)$$

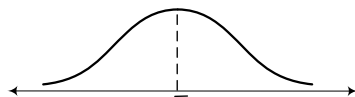
5.8. Teorijske distribucije kontinuirane slučajne varijable

5.8.1. Normalna distribucija

Normalna distribucija je simetrična ($\alpha_3 = 0$), unimodalna i ima oblik zvona. Normalno je zaobljena ($\alpha_4 = 3$):

Slika 5.6.

Normalna krivulja

*Izvor: Konstrukcija autora.*

Prema slici 5.6 može se uočiti da normalna krivulja u istoj točki točno na sredini ima aritmetičku sredinu, medijan i mod.

Funkcija vjerojatnosti ili funkcija gustoće vjerojatnosti kod normalne distribucije je:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in (-\infty, +\infty). \quad (5.42)$$

Općenito se kaže da se **kontinuirana slučajna varijabla X ravna po normalnoj distribuciji** koja je određena parametrima μ (očekivanje) i σ^2 (varijanca):

$$X \sim N(\mu, \sigma^2). \quad (5.43)$$

Za standardiziranu varijablu Z:

$$Z = \frac{x - \mu}{\sigma}, \quad (5.44)$$

funkcija vjerojatnosti je:

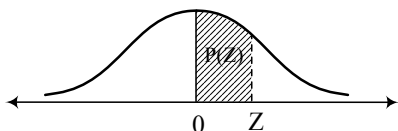
$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2}. \quad (5.45)$$

Ova normalna distribucija se naziva **standardizirana ili jedinična normalna distribucija** s očekivanjem 0 i jediničnom varijancom, pa vrijedi da je:

$$X \sim N(0,1). \quad (5.46)$$

Slika 5.7.

Površina ispod normalne krivulje

*Izvor: Konstrukcija autora.*

Jedinična normalna distribucija ima uvijek iste vrijednosti parametara, pa se vrijednosti za intervale vrijednosti varijable Z , od 0 do z kako je prikazano na slici 5.7, (u statističkoj literaturi ima i drugačije koncipiranih tablica), mogu prikazati u jedinstvenoj **tablici površina (vjerojatnosti) ispod normalne krivulje**.

Na primjer:

$$P(x_1 < X \leq x_2) = P_2\left(\frac{x_2 - \mu}{\sigma}\right) - P_1\left(\frac{x_1 - \mu}{\sigma}\right) = P_2(Z_2) - P_1(Z_1). \quad (5.47)$$

5.8.2. Studentova distribucija

Studentova distribucija još se naziva t-distribucija. Ima svoju primjenu u statistici kod procjene parametara osnovnog skupa i kod testiranja hipoteza na osnovi uzorka.

Varijabla "t" je definirana na području: $< -\infty, +\infty >$. Studentova distribucija je simetrična s obzirom na $t = 0$.

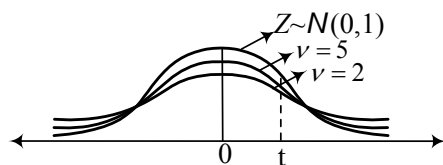
Ako je $Z \sim N(0,1)$ i χ^2 varijabla tzv. hi-kvadrat ili gama-distribucije s ν (ili *df* - degree of freedom) stupnjevima slobode i ako su one nezavisne slučajne varijable, tada je:

$$t_\nu = \frac{z}{\sqrt{\frac{\chi^2}{\nu}}}, \quad (5.48)$$

t - varijabla Studentove distribucije sa stupnjevima slobode ν (ili *df*).

Slika 5.8.

Studentova distribucija za različite stupnjeve slobode ν



Izvor: Konstrukcija autora.

Na slici 5.8 se vidi da se s povećanjem stupnjeva slobode ν , studentova distribucija približava normalnoj distribuciji. Već za $\nu = 30$ u praksi se umjesto t - distribucije upotrebljava normalna distribucija (pogreška aproksimacije u tom je slučaju manja od 0,03).

6. OCJENE PARAMETARA OSNOVNOG SKUPA NA OSNOVI UZORKA

6.1. Uzorak

Uzorak je podskup osnovnoga statističkog skupa, a uzima se u svrhu ispitivanja obilježja elemenata osnovnog skupa (ili populacije). Uzorak treba biti reprezentativan i slučajan. Slučajan izbor jedinica osnovnog skupa u uzorak može se postići pomoću tablice slučajnih brojeva. Jedinice u uzorak mogu se birati sistematski (npr. svaki 10. element osnovnog skupa).

Frakcija odabiranja (f) omjer je jedinica u uzorku i broja jedinica u osnovnom skupu:

$$f = \frac{n}{N}, \quad (6.1)$$

gdje je :

n - broj jedinica u uzorku,

N - broj jedinica u osnovnom skupu.

Korak izbora je recipročna vrijednost frakcije odabiranja ($1/f$) i upotrebljava se kod sistematskog izbora jedinica u uzorak.

Broj svih mogućih uzoraka (bez ponavljanja) veličine n iz osnovnog skupa veličine N jednak je broju kombinacija bez ponavljanja n -tog razreda:

$$K = \binom{N}{n}. \quad (6.2)$$

Pomoću uzorka vrši se procjena određenih parametara osnovnog skupa i testiraju se hipoteze o nepoznatim parametrima osnovnog skupa.

Ako se iz osnovnog skupa veličine N izaberu svi mogući uzorci veličine n , te se za svaki uzorak izračuna neki odgovarajući parametar, distribucija tih parametara naziva se **sampling distribucija**.

Priistranost je **razlika** između očekivane vrijednosti nekog parametra iz sampling distribucije i tog istog parametra iz osnovnog skupa:

$$E(\hat{\Theta}) \neq \Theta. \quad (6.3)$$

Ako između očekivane vrijednosti nekog parametra iz sampling distribucije i tog istog parametra iz osnovnog skupa **ne postoji razlika**, onda se to svojstvo naziva **nepriistranost**:

$$E(\hat{\Theta}) = \Theta. \quad (6.4)$$

Standardna devijacija sampling distribucije naziva se **standardna greška**.

6.2. Procjena aritmetičke sredine osnovnog skupa

Ako veličina uzorka teži prema beskonačno **sampling distribucija aritmetičkih sredina**, teži **normalnom obliku**. Kod malih uzoraka sampling distribucija aritmetičkih sredina ima oblik Studentove ili t-distribucije. Stoga vrijedi da ako je:

$n > 30 \Rightarrow$ koristi se Normalna distribucija i

$n \leq 30 \Rightarrow$ koristi se Studentova ili t-distribucija.

Standardna greška aritmetičke sredine je (standardna devijacija sampling distribucije aritmetičke sredine):

$$Se(\hat{\bar{X}}) = \frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad \text{ako je } f \leq 0,05 \text{ i}$$

$$Se(\hat{\bar{X}}) = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \quad \Rightarrow \quad \text{ako je } f > 0,05.$$

Ako nije poznata standardna devijacija osnovnog skupa σ , računa se **nepristrana** (točkasta) **ocjena varijance osnovnog skupa na osnovi uzorka** S^2 :

$$S^2 = \hat{\sigma} \cdot \left(\frac{n}{n-1} \right) \quad \Rightarrow \quad \text{ako je } n \leq 30 \text{ i}$$

$$S^2 = \hat{\sigma} \quad \Rightarrow \quad \text{ako je } n > 30 .$$

Točkasta procjena (procjena jednim brojem) **aritmetičke sredine** osnovnog skupa na osnovi uzorka je:

$$\hat{\bar{X}} .$$

Intervalna procjena aritmetičke sredine osnovnog skupa na osnovi uzorka je:

$$\Pr \left\{ \hat{\bar{X}} - Z \cdot Se(\hat{\bar{X}}) < \bar{X} < \hat{\bar{X}} + Z \cdot Se(\hat{\bar{X}}) \right\} = 1 - \alpha , \quad (6.5)$$

gdje je:

$\hat{\bar{X}}$ - aritmetička sredina izračunata na osnovi uzorka

Z - ako je ($n > 30$) računa se vrijednost iz tablice površina Normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$)

t - ako je ($n \leq 30$) umjesto (Z) računa se vrijednost iz tablice površina Studentove ili t-distribucije (pomoću $t_{\frac{\alpha}{2}, df=n-1}$)

$Se(\hat{\bar{X}})$ - standardna greška aritmetičke sredine

$1 - \alpha$ - nivo pouzdanosti procjene (ako nije određeno drukčije, najčešće se uzima da je $1 - \alpha = 95\%$).

Veličina uzorka uz zadani nivo pouzdanosti i maksimalnu grešku određuje se na sljedeći način:

$$n' = \left[\frac{Z \cdot \sigma}{greška} \right]^2 \Rightarrow \quad \text{ako je } f \leq 0,05 \text{ ili} \quad (6.6)$$

$$n = \frac{n'}{1 + \frac{n'}{N}} \quad \Rightarrow \quad \text{ako je } f > 0,05, \quad (6.7)$$

gdje je:

$Z_{\frac{1-\alpha}{2}}$ - vrijednost iz tablice površina Normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$)

σ - standardna devijacija iz osnovnog skupa

greška - maksimalna greška.

Ako je **maksimalna greška** određena u **relativnom izrazu** umjesto σ (standardne devijacije), veličina uzorka se računa pomoću V (koeficijenta varijacije):

$$V = \frac{\sigma}{\bar{X}} \cdot 100. \quad (6.8)$$



Analitički primjer 6.1.

Za zadani uzorak ispitanika potrebno je:

- Izračunati prosječnu visinu ispitanika u uzorku (varijabla *v1*)!
- Izračunati interval procjene za prosječnu visinu ispitanika osnovnog skupa uz pouzdanost procjene od 95%!
- Izračunati interval procjene za prosječnu visinu ispitanika osnovnog skupa uz pouzdanost procjene od 99%!



Rješenje 6.1:

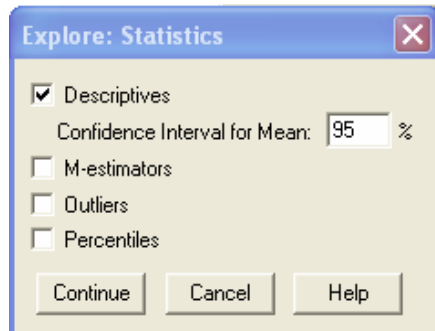
- Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Explore**.

U otvorenom prozoru **Explore** potrebno je zadanu numeričku varijablu **v1** - visina u cm pomoću odgovarajuće strelice prebaciti u **Dependent List**. U prostoru **Display** potrebno je aktivirati **Statistics**.

Klikom na ikonu **Statistics** otvara se novi prozor **Explore: Statistics** koji je prikazan na slici 6.1.

Slika 6.1.

Odabrani nivo pouzdanosti procjene za aritmetičku sredinu od 95%



Izvor: Simulirani podaci.

Prema zahtjevu zadatka od (b) ovdje je izabran nivo pouzdanosti procjene od 95%. Konačno, da bi se u **Outputu** programa SPSS dobili željeni podaci o visini ispitanika, potrebno je kliknuti na **Continue** i **OK**. Rezultat je prikazan u tablici 6.1.

Tablica 6.1.

Podaci o visini ispitanika

| Descriptives | | | Statistic | Std. Error |
|--------------|----------------------------------|-------------|-----------|------------|
| Visina u cm | Mean | | 175,77 | ,571 |
| | 95% Confidence Interval for Mean | Lower Bound | 174,64 | |
| | | Upper Bound | 176,89 | |
| | 5% Trimmed Mean | | 175,54 | |
| | Median | | 174,00 | ,159 |
| | Variance | | 76,539 | |
| | Std. Deviation | | 8,749 | |
| | Minimum | | 154 | |
| | Maximum | | 206 | |
| | Range | | 52 | |
| | Interquartile Range | | 12 | |
| | Skewness | | ,485 | |
| | Kurtosis | | ,184 | |
| | | | | |
| | | | | |

Izvor: Simulirani podaci.

Prema podacima iz tablice 6.1. može se vidjeti da je prosječna visina ispitanika u uzorku 175,77cm.

b) Intervalna procjena aritmetičke sredine osnovnog skupa pomoću zadanog uzorka uz nivo pouzdanosti procjene od 95% je:

$$\Pr\{174,64 < \bar{X} < 176,89\} = 95\% .$$

Uz traženu intervalnu procjenu u izlaznoj tablici **Descriptives** u **Outputu** programa **SPSS** dobiju se i sljedeći podaci uzorka ispitanika:

Median (medijan), Variance (varijanca), Std. Deviation (standardna devijacija), Minimum (najmanja vrijednost varijable), Maximum (najveća vrijednost varijable), Range (raspon varijacije obilježja), Interquartile Range (interkvartil), Skewness (mjera asimetrije), Kurtosis (mjera zaobljenosti), **Std. Error Mean (standardna greška aritmetičke sredine)**, Std. Error Skewness (standardna greška mjere asimetrije) i Std. Error Kurtosis (standardna greška mjere zaobljenosti).

c) Da bi se dobila intervalna procjena aritmetičke sredine visine osnovnog skupa pomoću zadanog uzorka uz nivo pouzdanosti procjene od 99%, potrebno je ponoviti sličan postupak, samo se u prozor **Explore: Statistics** upiše novi nivo pouzdanosti procjene 99%.

Tablica 6.2.

Podaci o visini ispitanika

| Descriptives | | | Statistic | Std. Error |
|--------------|----------------------------------|-------------|-----------|------------|
| Visina u cm | Mean | | 175,77 | ,571 |
| | 99% Confidence Interval for Mean | Lower Bound | 174,28 | |
| | | Upper Bound | 177,25 | |
| | 5% Trimmed Mean | | 175,54 | |
| | Median | | 174,00 | |
| | Variance | | 76,539 | |
| | Std. Deviation | | 8,749 | |
| | Minimum | | 154 | |
| | Maximum | | 206 | |
| | Range | | 52 | |
| | Interquartile Range | | 12 | |
| | Skewness | | ,485 | ,159 |
| | Kurtosis | | ,184 | ,316 |

Izvor: Simulirani podaci.

Rezultat je prikazan u tablici 6.2. Intervalna procjena aritmetičke sredine osnovnog skupa pomoću zadanog uzorka uz nivo pouzdanosti procjene od 99% je:

$$\Pr\{174,28 < \bar{X} < 177,25\} = 99\% .$$

Može se zaključiti da s povećanjem nivoa pouzdanosti procjene interval procjene postaje širi, odnosno procjena postaje manje precizna.



Analitički primjer 6.2.

Za zadani uzorak ispitanika potrebno je:

- Izračunati prosječnu težinu ispitanika u uzorku (varijabla *v1*)!
- Izračunati interval procjene za prosječnu težinu ispitanika osnovnog skupa uz pouzdanost procjene od 95%!
- Izračunati interval procjene za prosječnu težinu ispitanika osnovnog skupa uz pouzdanost procjene od 99%!



Rješenje 6.2:

- Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Explore**.

U otvorenom prozoru **Explore** potrebno je zadanu numeričku varijablu **v1** - težina u kg pomoću odgovarajuće strelice prebaciti u **Dependent List**. U prostoru **Display** potrebno je aktivirati **Statistics**.

Klikom na ikonu **Statistics** otvara se novi prozor **Explore: Statistics**.

Prema zahtjevu zadatka od (b) ovdje izabran nivo pouzdanosti procjene od 95%. Konačno, da bi se u **Outputu** programa SPSS dobili željeni podaci o težini ispitanika, potrebno je kliknuti na **Continue** i **OK**. Rezultat je prikazan u tablici 6.3.

Prema podacima iz tablice 6.3 može se vidjeti da je prosječna težina ispitanika u uzorku 61,18 kg.

Tablica 6.3.

Podaci o težini ispitanika

| Descriptives | | | Statistic | Std. Error |
|--------------|----------------------------------|-------------|-----------|------------|
| Tezina u kg | Mean | | 67,1795 | ,88431 |
| | 95% Confidence Interval for Mean | Lower Bound | 65,4372 | |
| | | Upper Bound | 68,9218 | |
| | 5% Trimmed Mean | | 66,3257 | |
| | Median | | 64,0000 | |
| | Variance | | 182,989 | |
| | Std. Deviation | | 13,52735 | |
| | Minimum | | 46,00 | |
| | Maximum | | 110,00 | |
| | Range | | 64,00 | |
| | Interquartile Range | | 19,00 | |
| | Skewness | | ,840 | ,159 |
| | Kurtosis | | ,214 | ,317 |

Izvor: Simulirani podaci.

b) Intervalna procjena aritmetičke sredine težine u osnovnom skupu pomoću zadanog uzorka uz nivo pouzdanosti procjene od 95% je:

$$\Pr\{65,44 < \bar{X} < 68,92\} = 95\%.$$

c) Da bi se dobila intervalna procjena aritmetičke sredine težine populacije osnovnog skupa pomoću zadanog uzorka uz nivo pouzdanosti procjene od 99%, potrebno je ponoviti sličan postupak, samo se u prozor **Explore: Statistics** upiše novi nivo pouzdanosti procjene 99%.

Rezultat je prikazan u tablici 6.4. Intervalna procjena aritmetičke sredine težine stanovnika u osnovnom skupu pomoću zadanog uzorka uz nivo pouzdanosti procjene od 99% je:

$$\Pr\{64,88 < \bar{X} < 69,48\} = 99\%.$$

Može se opet zaključiti da s povećanjem nivoa pouzdanosti procjene interval procjene postaje širi, odnosno procjena postaje manje precizna.

Tablica 6.4.

Podaci o težini ispitanika

| Descriptives | | | Statistic | Std. Error |
|--------------|----------------------------------|-------------|-----------|------------|
| Tezina u kg | Mean | | 67,1795 | ,88431 |
| | 99% Confidence Interval for Mean | Lower Bound | 64,8829 | |
| | | Upper Bound | 69,4761 | |
| | 5% Trimmed Mean | | 66,3257 | |
| | Median | | 64,0000 | |
| | Variance | | 182,989 | |
| | Std. Deviation | | 13,52735 | |
| | Minimum | | 46,00 | |
| | Maximum | | 110,00 | |
| | Range | | 64,00 | |
| | Interquartile Range | | 19,00 | |
| | Skewness | | ,840 | ,159 |
| | Kurtosis | | ,214 | ,317 |

*Izvor: Simulirani podaci.***6.3. Procjena proporcije osnovnog skupa**

Ako veličina uzorka teži prema beskonačno, **sampling distribucija proporcija** teži **normalnom obliku**. Kod malih uzoraka sampling distribucija proporcija ima oblik Studentove ili t-distribucije. Stoga vrijedi da ako je:

$n > 30 \Rightarrow$ koristi se Normalna distribucija i

$n \leq 30 \Rightarrow$ koristi se Studentova ili t-distribucija.

Standardna greška proporcije (standardna devijacija sampling distribucije proporcije) je:

- ako je $f \leq 0,05$:

$$Se(\hat{p}) = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \Rightarrow \text{ako je } n > 30 \text{ i}$$

$$Se(\hat{p}) = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n-1}} \quad \Rightarrow \quad \text{ako je } n \leq 30.$$

- ako je $f > 0,05$:

$$Se(\hat{p}) = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \cdot \sqrt{\frac{N-n}{N-1}} \quad \Rightarrow \quad \text{ako je } n > 30 \text{ i}$$

$$Se(\hat{p}) = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n-1}} \cdot \sqrt{\frac{N-n}{N-1}} \quad \Rightarrow \quad \text{ako je } n \leq 30.$$

Točkasta procjena (procjena jednim brojem) **proporcije** osnovnog skupa na osnovi uzorka je: \hat{p} , odnosno to je proporcija uzorka.

Intervalna procjena proporcije osnovnog skupa na osnovi uzorka je:

$$\Pr\{\hat{p} - Z \cdot Se(\hat{p}) < P < \hat{p} + Z \cdot Se(\hat{p})\} = 1 - \alpha \quad (6.9)$$

gdje je:

\hat{p} - proporcija izračunata na osnovi uzorka; ($\hat{q} = 1 - \hat{p}$)

Z - ako je ($n > 30$) računa se vrijednost iz tablice površina Normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$)

t - ako je ($n \leq 30$) umjesto (Z) računa se vrijednost iz tablice površina Studentove ili t-distribucije (pomoću $t_{\frac{\alpha}{2}, [df=n-1]}$)

$Se(\hat{p})$ - standardna greška proporcije

$1 - \alpha$ - nivo pouzdanosti procjene (ako nije određeno drugačije, najčešće se uzima da je $1 - \alpha = 95\%$).

Veličina uzorka uz zadani nivo pouzdanosti i maksimalnu grešku određuje se na sljedeći način:

$$n' = \left[\frac{Z \cdot \sqrt{P \cdot Q}}{\text{greška}} \right]^2 \quad \Rightarrow \quad \text{ako je } f \leq 0,05 \text{ ili} \quad (6.10)$$

$$n = \frac{n'}{1 + \frac{n'}{N}} \quad \Rightarrow \quad \text{ako je } f > 0,05, \quad (6.11)$$

gdje je:

Z - vrijednost iz tablice površina Normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$)

$P \cdot Q$ - varijanca osnovnog skupa

greška - maksimalna greška.

Ako je **maksimalna greška** određena u **relativnom izrazu** umjesto $\sqrt{P \cdot Q}$ (standardne devijacije), veličina uzorka se računa pomoću V (koeficijenta varijacije osnovnog skupa):

$$V = \sqrt{\frac{Q}{P}}, \quad (6.12)$$

pa vrijedi da je:

$$n' = \left[\frac{Z \cdot \sqrt{\frac{Q}{P}}}{\text{greška}} \right]^2. \quad (6.13)$$



Analitički primjer 6.3.

Za zadani uzorak ispitanika (koji su odgovorili na postavljeno pitanje) potrebno je:

- Izračunati proporciju uzorka ispitanika kojima su financijske prilike u obitelji takve da zadovoljavaju osnovne potrebe (varijabla v_1)!
- Izračunati interval procjene za proporciju ispitanika osnovnog skupa kojima su financijske prilike u obitelji takve da zadovoljavaju osnovne potrebe uz pouzdanost procjene od 95%!

- c) Izračunati interval procjene za proporciju ispitanika osnovnog skupa kojima su financijske prilike u obitelji takve da zadovoljavaju osnovne potrebe uz pouzdanost procjene od 99%!



Rješenje 6.3:

- a) Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

U otvorenom prozoru **Frequencies** potrebno je zadatu numeričku varijablu **v1** - financijske prilike u obitelji pomoću odgovarajuće strelice prebaciti u **Variable(s)**.

Da bi se u **Outputu** programa **SPSS** dobili željeni podaci o financijskim prilikama obitelji ispitanika, potrebno je kliknuti na **Continue** i **OK**. Rezultat je prikazan u tablici 6.5.

Tablica 6.5

Podaci o financijskim prilikama u obitelji ispitanika

Financijske prilike u obitelji

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------|---------------------------------|-----------|---------|---------------|--------------------|
| Valid | Nema novca za normalan standard | 4 | 1,7 | 1,7 | 1,7 |
| | Zadovoljava osnovne potrebe | 104 | 43,5 | 43,9 | 45,6 |
| | Nema materijalnih problema | 122 | 51,0 | 51,5 | 97,0 |
| | Zivi odlicno | 7 | 2,9 | 3,0 | 100,0 |
| | Total | 237 | 99,2 | 100,0 | |
| Missing | System | 2 | ,8 | | |
| Total | | 239 | 100,0 | | |

Izvor: Simulirani podaci.

Prema podacima u tablici 6.5. proporcija uzorka ispitanika kojima su financijske prilike u obitelji takve da zadovoljavaju osnovne potrebe omjer je povoljnih i svih mogućih ishoda, a može se i direktno očitati iz izlazne tablice u stupcu **Valid Percent**:

$$\hat{p} = \frac{m}{n} = \frac{104}{237} = 0,439; \quad \hat{q} = 1 - \hat{p} = 1 - 0,439 = 0,561.$$

S obzirom na to da je veličina uzorka 237, vrijedi da je $n > 30$, pa se za intervalnu procjenu koristi normalna distribucija;

b) Potrebno je napomenuti da program **SPSS** *nema mogućnost direktnog izračuna intervalne procjene proporcije osnovnog skupa na osnovi uzorka*.

Intervalna procjena proporcije osnovnog skupa na osnovi uzorka uz pouzdanost procjene od 95% je:

$$\Pr\{\hat{p} - Z \cdot Se(\hat{p}) < P < \hat{p} + Z \cdot Se(\hat{p})\} = 1 - \alpha,$$

$$n > 30 \quad \Rightarrow \quad Se(p) = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} = \sqrt{\frac{0,439 \cdot 0,561}{237}} = 0,032235879$$

Pomoću tablica odgovarajućih površina ispod normalne krivulje vrijedi da je:

$$Z_{\frac{1-\alpha}{2} = \frac{0,95}{2} = 0,475} \stackrel{tab}{\Rightarrow} Z = 1,96$$

Uvrštavanjem odgovarajućih veličina u početni izraz intervalne procjene dobije se:

$$\Pr\{0,439 - 1,96 \cdot 0,032235879 < P < 0,439 + 1,96 \cdot 0,032235879\} = 95\%,$$

Konačno je intervalna procjena proporcije ispitanika osnovnog skupa kojima su financijske prilike u obitelji takve da zadovoljavaju osnovne potrebe na osnovi zadanog uzorka uz pouzdanost procjene od 95%:

$$\Pr\{0,3758 < P < 0,50218\} = 95\%.$$

c) Na sličan način intervalna procjena proporcije ispitanika osnovnog skupa kojima su financijske prilike u obitelji takve da zadovoljavaju osnovne potrebe na osnovi uzorka uz pouzdanost procjene od 99% je:

$$Z_{\frac{1-\alpha}{2} = \frac{0,99}{2} = 0,495} \stackrel{tab}{\Rightarrow} Z = 2,58$$

$$\Pr\{0,439 - 2,58 \cdot 0,032235879 < P < 0,439 + 2,58 \cdot 0,032235879\} = 99\%,$$

$$\Pr\{0,35583 < P < 0,52217\} = 99\% .$$

Može se zaključiti da s povećanjem nivoa pouzdanosti procjene interval procjene postaje širi, odnosno procjena postaje manje precizna.



Analitički primjer 6.4.

Za zadani uzorak ispitanika (koji su odgovorili na postavljeno pitanje) potrebno je:

- Izračunati proporciju uzorka ispitanika kojima su roditelji razvedeni (varijabla v1)!
- Izračunati interval procjene za proporciju ispitanika osnovnog skupa kojima su roditelji razvedeni uz pouzdanost procjene od 95%!
- Izračunati interval procjene za proporciju ispitanika osnovnog skupa kojima su roditelji razvedeni uz pouzdanost procjene od 98%!



Rješenje 6.4:

- Na glavnom izborniku potrebno je izabrati ikonu *Analyze*, a na njezinu padajućem izborniku *Descriptive Statistics* i *Frequencies*.

U otvorenom prozoru *Frequencies* potrebno je zadanu numeričku varijablu *v1* - bračno stanje roditelja pomoću odgovarajuće strelice prebaciti u *Variable(s)*.

Tablica 6.6

Podaci o bračnom stanju roditelja ispitanika

Bracno stanje roditelja

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|-------------------------|-----------|---------|---------------|--------------------|
| Valid | U braku | 215 | 90,0 | 90,0 | 90,0 |
| | Rastavljeni | 10 | 4,2 | 4,2 | 94,1 |
| | Umro jedan od roditelja | 11 | 4,6 | 4,6 | 98,7 |
| | Ostalo | 3 | 1,3 | 1,3 | 100,0 |
| | Total | 239 | 100,0 | 100,0 | |

Izvor: Simulirani podaci.

Da bi se u **Outputu** programa **SPSS** dobili željeni podaci o bračnom stanju roditelja ispitanika, potrebno je kliknuti na **Continue** i **OK**. Rezultat je prikazan u tablici 6.6.

Prema podacima u tablici 6.6. proporcija uzorka ispitanika kojima su roditelji razvedeni omjer je povoljnih i svih mogućih ishoda, a može se i direktno očitati iz izlazne tablice u stupcu **Valid Percent**:

$$\hat{p} = \frac{m}{n} = \frac{10}{239} = 0,042 ; \hat{q} = 1 - \hat{p} = 1 - 0,042 = 0,958 .$$

S obzirom na to da je veličina uzorka 239, vrijedi da je $n > 30$, pa se za intervalnu procjenu koristi normalna distribucija.

b) Intervalna procjena proporcije u osnovnom skupu onih stanovnika kojima su roditelji razvedeni na osnovi uzorka uz pouzdanost procjene od 95% je:

$$\Pr\{\hat{p} - Z \cdot Se(\hat{p}) < P < \hat{p} + Z \cdot Se(\hat{p})\} = 1 - \alpha ,$$

$$n > 30 \quad \Rightarrow \quad Se(p) = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} = \sqrt{\frac{0,042 \cdot 0,958}{239}} = 0,012975032$$

Pomoću tablica odgovarajućih površina ispod normalne krivulje vrijedi da je:

$$Z_{\frac{1-\alpha}{2} = \frac{0,95}{2} = 0,475} \stackrel{tab}{\Rightarrow} Z = 1,96$$

Uvrštavanjem odgovarajućih veličina u početni izraz intervalne procjene dobije se:

$$\Pr\{0,042 - 1,96 \cdot 0,012975032 < P < 0,042 + 1,96 \cdot 0,012975032\} = 95\% .$$

Konačno je intervalna procjena proporcije ispitanika osnovnog skupa kojima su roditelji razvedeni na osnovi zadanog uzorka uz pouzdanost procjene od 95%:

$$\Pr\{0,0166 < P < 0,0674\} = 95\% .$$

c) Na sličan način intervalna procjena proporcije ispitanika osnovnog skupa kojima su roditelji razvedeni na osnovi uzorka uz pouzdanost procjene od 98% je:

$$Z_{\frac{1-\alpha}{2}=\frac{0,98}{2}=0,49} \stackrel{tab}{\Rightarrow} Z = 2,33 ,$$

$$\Pr\{0,042 - 2,33 \cdot 0,032235879 < P < 0,042 + 2,33 \cdot 0,032235879\} = 98\% ,$$

$$\Pr\{0,01177 < P < 0,07223\} = 98\% .$$

I ovdje vrijedi jednak zaključak da s povećanjem nivoa pouzdanosti procjene interval procjene postaje širi, odnosno procjena postaje manje precizna.

7. ANALITIČKI PRIMJERI ZA VJEŽBU



Analitički primjer 7.1.

Nakon provedene ankete potrebno je u dokument (Variable View) programa SPSS upisati varijablu (obilježje): Bračno stanje (v3) i kodirati je prema zadanoj tablici 7.1.

Tablica 7.1.

Kodna lista za varijablu Bračno stanje

| OBILJEŽJE | OBLICI OBILJEŽJA | KOD |
|--------------------|----------------------|-----|
| bracno stanje (v3) | neozenjen (neudana) | 1 |
| | ozenjen (udana) | 2 |
| | razveden (razvedena) | 3 |
| | udovac (udovica) | 4 |

Izvor: Podaci su simulirani.

Nakon kodiranja zadatak je unijeti konkretne podatke za 50 ispitanika zadane u tablici 7.2. u navedeni dokument (Data View).

Tablica 7.2.

Tablica s kodiranim podacima za varijablu Bračno stanje

| Bračno stanje | | | | |
|---------------|-----|-----|-----|-----|
| (1) | (2) | (3) | (4) | (5) |
| 2 | 1 | 1 | 1 | 2 |
| 2 | 1 | 4 | 1 | 2 |
| 1 | 2 | 1 | 1 | 2 |
| 1 | 2 | 1 | 1 | 1 |
| 1 | 1 | 2 | 2 | 2 |
| 1 | 1 | 2 | 2 | 2 |
| 1 | 3 | 1 | 2 | 1 |
| 2 | 1 | 1 | 2 | 1 |
| 2 | 2 | 3 | 3 | 1 |
| 2 | 1 | 1 | 1 | 2 |

Izvor: Podaci su simulirani.



Rješenje 7.1:

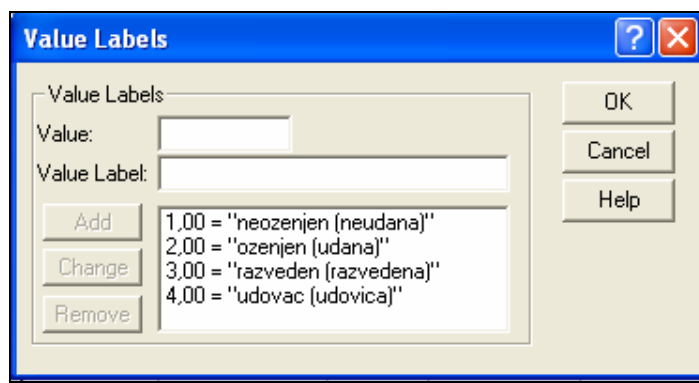
U programu SPSS pitanja, odnosno varijable se definiraju u dijelu **Variable view** koji se bira u donjem lijevom kutu ekrana.

Varijablu je potrebno nazvati (*Name*) v3. Radi lakšeg rada s podacima naznačena je vrsta podataka (*Type*) Numeric. U stupcu *Label* označeno je da je varijabla v3: Bračno stanje. Zadnji stupac definira skalu mjerenja (*Measure*), gdje je za v3 - Bračno stanje označena opisna, tj. *Nominal* skala.

U stupcu *Values* za varijablu v3 - Bračno stanje uneseni su kodovi prema tablici 7.1. Klikom miša u polje stupca *Values* otvara se prozor *Value Labels* u kojem se definira konkretno kodiranje. To je za ovaj primjer prikazano na slici 7.1.

Slika 7.1.

Prikaz definiranih varijabli u programu SPSS



Izvor: Podaci su simulirani.

U programu SPSS podaci prema definiranim varijablama se unose u dijelu **Data view** koji se bira u donjem lijevom kutu ekrana.

Na slici 7.2 prikazani su podaci o bračnom stanju dijela ispitanika prema zadanim podacima u tablici 7.2 (podaci su simulirani).

Naravno, nakon što su se podaci unijeli u program SPSS, potrebno ih je snimiti na standardan način kako se to radi kod većine klasičnih programa za računalo. Na glavnom izborniku odabire se **File**, a u njegovu

padajućem izborniku **Save As** te se uneseni podaci snimaju pod željenim imenom.

Slika 7.2.

Prikaz podataka varijable v3 - Bračno stanje za dio ispitanika u programu SPSS

| v3 | |
|------|--|
| 2,00 | |
| 2,00 | |
| 1,00 | |
| 1,00 | |
| 1,00 | |
| 1,00 | |
| 1,00 | |
| 1,00 | |
| 2,00 | |
| 2,00 | |
| 2,00 | |
| 1,00 | |
| 1,00 | |
| 2,00 | |
| 2,00 | |
| 1,00 | |
| 1,00 | |
| 3,00 | |
| 1,00 | |
| 2,00 | |
| 1,00 | |
| 1,00 | |
| 4,00 | |

Izvor: Podaci su simulirani.



Analitički primjer 7.2.

Zadatak je formirati statističku tablicu prema obilježju "Bračno stanje". Komentirajte drugu apsolutnu frekvenciju! Komentirajte prvi relativni broj strukture (Percent)!



Rješenje 7.2:

Potrebno je na glavnom izborniku odabrati *Analyze*, a na njegovu padajućem izborniku odabrati *Descriptive Statistics*. Dalje se bira *Frequencies*. Varijabla v3 - Bračno stanje se prebacuje u polje *Variable(s)*.

Klikom na *OK* u *Outputu* programa *SPSS* dobije se tablica frekvencija uzorka prema odabranom obilježju, odnosno varijabli. Rezultat je prikazan u tablicama 7.3.

Tablica 7.3.

| | | Bracno stanje | | | |
|-------|----------------------|---------------|---------|---------------|--------------------|
| | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | neozenjen (neudana) | 26 | 52,0 | 52,0 | 52,0 |
| | ozenjen (udana) | 20 | 40,0 | 40,0 | 92,0 |
| | razveden (razvedena) | 3 | 6,0 | 6,0 | 98,0 |
| | udovac (udovica) | 1 | 2,0 | 2,0 | 100,0 |
| | Total | 50 | 100,0 | 100,0 | |

Izvor: Podaci su simulirani.

Druga apsolutna frekvencija je 20 i ona u prvom slučaju znači da je od 50 odabranih ispitanika njih 20 bilo oženjenih (udanih). Prvi relativni broj strukture je 52 i on znači da je među odabranim ispitanicima njih 52% neoženjenih (neudanih).



Analitički primjer 7.3.

Zadatak je 50 anketiranih ispitanika prikazati jednostavnim stupcima prema obilježju "Bračno stanje" (upisati naslov grafikona)! Komentirajte sliku!

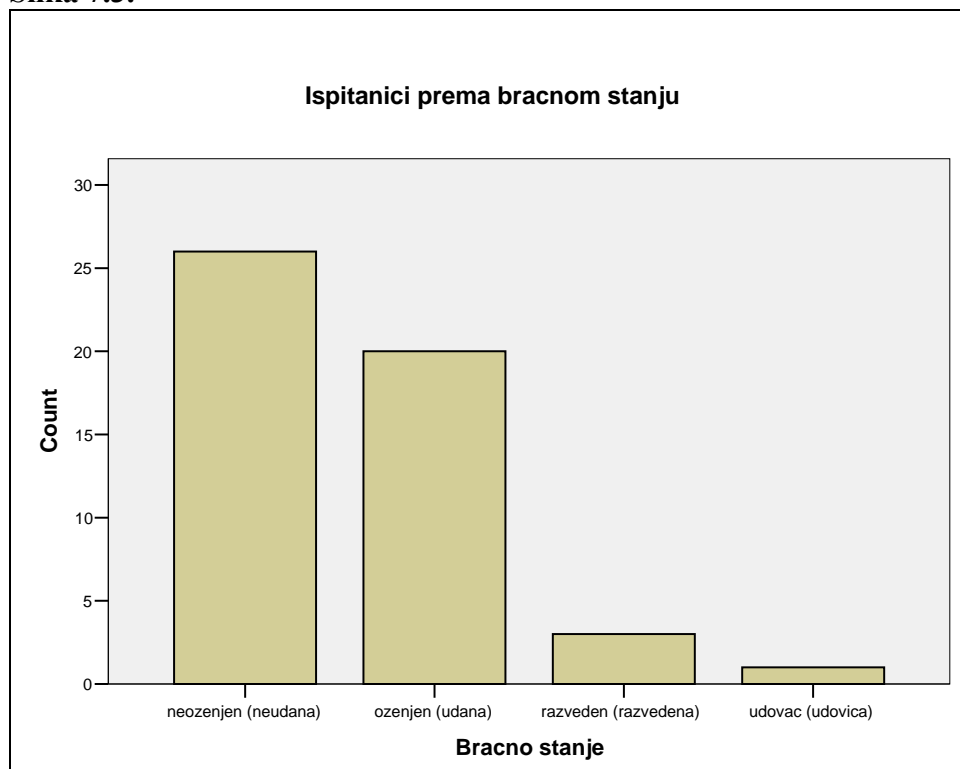


Rješenje 7.3:

Potrebno je na glavnom izborniku odabrati *Graphs*, a na njegovu padajućem izborniku *Bar chart* na kojem se aktivira ikona *Simple*.

Odabirom ikone **Define** otvara se prozor **Define Simple Bar: Summaries for Groups of Cases** gdje se u prostor **Category Axis** prebaci željena varijabla, što je u ovom slučaju Bračno stanje (v3). Grafikonu se može definirati i naslov na način da se klikne na ikonu **Titles...** Klikom na **OK** u **Outputu** programa **SPSS** dobije se traženi grafikon koji je prikazan na slici 7.3.

Slika 7.3.



Izvor: Podaci su simulirani.

Prema grafikonu jednostavnih stupaca može se vidjeti da je među odabranim ispitanicima najviše neoženjenih (neudanih), a najmanje onih koji su udovac ili udovica. Razvedenih ima manje u odnosu na broj oženjenih (udanih) u odabranom uzorku.

**Analitički primjer 7.4.**

Za ispitanike prema obilježju "Visina u cm" izračunajte: aritmetičku sredinu, medijan i mod! Komentirajte sve tri srednje vrijednosti na konkretnom primjeru!

**Rješenje 7.4:**

Na glavnom izborniku potrebno je izabrati ikonu *Analyze*, a na njezinu padajućem izborniku *Descriptive Statistics* i *Frequencies*.

U prozoru *Frequencies* pomoću odgovarajuće strelice varijabla v2 - visina u cm se prebacuje u polje *Variable(s)*.

Klikom na ikonu *Statistics* otvara se novi prozor *Frequencies Statistics* gdje je potrebno aktivirati određene statističke veličine prema zahtjevu zadatka, a koje se odnose na odabranu varijablu visina u cm (v2).

Aritmetička sredina dobit će se aktiviranjem veličine *Mean*, a za mod potrebno je označiti *Mode*, a za medijan *Median*.

Da bi se u *Outputu* programa SPSS dobili željeni podaci o visini ispitanika, potrebno je kliknuti na *Continue*, zatim na *OK*. Rezultat je prikazan u tablici 7.4.

Tablica 7.4.

Statistički podaci o visini ispitanika

| Statistics | | |
|-------------|---------|-----------|
| Visina u cm | | |
| N | Valid | 50 |
| | Missing | 0 |
| Mean | | 175,5800 |
| Median | | 174,5000 |
| Mode | | 172,00(a) |

a Multiple modes exist. The smallest value is shown

Izvor: Podaci su simulirani.

Prema podacima iz tablice 3.3. može se vidjeti da je u analizi 50 ispitanika. Prosječna visina ispitanika u odabranom uzorku (od onih koji su dali odgovor), odnosno aritmetička sredina je 175,58 cm.

Medijan je 174,5 cm, što znači da je polovica ispitanika s visinom manjom od 174,5 cm ili jednakom tom visinom, a preostala polovica ispitanika je s visinom većom od 174,5 cm.

Mod, odnosno najčešća visina ispitanika, 172 je cm. Ispod tablice je naznaka da u ovoj distribuciji to nije jedini mod, dakle ovo nije unimodalna distribucija.



Analitički primjer 7.5.

Zadatak je za samo ispitanike "muškog spola" formirati tablicu frekvencija prema varijabli (obilježju) "Školska sprema". Komentirajte 4. apsolutnu frekvenciju i opseg ovoga izabranog statističkog skupa!



Rješenje 7.5:

Nakon odabira (filtriranja) ispitanika muškog spola (**Data; Select cases; If condition is satisfied: v1 = 2**) koji je u obilježju spol kodiran s 2, potrebno je na glavnom izborniku odabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v2 - školska sprema se prebacuje u polje **Variable(s)**.

Da bi se u **Outputu** programa SPSS dobili željeni podaci o školskoj spremi odabranih ispitanika muškog spola, potrebno je kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 7.5.

Prema podacima iz tablice 7.5 može se vidjeti da su u analizi 22 ispitanika muškog spola. Četvrta apsolutna frekvencija je 13, što znači da je 13 ispitanika u analizi bilo sa srednjom stručnom spremom.

Tablica 7.5.

Statistički podaci o školskoj spremi ispitanika
Školska sprema

| | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------------------------|-----------|---------|---------------|--------------------|
| Valid Nisam pohađao skolu | 1 | 4,5 | 4,5 | 4,5 |
| Osnovna skola | 1 | 4,5 | 4,5 | 9,1 |
| Neka zanatska skola | 1 | 4,5 | 4,5 | 13,6 |
| Srednja skola | 13 | 59,1 | 59,1 | 72,7 |
| Visa skola | 1 | 4,5 | 4,5 | 77,3 |
| Visoka skola | 4 | 18,2 | 18,2 | 95,5 |
| Magisterij | 1 | 4,5 | 4,5 | 100,0 |
| Total | 22 | 100,0 | 100,0 | |

Izvor: Podaci su simulirani.



Analitički primjer 7.6.

Svih 50 anketiranih ispitanika prema varijabli (obilježju) "Školska sprema" prikazite grafički pomoću strukturnoga kruga (upisati naslov grafikona)! Komentirajte sliku!

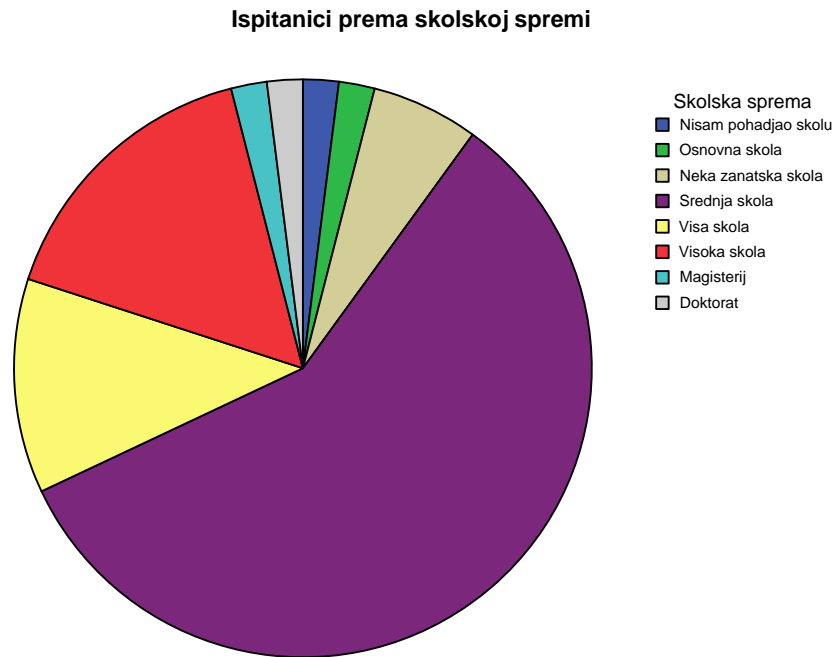


Rješenje 7.6:

Potrebno je na glavnom izborniku odabrati **Graphs**, a na njegovu padajućem izborniku **Pie chart**, na kojem se aktivira **Summaries of groups of case**.

Odabirom ikone **Define** otvara se prozor **Define Pie: Summaries for Groups of Cases** gdje se u prostor **Define Slices by** prebaci željena varijabla, što je u ovom slučaju Školska sprema (v2). Grafikonu se može definirati i naslov na način da se klikne na ikonu **Titles...** Klikom na **OK** u **Outputu** programa **SPSS** dobije se traženi grafikon koji je prikazan na slici 7.4.

Slika 7.4.



Izvor: Podaci su simulirani.

Prema grafikonu strukturnoga kruga može se vidjeti da je među odabranim ispitanicima najviše onih sa srednjom stručnom spremom, zatim slijede oni s završenom visokom i višom školom. Ostale kategorije stručne spreme u ovom su uzorku manje zastupljene.



Analitički primjer 7.7.

Ispitanike prema obilježju "težina u kg" potrebno je grupirati na sljedeći način:

| | | |
|----|------|------|
| 41 | - 50 | ⇒ 50 |
| 51 | - 60 | ⇒ 60 |
| 61 | - 70 | ⇒ 70 |
| 71 | - 80 | ⇒ 80 |
| 81 | - 90 | ⇒ 90 |

91 - 100 \Rightarrow 100
 101 - 110 \Rightarrow 110.

Nova varijabla je v4, a predstavlja "Grupiranu težinu".

Za ovu varijablu "Grupirana težina" formirajte tablicu frekvencija i komentirajte 3. relativni broj strukture (Percent)!



Rješenje 7.7:

Da bi se pomoću varijable v3 - "Težina u kg" formirala nova varijabla v4 - "Grupirana težina" po principu koji je zadan u ovom primjeru, potrebno je na glavnom izborniku odabrati **Transform**, a u njegovu padajućem izborniku odabrati **Recode**. U njegovu izborniku bira se **Into Different Variables** (jer bi se odabirom **Into Same Variables** izgubila postojeća varijabla). U polje **Numeric Variable** bira se v3 - "Težina u kg", a u **Output Variable, Name** je v4, a **Label** "Grupirana težina".

Tablica 7.6.

Statistički podaci o školskoj spremi ispitanika

Grupirana težina

| | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------------|-----------|---------|---------------|--------------------|
| Valid 50,00 | 4 | 8,0 | 8,0 | 8,0 |
| 60,00 | 14 | 28,0 | 28,0 | 36,0 |
| 70,00 | 15 | 30,0 | 30,0 | 66,0 |
| 80,00 | 7 | 14,0 | 14,0 | 80,0 |
| 90,00 | 5 | 10,0 | 10,0 | 90,0 |
| 100,00 | 4 | 8,0 | 8,0 | 98,0 |
| 110,00 | 1 | 2,0 | 2,0 | 100,0 |
| Total | 50 | 100,0 | 100,0 | |

Izvor: Podaci su simulirani.

Odabirom ikone **Old and New Values** otvara se novi prozor **Recode into Different Variables**. Aktiviranjem opcije **Range** upisuju se odgovarajuće grupe ocjena, kako je zadano. Nakon upisivanja svakog pojedinog **Range**, upisuje se odgovarajuća **New Value**, koja se svaki put kao nova grupa dodaje klikom na ikonu **Add**. Nakon što su unesene sve

grupe ikonom **Continue**, zatim **Change** i **OK**, u novom stupcu Data View kreirat će se nova varijabla "Grupirana težina", odnosno v4.

Treći relativni broj strukture iznosi 30 i znači da je 30% ispitanika u uzorku s težinom između 61 kg i 70 kg.



Analitički primjer 7.8.

Za sve ispitanike prema varijabli "težina u kg" izračunajte medijan, donji i gornji kvartil! Komentirajte sve 3 karakteristične vrijednosti niza!



Rješenje 7.8:

Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v3 - visina u cm se prebacuje u polje **Variable(s)**.

Klikom na ikonu **Statistics** otvara se novi prozor **Frequencies Statistics** gdje je potrebno aktivirati određene statističke veličine prema zahtjevu zadatka, a koje se odnose na odabranu varijablu težina u kg (v3).

Da bi se u **Outputu** programa SPSS dobili željeni podaci o težini ispitanika, potrebno je kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 7.7.

Tablica 7.7.

Statistički podaci o težini ispitanika

| Statistics | | |
|-------------|---------|---------|
| N | Valid | 50 |
| | Missing | 0 |
| Median | | 63,5000 |
| Percentiles | 25 | 55,7500 |
| | 50 | 63,5000 |
| | 75 | 77,2500 |

Izvor: Podaci su simulirani.

Prema podacima iz tablice 7.7. može se vidjeti da je u analizi bilo 50 ispitanika.

Medijan je 63,5 kg, što znači da je polovica ispitanika s težinom manjom od 63,5 kg ili jednakom težinom, a preostala polovica ispitanika je s težinom većom od 63,5 kg.

Donji kvartil je 55,75 kg, što znači da je 25% ispitanika s težinom manjom od 55,75 cm ili jednakom težinom, a preostalih 75% ispitanika je s težinom većom od 55,75 kg.

Gornji kvartil je 77,25 kg, što znači da je 75% ispitanika s težinom manjom od 77,25 kg ili jednakom težinom, a preostalih 25% ispitanika je s težinom većom od 77,25 kg.



Analitički primjer 7.9.

Nakon provedene ankete potrebno je u dokument (Variable View) programa SPSS upisati varijablu (obilježje): Materijalno stanje (v3) i kodirati je prema zadanoj tablici 7.8.

Tablica 7.8.

Kodna lista za varijablu Materijalno stanje

| OBILJEŽJE | OBLICI OBILJEŽJA | KOD |
|-------------------------|-----------------------|-----|
| materijalno stanje (P3) | Vrlo dobro | 1 |
| | Dobro | 2 |
| | Skromno, ali dovoljno | 3 |
| | Slabo | 4 |
| | Veoma slabo | 5 |

Izvor: Podaci su simulirani.

Nakon kodiranja zadatak je unijeti konkretne podatke za 50 ispitanika zadane u tablici 7.8. u navedeni dokument (Data View).

Tablica 7.9.

Tablica s kodiranim podacima za varijablu Materijalno stanje

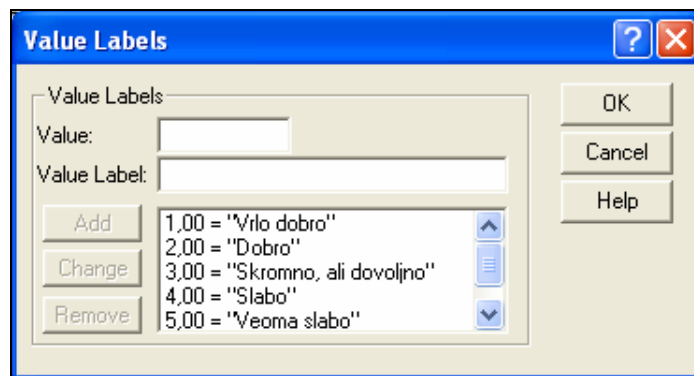
| Materijalno stanje | | | | |
|--------------------|-----|-----|-----|-----|
| (1) | (2) | (3) | (4) | (5) |
| 2 | 2 | 1 | 2 | 2 |
| 2 | 2 | 4 | 4 | 3 |
| 2 | 2 | 1 | 5 | 3 |
| 3 | 2 | 1 | 2 | 2 |
| 3 | 1 | 2 | 3 | 4 |
| 2 | 1 | 5 | 5 | 3 |
| 2 | 4 | 2 | 2 | 2 |
| 2 | 2 | 2 | 2 | 2 |
| 2 | 2 | 5 | 2 | 3 |
| 2 | 2 | 2 | 2 | 2 |

Izvor: Podaci su simulirani.**Rješenje 7.9:**

U programu SPSS pitanja, odnosno varijable se definiraju u dijelu **Variable view** koji se bira u donjem lijevom kutu ekrana.

Slika 7.5.

Prikaz definiranih varijabli u programu SPSS

*Izvor: Podaci su simulirani.*

Varijablu je potrebno nazvati (*Name*) v3. Radi lakšeg rada s podacima naznačena je vrsta podataka (*Type*) Numeric. U stupcu *Label*

označeno je da je varijabla v3: Materijalno stanje. Zadnji stupac definira skalu mjerenja (*Measure*) gdje je za v3 - Materijalno stanje označena opisna, tj. *Nominal* skala.

U stupcu *Values* za varijablu v3 - Materijalno stanje uneseni su kodovi prema tablici 7.8. Klikom miša u polje stupca *Values* otvara se prozor *Value Labels* u kojem se definira konkretno kodiranje. To je za ovaj primjer prikazano na slici 7.5.

U programu SPSS podaci prema definiranim varijablama unose se u dijelu **Data view** koji se bira u donjem lijevom kutu ekrana.

Na slici 7.6 prikazani su podaci o materijalnom stanju dijela ispitanika prema zadanim podacima u tablici 7.9 (podaci su simulirani).

Naravno, nakon što su se podaci unijeli u program SPSS, potrebno ih je snimiti na standardan način kako se to radi kod većine klasičnih programa za računalo. Na glavnom izborniku odabire se **File**, a u njegovu padajućem izborniku **Save As** te se uneseni podaci snimaju pod željenim imenom.

Slika 7.6.

Prikaz podataka varijable v3 - Materijalno stanje za dio ispitanika u programu SPSS

| v3 |
|------|
| 2,00 |
| 2,00 |
| 2,00 |
| 2,00 |
| 2,00 |
| 2,00 |
| 2,00 |
| 1,00 |
| 1,00 |
| 4,00 |
| 2,00 |
| 2,00 |
| 2,00 |
| 4,00 |

Izvor: Podaci su simulirani.



Analitički primjer 7.10.

Zadatak je formirati statističku tablicu prema obilježju "Materijalno stanje". Komentirajte drugu apsolutnu frekvenciju! Komentirajte četvrti i peti relativni broj strukture (Percent)!



Rješenje 7.10:

Potrebno je na glavnom izborniku odabrati *Analyze*, a na njegovu padajućem izborniku odabrati *Descriptive Statistics*. Dalje se bira *Frequencies*. Varijabla v3 - Materijalno stanje se prebacuje u polje *Variable(s)*.

Klikom na *OK* u *Outputu* programa *SPSS* dobije se tablica frekvencija uzorka prema odabranom obilježju, odnosno varijabli. Rezultat je prikazan u tablici 7.10.

Tablica 7.10.

| | | Materijalno stanje | | | |
|-------|-----------------------|--------------------|---------|---------------|--------------------|
| | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Vrlo dobro | 2 | 4,0 | 4,0 | 4,0 |
| | Dobro | 32 | 64,0 | 64,0 | 68,0 |
| | Skromno, ali dovoljno | 8 | 16,0 | 16,0 | 84,0 |
| | Slabo | 4 | 8,0 | 8,0 | 92,0 |
| | Veoma slabo | 4 | 8,0 | 8,0 | 100,0 |
| | Total | 50 | 100,0 | 100,0 | |

Izvor: Podaci su simulirani.

Druga apsolutna frekvencija je 32 i ona u ovom slučaju znači da je od 50 odabranih ispitanika njih 32 s dobrim materijalnim stanjem. Četvrti i peti relativni brojevi strukture su 8, što znači da je među odabranim ispitanicima njih po 8% sa slabim i veoma slabim materijalnim stanjem.



Analitički primjer 7.11.

Zadatak je 50 anketiranih ispitanika prikazati strukturnim krugom prema obilježju "Materijalno stanje" (upisati naslov grafikona)! Komentirajte sliku!

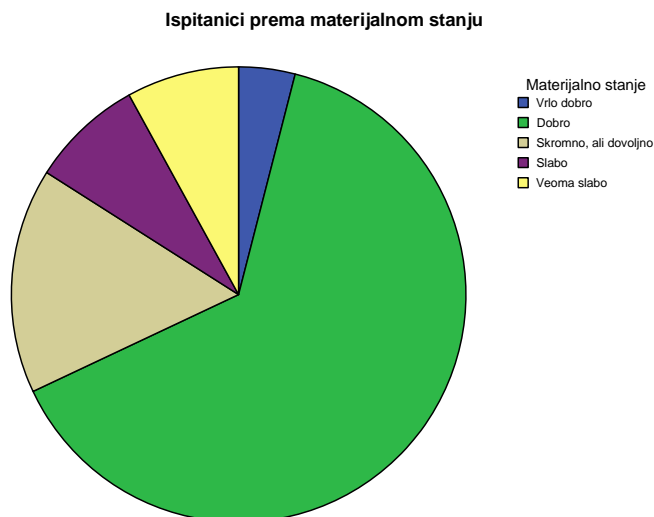


Rješenje 7.11:

Potrebno je na glavnom izborniku odabrati **Graphs**, a na njegovu padajućem izborniku **Pie chart** na kojem se aktivira **Summaries of groups of case**.

Odabirom ikone **Define** otvara se prozor **Define Pie: Summaries for Groups of Cases** gdje se u prostor **Define Slices by** prebaci željena varijabla, što je u ovom slučaju Materijalno stanje (v3). Grafikonu se može definirati i naslov na način da se klikne na ikonu **Titles...** Klikom na **OK** u **Outputu** programa **SPSS** dobije se traženi grafikon koji je prikazan na slici 7.7.

Slika 7.7.



Izvor: Podaci su simulirani.

Prema grafikonu strukturnog kruga može se vidjeti da je među odabranim ispitanicima najviše onih s dobrim materijalnim stanjem, zatim slijede oni koji su se izjasnili da im je materijalno stanje skromno, ali dovoljno. Ostale kategorije materijalnog stanja u ovom su uzorku manje zastupljene, a može se uočiti da onih čije je materijalno stanje vrlo dobro ima najmanje.



Analitički primjer 7.12.

Za ispitanike prema obilježju "Težina u kg" izračunajte: aritmetičku sredinu i medijan! Komentirajte te dvije srednje vrijednosti na konkretnom primjeru! Pokažite da vrijedi svojstvo aritmetičke sredine da je: $x_{\min} \leq \bar{X} \leq x_{\max}$.



Rješenje 7.12:

Na glavnom izborniku potrebno je izabrati ikonu *Analyze*, a na njezinu padajućem izborniku *Descriptive Statistics* i *Frequencies*.

U prozoru *Frequencies* pomoću odgovarajuće strelice varijabla v3 - težina u kg se prebacuje u polje *Variable(s)*.

Klikom na ikonu *Statistics* otvara se novi prozor *Frequencies Statistics* gdje je potrebno aktivirati određene statističke veličine prema zahtjevu zadatka, a koje se odnose na odabranu varijablu težina u kg (v3).

Tablica 7.11.

Statistički podaci o težini ispitanika
Statistics

| Težina u kg | | |
|-------------|---------|---------|
| N | Valid | 50 |
| | Missing | 0 |
| Mean | | 67,5400 |
| Median | | 63,5000 |
| Minimum | | 46,00 |
| Maximum | | 102,00 |

Izvor: Podaci su simulirani.

Da bi se u **Outputu** programa SPSS dobili željeni podaci o težini ispitanika, potrebno je kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 7.11.

Prema podacima iz tablice 7.11. može se vidjeti da je u analizi bilo 50 ispitanika.

Aritmetička sredina, odnosno prosječna težina odabranih ispitanika je 67,54 kg.

Medijan je 63,5 kg, što znači da je polovica ispitanika s težinom manjom od 63,5 ili jednakom tom težinom, a preostala polovica ispitanika je s težinom većom od 63,5 kg.

Da bi se dokazalo svojstvo aritmetičke sredine da se ona nalazi između najmanje i najveće vrijednosti niza, potrebno je pronaći najmanju i najveću vrijednost. Prema podacima u tablici 7.11 vrijedi da je:

$$x_{\min} \leq \bar{X} \leq x_{\max} \Rightarrow 46 \text{ kg} \leq \bar{X} \leq 102 \text{ kg} .$$


Analitički primjer 7.13.

Zadatak je za samo ispitanike "ženskog spola" formirati tablicu frekvencija prema varijabli (obilježju) "Abortus". Komentirajte 2. apsolutnu frekvenciju i 3. relativni broj strukture (Percent) ovog izabranoga statističkog skupa!



Rješenje 7.13:

Potrebno je na glavnom izborniku odabrati **Analyze**, a na njegovu padajućem izborniku odabrati **Descriptive Statistics**. Dalje se bira **Frequencies**. Varijabla v2 - Abortus se prebacuje u polje **Variable(s)**.

Klikom na **OK** u **Outputu** programa **SPSS** dobije se tablica frekvencija uzorka prema odabranom obilježju, odnosno varijabli. Rezultat je prikazan u tablici 7.12.

Druga apsolutna frekvencija je 17 i ona u ovom slučaju znači da od 50 odabranih ispitanika njih 17 smatra da je abortus dopušten u slučajevima silovanja. Treći relativni brojevi strukture je 4, što znači da je

među odabranim ispitanicima njih 4% koji su se izjasnili da su za to da abortus bude dopušten onda kada značajno remeti obiteljski život.

Tablica 7.12.

| | | Abortus | | | |
|-------|--|-----------|---------|---------------|--------------------|
| | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Ubojstvo | 26 | 52,0 | 52,0 | 52,0 |
| | Dopusten u slucajevima silovanja | 17 | 34,0 | 34,0 | 86,0 |
| | Dopusten kada znacajno remeti obiteljski zivot | 2 | 4,0 | 4,0 | 90,0 |
| | Iskljuciva odluka zene | 5 | 10,0 | 10,0 | 100,0 |
| | Total | 50 | 100,0 | 100,0 | |

Izvor: Podaci su simulirani.



Analitički primjer 7.14.

Svih 50 anketiranih ispitanika prema varijabli (obilježju) "Abortus" prikažite grafički pomoću jednostavnih stupaca (upisati naslov grafikona)! Komentirajte sliku!



Rješenje 7.14:

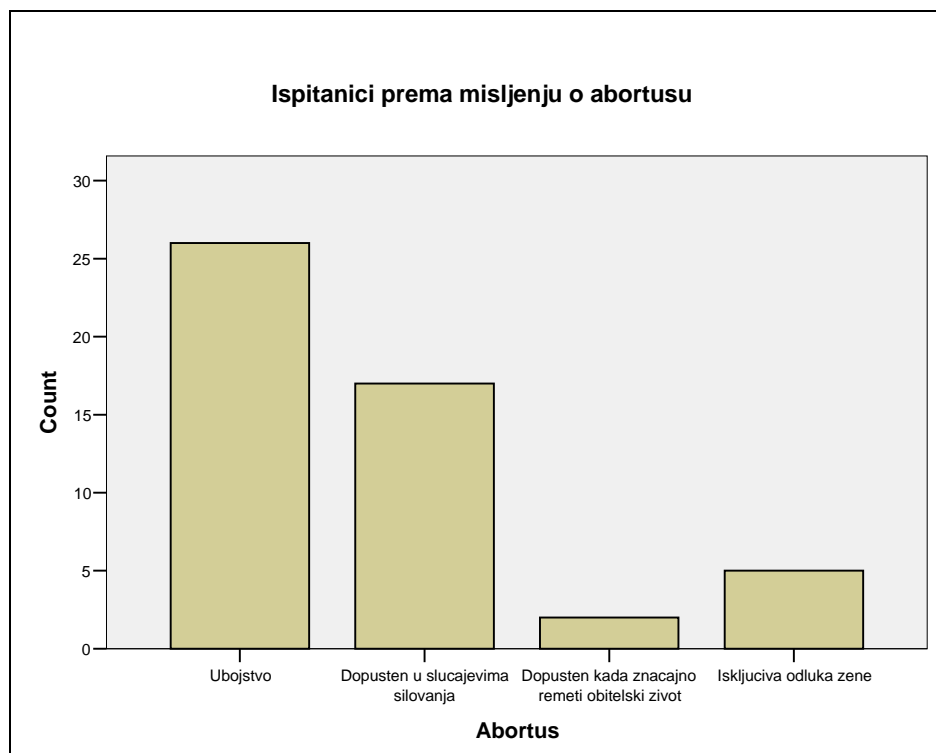
Potrebno je na glavnom izborniku odabrati **Graphs**, a na njegovu padajućem izborniku **Bar chart** na kojem se aktivira ikona **Simple**.

Odabirom ikone **Define** otvara se prozor **Define Simple Bar: Summaries for Groups of Cases** gdje se u prostor **Category Axis** prebaci željena varijabla, što je u ovom slučaju Abortus (v2). Grafikonu se može definirati i naslov na način da se klikne na ikonu **Titles...** Klikom na **OK** u **Outputu** programa **SPSS** dobije se traženi grafikon koji je prikazan na slici 7.8.

Prema grafikonu jednostavnih stupaca može se vidjeti da je među odabranim ispitanicima najviše onih koji smatraju da je abortus ubojstvo, a najmanje onih koji smatraju da je abortus dopušten kada značajno

remeti obiteljski život. Veliki dio ispitanika smatra da je abortus dopušten u slučajevima silovanja.

Slika 7.8.



Izvor: Podaci su simulirani.



Analitički primjer 7.15.

Za sve ispitanike prema varijabli "visina u cm" izračunajte aritmetičku sredinu! Komentirajte izračunatu srednju vrijednost! Pokažite da vrijedi svojstvo aritmetičke sredine da je: $\sum_{i=1}^N (x_i - \bar{X}) = 0$ (za negrupirani niz). (Neka to bude preko nove varijable v4, a traži se njezin zbroj: $Sum = 0$).

**Rješenje 7.15:**

Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v3 - visina u cm se prebacuje u polje **Variable(s)**.

Klikom na ikonu **Statistics** otvara se novi prozor **Frequencies Statistics** gdje je potrebno aktivirati određene statističke veličine prema zahtjevu zadatka, a koje se odnose na odabranu varijablu visina u cm (v3).

Da bi se u **Outputu** programa SPSS dobili željeni podaci o visini ispitanika, potrebno je kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 7.13.

Tablica 7.13.

Statistički podaci o visini ispitanika

| Statistics | | |
|-------------|---------|----------|
| Visina u cm | | |
| N | Valid | 50 |
| | Missing | 0 |
| Mean | | 175,0400 |

Izvor: Podaci su simulirani.

Prema podacima iz tablice 7.13 može se vidjeti da je u analizi bilo 50 ispitanika.

Aritmetička sredina odnosno prosječna visina odabranih ispitanika je 175,04 cm.

Da bi se dokazalo svojstvo aritmetičke sredine da je zbroj odstupanja vrijednosti obilježja od prosjeka jednak 0, potrebno je formirati novi stupac (varijablu) gdje će se izračunati ta odstupanja. Zbroj svih podataka iz toga novog stupca mora biti 0. Na glavnom izborniku potrebno je odabrati **Transform**, a na njegovu padajućem izborniku bira se **Compute**. Na taj se način otvori prozor **Compute Variable** gdje se računaju odstupanja od prosjeka kao nova varijabla v4.

Pod **Target Variable** upisuje se novo ime v4, a u **Numeric Expression** se upisuje: $v3 - 175,04$, odnosno da se od svih vrijednosti

varijable v3 oduzme njihov prosjek 175,04. Potrebno je napomenuti da se **decimalni zarez** u programu SPSS **označava točkom**. Klikom na **OK** u prostoru **Data View** formira se nova varijabla v4.

Da bi se vidio zbroj podataka u ovom stupcu na glavnom izborniku, potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Descriptives**. Varijabla v4 prebacuje se u **Variable(s)**. U **Options** je potrebno aktivirati **Sum** (dok se druge veličine mogu deaktivirati jer u ovom slučaju nisu potrebne).

Tablica 7.14.

Zbroj odstupanja vrijednosti obilježja od prosjeka

| Descriptive Statistics | | |
|------------------------|----|-----|
| | N | Sum |
| v4 | 50 | ,00 |
| Valid N (listwise) | 50 | |

Izvor: Podaci su simulirani.

Suma vrijednosti novoformirane varijable v4 u zadnjem stupcu tablice 7.14 je 0, čime je dokazano navedeno svojstvo aritmetičke sredine.



Analitički primjer 7.16.

Za varijablu (obilježje) *prosječna ocjena na I. godini studija (v1)* odredite: standardnu devijaciju i asimetriju distribucije (Skewness). Komentirajte izračunate vrijednosti na konkretnom primjeru!



Rješenje 7.16:

Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v1 - *prosječna ocjena na I. godini studija* prebacuje se u polje **Variable(s)**.

Klikom na ikonu **Statistics** otvara se novi prozor **Frequencies Statistics** gdje je potrebno aktivirati određene statističke veličine prema zahtjevu zadatka, a koje se odnose na odabranu varijablu (v1).

Da bi se u **Outputu** programa SPSS dobili željeni podaci o varijabli *prosječna ocjena na I. godini studija* ispitanika, potrebno je kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 7.15.

Tablica 7.15.

| Statistics | | |
|------------------------------|---------|--------|
| PROSJEČNA OCJENA NA I.GODINI | | |
| N | Valid | 40 |
| | Missing | 0 |
| Std. Deviation | | ,88252 |
| Skewness | | -,018 |
| Std. Error of Skewness | | ,374 |

Izvor: Podaci su simulirani.

Prema podacima iz tablice 7.15 može se vidjeti da je u analizi 40 ispitanika.

Standardna devijacija, odnosno prosječno odstupanje vrijednosti zadanog obilježja od aritmetičke sredine je 0,88.

Prema pokazatelju asimetrije (Skewness) može se zaključiti da je zadana distribucija lagano negativno ili ljevostrano asimetrična.



Analitički primjer 7.17.

Izračunati dvodimenzionalnu distribuciju vjerojatnosti: za

- varijablu (obilježje) *prosječna ocjena na I godini studija (v1)* koja se mijenja po redcima (row) i za
- varijablu (obilježje) *grupirani džeparac (v2)* koja se mijenja po stupcima (column).

a) Komentirajte marginalnu vjerojatnost da je džeparac 400 kn u izračunatoj dvodimenzionalnoj distribuciji!

b) Komentirajte vjerojatnost da ispitanik ima džeparac 400 kn i prosječnu ocjenu na I. godini studija 4 prema izračunatoj dvodimenzionalnoj distribuciji!



Rješenje 7.17:

a) Na glavnom izborniku potrebno je odabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Crosstabs**. Nakon toga otvara se prozor **Crosstabs**. Odgovarajućim strelicama prema zahtjevu zadatka potrebno je varijablu (obilježje) *prosječna ocjena na I godini studija (v1)* prebaciti u prozor za retke Row(s) i varijablu (obilježje) *grupirani džeparac (v2)* prebaciti u prozor za stupce Column(s).

Klikom na ikonu **Cells...** otvara se prozor **Crosstabs: Cell Display** gdje je uz postojeće originalne frekvencije (**Counts**) potrebno aktivirati **Percentages; Total**.

Tablica 7.16.

Dio dvostruke tablice obilježja prosječna ocjena na I. godini studija i grupirani džeparac

| PROSJEČNA OCJENA NA I GODINI * GRUPIRANI DŽEPARAC Crosstabulation | | | | | | | |
|---|------|------------|--------------------|--------|--------|--------|--------|
| | | | GRUPIRANI DŽEPARAC | | | | |
| | | | 200,00 | 300,00 | 400,00 | 600,00 | 800,00 |
| PROSJEČNA OCJENA NA I GODINI | 1,00 | Count | 1 | 0 | 0 | 0 | 0 |
| | | % of Total | 2,5% | ,0% | ,0% | ,0% | ,0% |
| | 2,00 | Count | 4 | 0 | 0 | 2 | 1 |
| | | % of Total | 10,0% | ,0% | ,0% | 5,0% | 2,5% |
| | 3,00 | Count | 10 | 0 | 3 | 3 | 1 |
| | | % of Total | 25,0% | ,0% | 7,5% | 7,5% | 2,5% |
| | 4,00 | Count | 4 | 1 | 5 | 1 | 0 |
| | | % of Total | 10,0% | 2,5% | 12,5% | 2,5% | ,0% |
| | 5,00 | Count | 0 | 1 | 0 | 1 | 0 |
| | | % of Total | ,0% | 2,5% | ,0% | 2,5% | ,0% |
| Total | | Count | 19 | 2 | 8 | 7 | 2 |
| | | % of Total | 47,5% | 5,0% | 20,0% | 17,5% | 5,0% |

Izvor: Podaci su simulirani.

Da bi se u **Outputu** programa **SPSS** dobili traženi podaci, potrebno je kliknuti na **Continue** i **OK**.

Rezultati su prikazani u tablici 7.16.

Izračunate vjerojatnosti marginalne distribucije nalaze se u zbirnom retku i zbirnom stupcu (koji nije prikazan u tablici) tablice 7.16.

Treća marginalna vjerojatnost u zbirnom retku iznosi 20%, tj. vjerojatnost da ispitanik ima džeparac 400 kn, bez obzira na prosječnu ocjenu na I. godini studija iznosi 0,2.

b) Vjerojatnost da ispitanik ima džeparac 400 kn i prosječnu ocjenu na I. godini studija 4 prema izračunatoj dvodimenzionalnoj distribuciji u tablici 7.16 je 12,5%, odnosno 0,125.

Analitički primjer 7.18.

Uz pretpostavku da promatrane slučajne varijable *prosječna ocjena na I. godini studija* (v_1) i *grupirani džeparac* (v_2) nisu nezavisne, izračunajte uvjetnu distribuciju vjerojatnosti: $P(v_1 / v_2 = 200)$!

Rješenje 7.18:

Uz pretpostavku da promatrane slučajne varijable nisu nezavisne, uvjetna distribucija vjerojatnosti $P(v_1 / v_2 = 200)$ pomoću podataka u tablici 7.16 izračunata je na sljedeći način i prikazana je u tablici 7.17:

$$P(v_1 = 1 / v_2 = 200) = \frac{P(v_1 = 1, v_2 = 200)}{P(v_2 = 200)} = \frac{0,025}{0,475} = 0,052632,$$

.....

$$P(v_1 = 5 / v_2 = 200) = \frac{P(v_1 = 5, v_2 = 200)}{P(v_2 = 200)} = \frac{0,000}{0,475} = 0,000.$$

Zbroj vjerojatnosti u ovoj marginalnoj distribuciji vjerojatnosti jednak je 1.

Tablica 7.17.

Uvjetna distribucija vjerojatnosti

| v1 | 1 | 2 | 3 | 4 | 5 | Σ |
|------------------|---------|---------|---------|---------|-------|----------|
| $P(v1/v2 = 200)$ | 0,05263 | 0,21053 | 0,52632 | 0,21053 | 0,000 | 1 |

Izvor: Izračunato prema podacima u tablici 7.16.

**Analitički primjer 7.19.**

Za zadani uzorak ispitanika izračunajte aritmetičku sredinu *prosječna ocjena na I godini studija (v1)*. Za izračunatu srednju vrijednost potrebno je izračunati interval procjene za prosjek osnovnog skupa uz pouzdanost procjene od 99%! Komentirajte rezultate!

**Rješenje 7.19:**

Na glavnom izborniku potrebno je izabrati ikonu *Analyze*, a na njezinu padajućem izborniku *Descriptive Statistics* i *Explore*.

U otvorenom prozoru *Explore* potrebno je zadanu numeričku varijablu *v1 - prosječna ocjena na I godini studija* pomoću odgovarajuće strelice prebaciti u *Dependent List*. U prostoru *Display* potrebno je aktivirati *Statistics*.

Klikom na ikonu *Statistics* otvara se novi prozor *Explore: Statistics*.

Prema zahtjevu zadatka ovdje je izabran nivo pouzdanosti procjene od 99%. Konačno, da bi se u *Outputu* programa SPSS dobili željeni podaci o *prosječna ocjena na I. godini studija* ispitanika, potrebno je kliknuti na *Continue* i *OK*. Rezultat je prikazan u tablici 7.18.

Prema podacima iz tablice 7.18 može se vidjeti da je prosječna ocjena na I. godini studija ispitanika u uzorku 3,125.

Tablica 7.18.

Podaci o je prosječnoj ocjeni na I. godini studija ispitanika

| Descriptives | | | Statistic | Std. Error |
|---------------------------------|-------------------------------------|-------------|-----------|------------|
| PROSJEČNA OCJENA NA I GODINI | Mean | | 3,1250 | ,13954 |
| | 99% Confidence Interval for Mean | Lower Bound | 2,7471 | |
| | | Upper Bound | 3,5029 | |
| | 5% Trimmed Mean | | 3,1111 | |
| | Median | | 3,0000 | |
| | Variance | | ,779 | |
| | Std. Deviation | | ,88252 | |
| | Minimum | | 1,00 | |
| | Maximum | | 5,00 | |
| | Range | | 4,00 | |
| | Interquartile Range | | 1,00 | |
| | Skewness | | -,018 | ,374 |
| | Kurtosis | | -,071 | ,733 |

Izvor: Podaci su simulirani.

Intervalna procjena aritmetičke sredine težine u osnovnom skupu pomoću zadanog uzorka uz nivo pouzdanosti procjene od 99% je:

$$\Pr\{2,75 < \bar{X} < 3,50\} = 99\% .$$

**Analitički primjer 7.20.**

Za varijablu (obilježje) *dob ispitanika (v1)* odredite: koeficijent kvartilne devijacije. Komentirajte izračunatu vrijednost na konkretnom primjeru!

**Rješenje 7.20:**

Na glavnom izborniku potrebno je izabrati ikonu *Analyze*, a na njezinu padajućem izborniku *Descriptive Statistics* i *Frequencies*.

U prozoru *Frequencies* pomoću odgovarajuće strelice varijabla v1 - *dob ispitanika* se prebacuje u polje *Variable(s)*.

Klikom na ikonu **Statistics** otvara se novi prozor **Frequencies Statistics** gdje je potrebno aktivirati određene statističke veličine prema zahtjevu zadatka, a koje se odnose na odabranu varijablu (v1).

Tablica 7.19.

| Statistics | | |
|----------------------------|---------|-------|
| Dob ispitanika (grupirano) | | |
| N | Valid | 50 |
| | Missing | 21 |
| Percentiles | 25 | 20,00 |
| | 50 | 30,00 |
| | 75 | 40,00 |

Izvor: Podaci su simulirani.

Da bi se u **Outputu** programa SPSS dobili željeni podaci o ispitanicima, potrebno je kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 7.19.

Prema podacima iz tablice 7.19 može se vidjeti da je u analizi 50 ispitanika.

Koeficijent kvartilne devijacije je:

$$V_q = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{40 - 20}{40 + 20} = 0,333.$$



Analitički primjer 7.21.

Vjerojatnost da se među studentskom populacijom konzumira alkohol samo u posebnoj prigodi je: $p = 0,038$.

- Kolika je vjerojatnost da između 70 slučajno odabranih studenata (X1) budu 2 studenta koji konzumiraju alkohol samo u posebnoj prigodi?
- Kolika je vjerojatnost da između 70 slučajno odabranih studenata (X1) najviše 15 studenata konzumira alkohol samo u posebnoj prigodi?

**Rješenje 7.21:**

Zadana vjerojatnost slučajnog događaja da se među studentskom populacijom konzumira alkohol je manja od 0,10, a uzorak je veći od 50, pa se u ovom primjeru pretpostavlja da se slučajna varijabla X_1 ravna po Poissonovu zakonu.

Na glavnom izborniku potrebno je odabrati ikonu **Transform**, a na njezinu padajućem izborniku **Compute**. Nakon toga otvara se prozor **Compute Variable**. Da bi se aktivirala naredba za izračun vjerojatnosti po Poissonovu zakonu, potrebno je u izborniku **Function group**: odabrati **PDF & Noncentral PDF**. Nakon toga se u izborniku **Function and Special Variables**: bira **Pdf.Poisson** i klikom dva puta na tu funkciju ona se pojavljuje u prozoru **Numeric Expression**: na sljedeći način: **PDF.POISSON(?,?)**.

U navedenoj funkciji prvi upitnik predstavlja **quant**, tj. x vrijednost koju poprima slučajna varijabla u ovom primjeru X_1 . Drugi upitnik funkcije predstavlja **mean**, odnosno očekivanje slučajne varijable koja se po Poissonovu zakonu računa na sljedeći način:

$$n = 70; \text{prob} = 0,038: \Rightarrow \mu = n \cdot p = 70 \cdot 0,038 = 2,66.$$

U ovom primjeru vrijedi da je:

$$\text{quant} = X_1; \mu = 2,66,$$

pa se može upisati:

$$\text{PDF.POISSON}(X_1, 2.66).$$

U **Target Variable**: potrebno je imenovati novu varijablu izračunatih vjerojatnosti po Poissonovu zakonu i u ovom primjeru nazvana je **X1Poiss**.

Konačno, da bi se u **dodatnom stupcu** programa SPSS dobili željeni podaci **vjerojatnosti slučajne varijable X_1 po Poissonovoj distribuciji**, potrebno je kliknuti na **OK**.

Da bi se dobile kumulativne vjerojatnosti, potrebno je ponoviti sličan postupak.

Na glavnom izborniku potrebno je odabrati ikonu **Transform**, a na njezinu padajućem izborniku **Compute**. Nakon toga otvara se prozor **Compute Variable**. Da bi se aktivirala naredba za izračun kumulativnih vjerojatnosti po Poissonovu zakonu, potrebno je u izborniku **Function group**: odabrati **CDF & Noncentral CDF**. Nakon toga se u izborniku **Function and Special Variables**: bira **Cdf.Poisson** i klikom dva puta na tu funkciju ona se pojavljuje u prozoru **Numeric Expression**: na sljedeći način: **CDF.POISSON(?,?)**.

U navedenoj funkciji prvi upitnik opet predstavlja **quant**, tj. **x** vrijednost koju poprima slučajna varijabla u ovom primjeru X2. Drugi upitnik funkcije predstavlja **mean**, odnosno očekivanje slučajne varijable:

$$quant = X1; \mu = 2,66,$$

pa se može upisati:

$$\mathbf{CDF.POISSON(X1,2.66)}.$$

U **Target Variable**: potrebno je imenovati novu varijablu izračunatih vjerojatnosti po Poissonovu zakonu i u ovom primjeru nazvana je **X1CumPoiss**.

Konačno, da bi se u **dodatnom stupcu** programa SPSS dobili željeni podaci **kumulativnih vjerojatnosti slučajne varijable X1 po Poissonovoj distribuciji**, potrebno je kliknuti na **OK**. Rezultat je prikazan u tablici 7.20.

Prema rezultatima iz tablice 7.20. (prikazano je samo 13 vrijednosti od ukupno 70 vrijednosti zadane diskontinuirane slučajne varijable X1) mogu se očitati i/ili izračunati tražene vjerojatnosti.

a) Vjerojatnost da između 70 slučajno odabranih studenata (X1) budu 2 studenta koji konzumiraju alkohol samo u posebnoj prigodi je:

$$p(X1 = 2) = 0,2474628.$$

b) Vjerojatnost da između 70 slučajno odabranih studenata (X1) najviše 15 studenata konzumira alkohol samo u posebnoj prigodi je:

$$p(X1 \leq 15) = p(X1 = 0) + p(X1 = 1) + \dots + p(X1 = 15) \approx 1.$$

Izračunata vjerojatnost je približno jednaka 1 jer je zbroj svih 71 vjerojatnosti 1, ali zbog manjeg broja decimalnih mjesta ovaj je rezultat izjednačen s 1.

Tablica 7.20.

Podaci u SPSS-u vrijednosti, vjerojatnosti i kumulativne vjerojatnosti diskontinuirane slučajne varijable po Poissonovu zakonu

| X1 | v1 | X1poiss | X1cumpoiss |
|-------|----|----------|------------|
| ,00 | 3 | ,0699482 | ,0699482 |
| 1,00 | 1 | ,1860623 | ,2560105 |
| 2,00 | 3 | ,2474628 | ,5034733 |
| 3,00 | 3 | ,2194170 | ,7228903 |
| 4,00 | 2 | ,1459123 | ,8688027 |
| 5,00 | 3 | ,0776254 | ,9464280 |
| 6,00 | 3 | ,0344139 | ,9808419 |
| 7,00 | 3 | ,0130773 | ,9939192 |
| 8,00 | 3 | ,0043482 | ,9982674 |
| 9,00 | 3 | ,0012851 | ,9995526 |
| 10,00 | 1 | ,0003418 | ,9998944 |
| 11,00 | 1 | ,0000827 | ,9999771 |
| 12,00 | 1 | ,0000183 | ,9999954 |
| 13,00 | 3 | ,0000037 | ,9999991 |
| 14,00 | 1 | ,0000007 | ,9999998 |
| 15,00 | 3 | ,0000001 | 1,0000000 |

Izvor: Podaci su simulirani.



Analitički primjer 7.22.

Za zadani uzorak ispitanika (onih koji su odgovorili na postavljeno pitanje) potrebno je izračunati proporciju uzorka ispitanika koji nikada ne konzumiraju alkohol (varijabla v1) i izračunati interval procjene za proporciju ispitanika osnovnog skupa koji nikada ne konzumiraju alkohol uz pouzdanost procjene od 95%!

**Rješenje 7.22:**

Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

U otvorenom prozoru **Frequencies** potrebno je zadatu numeričku varijablu **v1** - *Konsumirate li alkoholna pića* pomoću odgovarajuće strelice prebaciti u **Variable(s)**.

Da bi se u **Output** programa **SPSS** dobili željeni podaci o konzumiranju alkohola ispitanika, potrebno je kliknuti na **Continue** i **OK**. Rezultat je prikazan u tablici 7.21.

Tablica 7.21

Podaci o konzumiranju alkohola

Konsumirate li alkoholna pica

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------|----------------------|-----------|---------|---------------|--------------------|
| Valid | NIKADA | 15 | 21,1 | 30,0 | 30,0 |
| | SAMO UZ OBROKE | 3 | 4,2 | 6,0 | 36,0 |
| | U POSEBNIM PRILIKAMA | 32 | 45,1 | 64,0 | 100,0 |
| | Total | 50 | 70,4 | 100,0 | |
| Missing | System | 21 | 29,6 | | |
| Total | | 71 | 100,0 | | |

Izvor: Podaci su simulirani.

Prema podacima u tablici 7.21 proporcija uzorka ispitanika koji nikada ne konzumiraju alkohol omjer je povoljnih i svih mogućih ishoda, a može se i direktno očitati iz izlazne tablice u stupcu **Valid Percent**:

$$\hat{p} = \frac{m}{n} = \frac{15}{50} = 0,3; \quad \hat{q} = 1 - \hat{p} = 1 - 0,3 = 0,7.$$

S obzirom na to da je veličina uzorka 50, vrijedi da je $n > 30$, pa se za intervalnu procjenu koristi normalna distribucija.

Intervalna procjena proporcije u osnovnom skupu onih stanovnika koji nikada ne konzumiraju alkohol na osnovi uzorka uz pouzdanost procjene od 95% je:

$$\Pr\{\hat{p} - Z \cdot Se(\hat{p}) < P < \hat{p} + Z \cdot Se(\hat{p})\} = 1 - \alpha,$$

$$n > 30 \quad \Rightarrow \quad Se(p) = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} = \sqrt{\frac{0,3 \cdot 0,7}{50}} = 0,064807406$$

Pomoću tablica odgovarajućih površina ispod normalne krivulje vrijedi da je:

$$Z_{\frac{1-\alpha}{2} = \frac{0,95}{2} = 0,475} \stackrel{tab}{\Rightarrow} Z = 1,96$$

Uvrštavanjem odgovarajućih veličina u početni izraz intervalne procjene dobije se:

$$\Pr\{0,3 - 1,96 \cdot 0,064807406 < P < 0,3 + 1,96 \cdot 0,064807406\} = 95\% ,$$

Konačno je intervalna procjena proporcije ispitanika osnovnog skupa koji nikada ne konzumiraju alkohol na osnovi zadanog uzorka uz pouzdanost procjene od 95%:

$$\Pr\{0,17298 < P < 0,427\} = 95\% .$$



Analitički primjer 7.23.

Za varijablu (obilježje) *prosječna ocjena na I godini studija (p1)* odredite: interkvartil i zaobljenost distribucije (Kurtosis). Komentirajte izračunate vrijednosti na konkretnom primjeru!



Rješenje 7.23:

Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v1 - *prosječna ocjena na I godini studija* prebacuje se u polje **Variable(s)**.

Klikom na ikonu **Statistics** otvara se novi prozor **Frequencies Statistics** gdje je potrebno aktivirati određene statističke veličine prema zahtjevu zadatka, a koje se odnose na odabranu varijablu (v1).

Da bi se u **Outputu** programa SPSS dobili željeni podaci o *prosječnoj ocjeni na I. godini studija* ispitanika, potrebno je kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 7.22.

Tablica 7.22.

PROSJEČNA OCJENA NA I. GODINI

| | | |
|------------------------|---------|--------|
| N | Valid | 40 |
| | Missing | 0 |
| Kurtosis | | -,071 |
| Std. Error of Kurtosis | | ,733 |
| Percentiles | 25 | 3,0000 |
| | 50 | 3,0000 |
| | 75 | 4,0000 |

Izvor: Podaci su simulirani.

Prema podacima iz tablice 7.22 može se vidjeti da je u analizi 40 ispitanika.

Interkvartil zadane distribucije može se izračunati pomoću donjeg i gornjeg kvartila:

$$I_q = Q_3 - Q_1 = 4 - 3 = 1.$$

Raspršenost prosječne ocjene na I. godini za srednjih 50% ispitanika je 1 ocjena.

Zaobljenost zadane distribucije može se komentirati pomoću pokazatelja Kurtosis koji u ovom primjer iznosi -0,071, što upućuje na blago tupi vrh distribucije u odnosu na normalno zaobljenu distribuciju.

**Analitički primjer 7.24.**

Za varijablu (obilježje) *dob ispitanika (v1)* odredite: varijancu. Komentirajte izračunatu vrijednost na konkretnom primjeru!

**Rješenje 7.24:**

Na glavnom izborniku potrebno je izabrati ikonu **Analyze**, a na njezinu padajućem izborniku **Descriptive Statistics** i **Frequencies**.

U prozoru **Frequencies** pomoću odgovarajuće strelice varijabla v1 - *dob ispitanika* prebacuje se u polje **Variable(s)**.

Klikom na ikonu **Statistics** otvara se novi prozor **Frequencies Statistics** gdje je potrebno aktivirati određene statističke veličine prema zahtjevu zadatka, a koje se odnose na odabranu varijablu (v1).

Da bi se u **Outputu** programa SPSS dobili željeni podaci o *dobi* ispitanika, potrebno je kliknuti na **Continue**, zatim na **OK**. Rezultat je prikazan u tablici 7.23.

Tablica 7.23.

| Statistics | | |
|----------------------------|---------|---------|
| Dob ispitanika (grupirano) | | |
| N | Valid | 50 |
| | Missing | 0 |
| Variance | | 138,612 |

Izvor: Podaci su simulirani.

Prema podacima iz tablice 7.23 može se vidjeti da je u analizi 50 ispitanika.

Varijanca, odnosno prosječno kvadratno odstupanje dobi ispitanika od aritmetičke sredine je $138,612 \text{ god}^2$.



Analitički primjer 7.25.

Vjerojatnost da se među studentskom populacijom ne konzumira alkohol je: $p = 0,228$.

- Kolika je vjerojatnost da između 6 slučajno odabranih studenata (X_1) budu 2 studenta koji ne konzumiraju alkohol?
- Kolika je vjerojatnost da između 6 slučajno odabranih studenata (X_1) najviše 3 studenta ne konzumiraju alkohol?



Rješenje 7.25:

Zadana vjerojatnost slučajnog događaja da se među studentskom populacijom ne konzumira alkohol je veća od 0,10, a uzorak je manji od 50, pa se u ovom primjeru pretpostavlja da se slučajna varijabla X_1 ravna po Binomom zakonu.

Na glavnom izborniku potrebno je odabrati ikonu **Transform**, a na njezinu padajućem izborniku **Compute**. Nakon toga otvara se prozor **Compute Variable**. Da bi se aktivirala naredba za izračun vjerojatnosti po Binomnom zakonu, potrebno je u izborniku **Function group**: odabrati **PDF & Noncentral PDF**. Nakon toga se u izborniku **Function and Special Variables**: bira **Pdf.Binom** i klikom dva puta na tu funkciju ona se pojavljuje u prozoru **Numeric Expression**: na sljedeći način: **PDF.BINOM(?,?,?)**.

U navedenoj funkciji prvi upitnik predstavlja **quant**, tj. x vrijednost koju poprima slučajna varijabla u ovom primjeru $X1$. Drugi upitnik funkcije predstavlja veličinu uzorka n , a treći upitnik je vjerojatnost realizacije slučajnog događaja p :

$quant = X1 ; n = 6 ; prob = 0,228,$

pa se može upisati:

PDF.BINOM(X1,6,0.228).

U **Target Variable**: potrebno je imenovati novu varijablu izračunatih vjerojatnosti po Binomnom zakonu i u ovom primjeru nazvana je **X1Bin**.

Konačno, da bi se u **dodatnom stupcu** programa SPSS dobili željeni podaci **vjerojatnosti slučajne varijable $X1$ po Binomnoj distribuciji**, potrebno je kliknuti na **OK**.

Da bi se dobile kumulativne vjerojatnosti, potrebno je ponoviti sličan postupak.

Na glavnom izborniku potrebno je odabrati ikonu **Transform**, a na njezinu padajućem izborniku **Compute**. Nakon toga otvara se prozor **Compute Variable**. Da bi se aktivirala naredba za izračun kumulativnih vjerojatnosti po Binomnom zakonu, potrebno je u izborniku **Function group**: odabrati **CDF & Noncentral CDF**. Nakon toga se u izborniku **Function and Special Variables**: bira **Cdf.Binom** i klikom dva puta na tu funkciju ona se pojavljuje u prozoru **Numeric Expression**: na sljedeći način: **CDF.BINOM(?,?,?)**.

U navedenoj funkciji prvi upitnik opet predstavlja **quant**, tj. x vrijednost koju poprima slučajna varijabla u ovom primjeru $X1$. Drugi

upitnik funkcije predstavlja veličinu uzorka n , a treći upitnik je vjerojatnost realizacije slučajnog događaja p :

$quant = X1; n = 6; prob = 0,228,$

pa se može upisati:

CDF.BINOM(X1,6,0.228).

U **Target Variable**: potrebno je imenovati novu varijablu izračunatih vjerojatnosti po Binomnom zakonu i u ovom primjeru nazvana je ***X1CumBin***.

Konačno, da bi se u **dodatnom stupcu** programa SPSS dobili željeni podaci **kumulativnih vjerojatnosti slučajne varijable X1 po Binomnoj distribuciji**, potrebno je kliknuti na **OK**. Rezultat je prikazan u tablici 7.24.

Tablica 7.24.

Podaci u SPSS-u vrijednosti, vjerojatnosti i kumulativne vjerojatnosti diskontinuirane slučajne varijable po Binomnom zakonu

| X1 | binomp | binomcum |
|------|----------|-----------|
| ,00 | ,2116917 | ,2116917 |
| 1,00 | ,3751221 | ,5868137 |
| 2,00 | ,2769684 | ,8637821 |
| 3,00 | ,1090653 | ,9728474 |
| 4,00 | ,0241582 | ,9970056 |
| 5,00 | ,0028539 | ,9998595 |
| 6,00 | ,0001405 | 1,0000000 |

Izvor: Podaci su simulirani.

Prema rezultatima iz tablice 7.24 mogu se očitati i/ili izračunati tražene vjerojatnosti.

a) Vjerojatnost da između 6 slučajno odabranih studenata ($X1$) budu 2 studenta koji ne konzumiraju alkohol je:

$$p(X1 = 2) = 0,277.$$

b) Vjerojatnost da između 6 slučajno odabranih studenata ($X1$) najviše 3 studenta ne konzumiraju alkohol je:

$$p(X1 \leq 3) = p(X1 = 0) + p(X1 = 1) + \dots + p(X1 = 3) = 0,973.$$



Analitički primjer 7.26.

Za zadani uzorak ispitanika (onih koji su odgovorili na postavljeno pitanje) potrebno je izračunati proporciju uzorka ispitanika koji konzumiraju alkohol samo u posebnoj prigodi (varijabla *v1*) i izračunati interval procjene za proporciju ispitanika osnovnog skupa koji konzumiraju alkohol samo u posebnoj prigodi uz pouzdanost procjene od 99%!



Rješenje 7.26:

Na glavnom izborniku potrebno je izabrati ikonu *Analyze*, a na njezinu padajućem izborniku *Descriptive Statistics* i *Frequencies*.

U otvorenom prozoru *Frequencies* potrebno je zadanu numeričku varijablu *v1* - *Konzumirate li alkoholna pića* pomoću odgovarajuće strelice prebaciti u *Variable(s)*.

Da bi se u *Outputu* programa *SPSS* dobili željeni podaci o konzumiranju alkohola ispitanika, potrebno je kliknuti na *Continue* i *OK*. Rezultat je prikazan u tablici 7.25.

Tablica 7.25

Podaci o konzumiranju alkohola

Konzumirate li alkoholna pica

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|----------------------|-----------|---------|---------------|--------------------|
| Valid | NIKADA | 15 | 30,0 | 30,0 | 30,0 |
| | SAMO UZ OBROKE | 3 | 6,0 | 6,0 | 36,0 |
| | U POSEBNIM PRILIKAMA | 32 | 64,0 | 64,0 | 100,0 |
| | Total | 50 | 100,0 | 100,0 | |

Izvor: Podaci su simulirani.

Prema podacima u tablici 7.25 proporcija uzorka ispitanika koji konzumiraju alkohol samo u posebnoj prigodi omjer je povoljnih i svih

mogućih ishoda, a može se i direktno očitati iz izlazne tablice u stupcu **Valid Percent**:

$$\hat{p} = \frac{m}{n} = \frac{32}{50} = 0,64; \quad \hat{q} = 1 - \hat{p} = 1 - 0,64 = 0,36.$$

S obzirom na to da je veličina uzorka 50, vrijedi da je $n > 30$, pa se za intervalnu procjenu koristi normalna distribucija.

Intervalna procjena proporcije u osnovnom skupu onih stanovnika koji konzumiraju alkohol samo u posebnim prigodama na osnovi uzorka uz pouzdanost procjene od 99% je:

$$\Pr\{\hat{p} - Z \cdot Se(\hat{p}) < P < \hat{p} + Z \cdot Se(\hat{p})\} = 1 - \alpha,$$

$$n > 30 \quad \Rightarrow \quad Se(p) = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} = \sqrt{\frac{0,64 \cdot 0,36}{50}} = 0,0678822$$

Pomoću tablica odgovarajućih površina ispod normalne krivulje vrijedi da je:

$$Z_{\frac{1-\alpha}{2} = \frac{0,99}{2} = 0,495} \stackrel{tab}{\Rightarrow} Z = 2,58$$

Uvrštavanjem odgovarajućih veličina u početni izraz intervalne procjene dobije se:

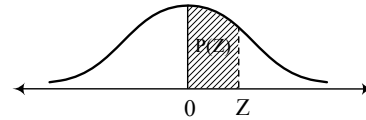
$$\Pr\{0,64 - 2,58 \cdot 0,0678822 < P < 0,64 + 2,58 \cdot 0,0678822\} = 99\%,$$

Konačno je intervalna procjena proporcije ispitanika osnovnog skupa koji konzumiraju alkohol samo u posebnoj prigodi na osnovi zadanog uzorka uz pouzdanost procjene od 99%:

$$\Pr\{0,46846 < P < 0,815136\} = 99\%.$$

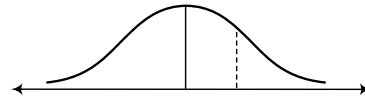
8. TABLICE ODABRANIH STATISTIČKIH DISTRIBUCIJA

A. Površine ispod normalne krivulje



| Z | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|-----|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 0.0 | 00000 | 00399 | 00798 | 01197 | 01595 | 01994 | 02392 | 02790 | 03188 | 03586 |
| 0.1 | 03983 | 04380 | 04776 | 05172 | 05567 | 05962 | 06356 | 06749 | 07142 | 07535 |
| 0.2 | 07926 | 08317 | 08706 | 09095 | 09483 | 09871 | 10257 | 10642 | 11026 | 11409 |
| 0.3 | 11791 | 12172 | 12552 | 12930 | 13307 | 13683 | 14058 | 14431 | 14803 | 15173 |
| 0.4 | 15542 | 15910 | 16276 | 16640 | 17003 | 17364 | 17724 | 18082 | 18439 | 18793 |
| 0.5 | 19146 | 19497 | 19847 | 20194 | 20540 | 20884 | 21226 | 21566 | 21904 | 22240 |
| 0.6 | 22575 | 22907 | 23237 | 23565 | 23891 | 24215 | 24537 | 24857 | 25175 | 25490 |
| 0.7 | 25804 | 26115 | 26424 | 26730 | 27035 | 27337 | 27637 | 27935 | 28230 | 28524 |
| 0.8 | 28818 | 29103 | 29389 | 29673 | 29955 | 30234 | 30511 | 30785 | 31057 | 31327 |
| 0.9 | 31594 | 31859 | 32121 | 32381 | 32639 | 32894 | 33147 | 33398 | 33646 | 33891 |
| 1.0 | 34134 | 34375 | 34614 | 34850 | 35083 | 35314 | 35543 | 35769 | 35993 | 36214 |
| 1.1 | 36433 | 36650 | 36864 | 37076 | 37286 | 37493 | 37698 | 37900 | 38100 | 38298 |
| 1.2 | 38493 | 38686 | 38877 | 39035 | 39251 | 39435 | 39617 | 39796 | 39973 | 40147 |
| 1.3 | 40320 | 40490 | 40658 | 40824 | 40988 | 41149 | 41309 | 41466 | 41621 | 41774 |
| 1.4 | 41924 | 42073 | 42220 | 42364 | 42507 | 42647 | 42786 | 42922 | 43056 | 43189 |
| 1.5 | 43319 | 43448 | 43574 | 43699 | 43822 | 43943 | 44062 | 44179 | 44295 | 44408 |
| 1.6 | 44520 | 44630 | 44738 | 44845 | 44950 | 45053 | 45154 | 45254 | 45352 | 45449 |
| 1.7 | 45543 | 45637 | 45728 | 45818 | 45907 | 45994 | 46080 | 46164 | 46246 | 46327 |
| 1.8 | 46407 | 46485 | 46562 | 46638 | 46712 | 46784 | 46856 | 46926 | 46995 | 47062 |
| 1.9 | 47128 | 47193 | 47257 | 47320 | 47381 | 47441 | 47500 | 47558 | 47615 | 47670 |
| 2.0 | 47725 | 47778 | 47831 | 47882 | 47932 | 47982 | 48030 | 48077 | 48124 | 48169 |
| 2.1 | 48214 | 48257 | 48300 | 48341 | 48382 | 48422 | 48461 | 48500 | 48537 | 48574 |
| 2.2 | 48610 | 48645 | 48679 | 48713 | 48745 | 48778 | 48809 | 48840 | 48870 | 48899 |
| 2.3 | 48928 | 48956 | 48983 | 49010 | 49036 | 49061 | 49086 | 49111 | 49134 | 49158 |
| 2.4 | 49180 | 49202 | 49224 | 49245 | 49266 | 49286 | 49305 | 49324 | 49343 | 49361 |
| 2.5 | 49379 | 49396 | 49413 | 49430 | 49446 | 49461 | 49477 | 49492 | 49506 | 49520 |
| 2.6 | 49534 | 49547 | 49560 | 49573 | 49585 | 49598 | 49609 | 49621 | 49632 | 49643 |
| 2.7 | 49653 | 49664 | 49674 | 49683 | 49693 | 49702 | 49711 | 49720 | 49728 | 49736 |
| 2.8 | 49744 | 49752 | 49760 | 49767 | 49774 | 49781 | 49788 | 49795 | 49801 | 49807 |
| 2.9 | 49813 | 49819 | 49825 | 49831 | 49836 | 49841 | 49846 | 49851 | 49856 | 49861 |
| 3.0 | 49865 | 49869 | 49874 | 49878 | 49882 | 49886 | 49889 | 49893 | 49897 | 49900 |
| 3.1 | 49903 | 49906 | 49910 | 49913 | 49916 | 49918 | 49921 | 49924 | 49926 | 49929 |
| 3.2 | 49931 | 49934 | 49936 | 49938 | 49940 | 49942 | 49944 | 49946 | 49948 | 49950 |
| 3.3 | 49952 | 49953 | 49955 | 49957 | 49958 | 49960 | 49961 | 49962 | 49964 | 49965 |
| 3.4 | 49966 | 49968 | 49969 | 49970 | 49971 | 49972 | 49973 | 49974 | 49975 | 49976 |
| 3.5 | 4997674 | | | | | | | | | |
| 4.0 | 4999683 | | | | | | | | | |
| 4.5 | 4999966 | | | | | | | | | |
| 5.0 | 499999713 | | | | | | | | | |

Napomena: Ispred svakog broja u polju tabele dolazi decimalna točka.

B. Kritične vrijednosti t, Studentove distribucije

| <i>P</i> | 40 | 30 | 25 | 20 | 15 | 10 | 5 | 2.5 | 1 | 0.5 | 0.1 | 0.05 |
|-------------|--------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|-------|
| <i>v</i> =1 | 0.3249 | 0.7265 | 1.0000 | 1.3764 | 1.963 | 3.078 | 6.314 | 12.71 | 31.82 | 63.66 | 318.1 | 636.6 |
| 2 | 0.2887 | 0.6172 | 0.8165 | 1.0607 | 1.386 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 22.33 | 31.60 |
| 3 | 0.2676 | 0.5844 | 0.7649 | 0.9785 | 1.250 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 10.21 | 12.92 |
| 4 | 0.2707 | 0.2686 | 0.7407 | 0.9410 | 1.190 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 7.173 | 8.610 |
| 5 | 0.2672 | 0.5594 | 0.7267 | 0.9195 | 1.156 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5.893 | 6.869 |
| 6 | 0.2648 | 0.5534 | 0.7176 | 0.9057 | 1.134 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 5.208 | 5.959 |
| 7 | 0.2632 | 0.5491 | 0.7111 | 0.8960 | 1.119 | 1.415 | 1.895 | 2.365 | 2.998 | 3.449 | 4.785 | 5.408 |
| 8 | 0.2619 | 0.5459 | 0.7064 | 0.8889 | 1.108 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 4.501 | 5.041 |
| 9 | 0.2610 | 0.5435 | 0.7027 | 0.8834 | 1.100 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 4.297 | 4.781 |
| 10 | 0.2602 | 0.5415 | 0.6998 | 0.8791 | 1.093 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.144 | 4.587 |
| 11 | 0.2596 | 0.5399 | 0.6974 | 0.8755 | 1.088 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 4.025 | 4.437 |
| 12 | 0.2590 | 0.5386 | 0.6955 | 0.8726 | 1.083 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 3.930 | 4.318 |
| 13 | 0.2586 | 0.5375 | 0.6938 | 0.8702 | 1.079 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 3.852 | 4.221 |
| 14 | 0.2582 | 0.5366 | 0.6924 | 0.8681 | 1.076 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 3.787 | 4.140 |
| 15 | 0.2579 | 0.5357 | 0.6912 | 0.8662 | 1.074 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 3.733 | 4.073 |
| 16 | 0.2576 | 0.5350 | 0.6901 | 0.8647 | 1.071 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 3.686 | 4.015 |
| 17 | 0.2573 | 0.5344 | 0.6982 | 0.8633 | 1.069 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.646 | 3.965 |
| 18 | 0.2571 | 0.5338 | 0.6892 | 0.8620 | 1.067 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.610 | 3.922 |
| 19 | 0.2569 | 0.5333 | 0.6876 | 0.8610 | 1.066 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.579 | 3.883 |
| 20 | 0.2567 | 0.5329 | 0.6870 | 0.8600 | 1.064 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.552 | 3.850 |
| 21 | 0.2566 | 0.5325 | 0.6864 | 0.8591 | 1.063 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | 3.527 | 3.819 |
| 22 | 0.2564 | 0.5321 | 0.6858 | 0.8583 | 1.061 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 3.505 | 3.792 |
| 23 | 0.2563 | 0.5317 | 0.6853 | 0.8575 | 1.060 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | 3.485 | 3.768 |
| 24 | 0.2562 | 0.5314 | 0.6848 | 0.8569 | 1.059 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 3.467 | 3.745 |
| 25 | 0.2561 | 0.5312 | 0.6844 | 0.8562 | 1.058 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 3.450 | 3.725 |
| 26 | 0.2560 | 0.5309 | 0.6840 | 0.8557 | 1.058 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 3.435 | 3.707 |
| 27 | 0.2559 | 0.5306 | 0.6837 | 0.8551 | 1.057 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 3.421 | 3.690 |
| 28 | 0.2558 | 0.5304 | 0.6834 | 0.8546 | 1.056 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 3.408 | 3.674 |
| 29 | 0.2557 | 0.5302 | 0.6830 | 0.8542 | 1.055 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 3.396 | 3.659 |
| 30 | 0.2556 | 0.5300 | 0.6828 | 0.8538 | 1.055 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 | 3.385 | 3.646 |
| 32 | 0.2555 | 0.5297 | 0.6822 | 0.8530 | 1.054 | 1.309 | 1.694 | 2.037 | 2.449 | 2.738 | 3.365 | 3.622 |
| 34 | 0.2553 | 0.5294 | 0.6818 | 0.8523 | 1.052 | 1.307 | 1.691 | 2.032 | 2.441 | 2.728 | 3.348 | 3.601 |
| 36 | 0.2552 | 0.5291 | 0.6814 | 0.8517 | 1.052 | 1.306 | 1.688 | 2.028 | 2.434 | 2.719 | 3.333 | 3.582 |
| 38 | 0.2551 | 0.5288 | 0.6810 | 0.8512 | 1.051 | 1.304 | 1.686 | 2.024 | 2.429 | 2.712 | 3.319 | 3.566 |
| 40 | 0.2550 | 0.5286 | 0.6807 | 0.8507 | 1.050 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 | 3.307 | 3.551 |
| 50 | 0.2547 | 0.5278 | 0.6794 | 0.8489 | 1.047 | 1.299 | 1.676 | 2.009 | 2.403 | 2.678 | 3.261 | 3.496 |
| 60 | 0.2545 | 0.5272 | 0.6786 | 0.8477 | 1.045 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 | 3.232 | 3.460 |
| 120 | 0.2539 | 0.5258 | 0.6765 | 0.8446 | 1.041 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 | 3.160 | 3.373 |
| ∞ | 0.2533 | 0.5244 | 0.6745 | 0.8416 | 1.036 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 | 3.291 |

9. LITERATURA

1. Anderson, D. R. et al. (2002). Statistics for Business and Economics. 8th. Edt. St Paul: West.
2. Anderson, D. R. et al. (1998). Quantitative Methods for Business. 7th. Edt. South-Western College Publishing.
3. Babbie, E. Halley, F. Zaino, J. (2000). *Adventures in Social Research. Data Analysis Using SPSS^X for Windows 95/98*. USA: Pine Forge Press.
4. Bešter, M. i Bregar, L. (1986). Ekonomska statistika. Ljubljana: Ekonomska fakulteta.
5. Blejec, M. (1976). Statističke metode za ekonomiste. Ljubljana: Ekonomska fakulteta.
6. Canavos, G. C. and Miller, D. M. (1993). An Introduction to Modern Business Statistics. Belmont. Wadsworth.
7. Cochran, W. G. (1977). Sampling Techniques. New York: Wiley.
8. Dillon, W. R. and Goldstein, M. (1984). Multivariate Analysis, Methods and Applications. New York: Wiley.
9. Dodge, M., Kinata, C. i Stison, C. (1997). Kako koristiti Microsoft Excel 97, Zagreb: Znak.
10. Everit, B. S. (1996). Making Sense of Statistics in Psychology, Oxford: Oxford University Press.
11. Feller, W. (1971). An Introduction to Probability Theory and its Applications. Vol. I and Vol. II, New York: Wiley.

-
12. Fisher, R. A. (1950). Statistical Methods for Research Workers. 11 th Edt. Edinburgh: Oliver and Boyd.
 13. Fisher, R. A. (1993). Statistical Methods, Eksperimental Design and Scientific Inference. Oxford: Oxford University Press.
 14. Frye, C. (2003). Microsoft Excel 2002 Korak po korak. Zagreb: Algoritam.
 15. Fulton, J. (1997). Vodič kroz Excel 97. Zagreb: Znak.
 16. Halmi, A. (2003). Multivarijatna analiza u društvenim znanostima. Alinea. Zagreb
 17. Hadživuković, S. (1975). Tehnika metoda uzorka. Beograd: Naučna knjiga
 18. Heij, C. et. al.(2004). *Econometrisc Methods with Applications in Business and Economics*, New York : Oxford University Press.
 19. Kendall, M. and Stuart, A. (1973). The Advanced Theory of Statistics. Vol I. London: Griffin.
 20. Kendall, M. and Stuart, A. (1974). The Advanced Theory of Statistics. Vol II. London: Griffin.
 21. Kendall, M. and Stuart, A. (1976). The Advanced Theory of Statistics. Vol III. London: Griffin.
 22. Kish, L. (1965). Survey Sampling. New Yor: Wiley.
 23. Kolesarić, V. i Petz. B. (2003). Statistički rječnik. Tumač statističkih pojmova. Naklada Slap.
 24. Leburić, A. i Krneta, M. (2004). Profil poduzetnika u Hrvatskoj. Socio-ekonomsko istraživanje 2003. Split: Naklada Bošković.

-
25. de Levie, R. (2004). Advanced Excel for Scientific Data Analysis. Oxford: Oxford University Press.
26. Martić, Lj. (1986). Mjere nejednakosti i siromaštva. Zagreb. Birotehnika.
27. Mason, D. R. and Lind, D. A. (1993). Statistical Techniques in Business and Economics. Boston, Massachusetts: IRVIN Publishing.
28. Metre, J. G. and Gilbreath, G. H. (1983). Statistics for Business and Economics. Plano. Texas: Business Publications.
29. Milton, J. S. et al. (1986). Introduction to Statistics. Lexington: D. C. Heath.
30. Momirović, K. (1998). Uvod u analizu nominalnih varijabli. Metodološke sveske 2. Ljubljana: JUS.
31. Mood, A. M. and Graybill, F. P. (1963). Introduction to the Theory of Statistics. New York: McGraw-Hill.
32. Neter, J. et al. (1993). Applied Statistics. 4th Edt. Boston: Allyn and Bacon.
33. Newbold, P. (1991). Statistics for Business and Economics. Englewood Cliffs: Prentice-Hall.
34. Pauše, Ž. (2003). Vjerojatnost, informacija, stohastički procesi. Zagreb: Školska knjiga.
35. Pavlić, I. (1971). Statistička teorija i primjena. Zagreb: Tehnička knjiga.
36. Pivac, S. and Jurun, E. (2005). Parameter Estimation in Excel, Proceedings of the 28th International Convention MIPRO 2005. Computers in Education, Opatija, pp 168-173.

-
37. Pivac, S. Jurun, E. and Jujnović, I. (2006). Ocjena parametara nelinearnih funkcija u programskom paketu Statistica na primjeru nezaposlenosti, Proceedings of the 29th International Convention MIPRO 2006. Computers in Education, Opatija, pp 241-245.
38. Pivac, S. Jurun, E. and Jerak, N. (2006). Rješavanje simultanih jednadžbi kao ekonometrijskog modela pomoću programskog paketa Excel, Proceedings of the 29th International Convention MIPRO 2006. Computers in Education, Opatija, pp 246-251.
39. Pivac, S. i Šego, B. (2005). Statistika, udžbenik sa zbirkom zadataka za IV. razred srednje Ekonomske škole, Zagreb: Alkascript.
40. Rozga, A. (2003). Statistika za ekonomiste, Sveučilište u Splitu. Ekonomski fakultet Split.
41. Rozga, A. i Grčić, B. (2003). Poslovna statistika, Ekonomski fakultet Split.
42. Seplaki, L. (1991). Attorneys' Dictionary and Handbook of Economics and Statistics, New York: Professional Horizons Press.
43. Siegel, A. F. (1994). Practical Business Statistics. Boston, Massachusetts: IRVIN Publishing.
44. Šošić, I. (1983). Metode statističke analize. Zagreb: Sveučilišna naklada Liber.
45. Šošić, I. (1985). Zbirka zadataka iz osnova statistike. Zagreb: Sveučilišna naklada Liber.
46. Šošić, I. i Serdar, V. (2002). Uvod u statistiku. XII. izdanje, Zagreb: Školska knjiga.
47. Šošić, I. (2004). Primijenjena statistika. Zagreb: Školska knjiga.

-
48. Tenjović, L. (2002). Statistika u psihologiji. Beograd: Centar za primenjenu psihologiju.
49. Žarković, S. S. (1965). Sampling Methods and Censuses. Rim:FAO
50. Whigham, D. (1998). Quantitative Business Methods Using Excel. Oxford University Press.
51. Wonnacott, T. H. and Wonnacott, R. J. (1990). Introductory Statistics for Business and Economics (fourth edition). New York: Wiley.
52. The Central Bureau of Statistics. Republic of Croatia: URL: <http://www.dzs.hr>

BILJEŠKA O AUTORICI I AUTORU

SNJEŽANA PIVAC, rođena 1972. u Splitu, gdje je završila osnovnu i srednju školu. Nakon studija na Ekonomskom fakultetu u Splitu, gdje je kao izvrstan student *diplomirala u roku od tri godine*, magistrira i doktorira pri Ekonomskom fakultetu u Zagrebu u znanstvenom polju ekonomije, grani kvantitativna ekonomija. Trenutno radi kao docentica na katedri kvantitativnih metoda u ekonomiji, na Ekonomskom fakultetu Sveučilišta u Splitu. Istražuje i objavljuje znanstvene radove iz ekonomske statistike, ekonometrije, ocjene parametara nelinearnih funkcija, modelske analize racionalnih očekivanja, modeliranja u uvjetima narušenih osnovnih pretpostavki, statistike u društvenim istraživanjima i kvantitativnih financija. Koautorica je knjige: *Statistika*, udžbenik sa zbirkom zadataka za IV. razred srednje Ekonomske škole, (Zagreb, Alkascript, 2005.). Trenutno sudjeluje u međunarodnom projektu *The impact of the current and future Europe enlargement on a financial integrated market*, Facolta di Economia, Udine, Italija, i projektu MZOS R Hrvatske *Matematički modeli u financijskom upravljanju*. Tajnica je Hrvatskog društva za operacijska istraživanja i članica American Statistical Association.

e-mail: spivac@efst.hr

ANTE ROZGA, rođen 1951. u Splitu. Diplomirao statistiku na Sveučilištu u Zagrebu. Poslijediplomske studije pohađao na Ekonomskom fakultetu u Zagrebu i *The London School of Economics*. Magistrirao i doktorirao iz vremenskih nizova. Izabran kao izvanredni profesor na Ekonomskom fakultetu u Splitu i Medicinskom fakultetu u Zagrebu. Predaje na Ekonomskom fakultetu, FESB-u, Filozofskom fakultetu, te studiju farmacije. Između ostalih držao pozvana predavanja u Hrvatskoj narodnoj banci i na PMF-u Splitu. Gostovao na Sveučilištima u Padovi, Veneciji, Londonu i Cambridgeu. Sudjelovao na više znanstvenih skupova. Član je: *American Statistical Association, Royal Statistical Society, International Association for Statistical Education, Statistical Society of Canada, Bernoulli Society for Mathematical Statistics and Probability*. Područje interesa: ekonomska statistika, ekonometrija, biostatistika, vremenski nizovi, statistika u zaštiti čovjekove okoline, multivarijantna analiza.

e-mail: rozga@efst.hr