Valentina Sokolovska
DESKRIPTIVNA STATISTIKA

Univerzitet u Novom Sadu
Centar za primenjenu statistiku

Valentina Sokolovska

DESKRIPTIVNA STATISTIKA

Novi Sad, 2013. godine

DESKRIPTIVNA STATISTIKA
dr Valentina Sokolovska, docent

Recenzenti
dr Zorana Lužanin, redovni profesor
dr Mirko Savić, redovni profesor
dr Dušan Marinković, redovni profesor

Izdavač
Univerzitet u Novom Sadu
Filozofski fakultet Novi Sad
21000 Novi Sad
Dr Zorana Đinđića 2

Za izdavača
Prof. dr Ivana Živančević Sekeruš

CIP-Каталогизација у публикацији
Библиотека Матице српске, Нови Сад

519.2(075.8)

СОКОЛОВСКА, Валентина
Deskriptivna statistika / Valentina Sokolovska. - Novi Sad : Filozofski fakultet, 2013 (Novi Sad : Grid). - 140 str. : ilustr. ; 24 cm

Tiraž 150. - Bibliografija.

ISBN 978-86-6065-192-3

a) Дескриптивна статистика
COBISS.SR-ID 282308615

Ova publikacija objavljena je u okviru Tempus projekta „Master program iz primenjene statistike" MAS 511140-Tempus-1-2010-1-RS-Tempus-JPCR.
The publishing of this booklet is a part of the Tempus project "Master programme in applied statistics" MAS 511140-Tempus-1-2010-1-RS-Tempus-JPCR.
Ovaj projekat je finansiran od strane Evropske komisije. Ova publikacija odražava samo stavove autora i Komisija ne može biti odgovorna za bilo kakvu upotrebu informacija koje se u publikaciji sadrže.
This project has been funded with support from the European Commission.
This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

SADRŽAJ

PREDGOVOR

Ovaj udžbenik posvećen je isključivo deskriptivnoj statistici i namenjen je najpre studentima društvenih nauka, a zatim i širem krugu korisnika. Pri njegovom pisanju rukovodili smo se potrebom da studentima, koji se u ranijem obrazovanju nisu susretali sa statistikom, na pristupačan način predstavimo postupke za opisivanje statističkih podataka.

Udžbenik sadrži sledeće delove: uvod, mere centriranosti, varijabiliteta i oblika rasporeda, kao i korišćenu literaturu.

U uvodnom delu predstavljen je, u najkraćim crtama, razvoj statistike, opisane su osnovne etape statističkog istraživanja, objašnjene vrste statističkih podataka, dok je najveći deo posvećen sređivanju i grafičkom prikazivanju statističkih podataka.

U drugom delu objašnjavaju se izračunate i pozicione mere centriranosti. Cilj ovog poglavlja je da prikaže način izračunavanja najčešće korišćenih mera centriranosti iz različitih vrsta podataka.

Deo posvećen varijabilitetu podeljen je na varijabilitet kvantitativnih i kvalitativnih podataka. Kako se društvena istraživanja uglavnom susreću sa kvalitativnim merenjima, analiza njihovog varijabiliteta je stoga obavezni deo koji udžbenik iz statistike treba da ima.

Opisivanje podataka prikupljenih istraživanjem završavamo poglavljem sa merama oblika rasporeda. U njemu objašnjavamo mere asimetrije i zaobljenosti.

Pri opisivanju svake mere navode se primeri, koji su većinom objavljeni u publikacijama Republičkog zavoda za statistiku, a namera je da čitaocima predstavimo rezultate na realnim podacima. Svi primeri su detaljno obrađeni, što se može činiti suvišnim, naročito ako se ima na umu dostupnost statističkih softvera u savremenim istraživanjima. No, iskustvo u pedagoškom radu pokazalo je da odabrani primeri i opis postupka olakšava razumevanje, kako primenjenih mera tako i interpretaciju dobijenih rezultata.

Veliku zahvalnost dugujem recenzentima, profesorima Univerziteta u Novom Sadu – dr Zorani Lužanin, dr Mirku Saviću i dr Dušanu Marinkoviću, koji su svojim sugestijama i primedbama umnogome poboljšali ovaj rukopis.

Nadamo se da će ovaj udžbenik omogućiti pre svega studentima da uspešno savladaju mere deskriptivne statistike, koje će im biti osnova za razumevanje i statističkih metoda. Razumevanje mera deskriptivne statistika je osnova za sticanje sigurnosti u njihovoj primeni i tumačenju rezultata.

U Novom Sadu, 2013. godine *Autorka*

UVOD

Razvoj statistike

Statističko istraživanje

Vrste statističkih podataka

Grupisanje i prikazivanje podataka

1 UVOD

1.1 RAZVOJ STATISTIKE

Etimologija pojma statistika potiče od novolatinskog *ratio status* i italijanskog ekvivalenta *regione di stato*, što u prevodu znači državni interes, i izvedenice *statista* (osoba vešta u vođenju državnih poslova) (Šošić, Serdar, 1997). Već samo etimološko razjašnjenje pojma statistika pomaže nam da razumemo njenu ulogu koju je imala do sredine XIX veka. Naime, do tog perioda statistika je obuhvatala podatke koji su bili važni za državu, kao što su brojno stanje stanovnika (vojnika ili poreskih obveznika), imovine itd. Tako su prvi organizovaniji popisi ove vrste sprovedeni još u starom veku u rimskoj republici i oni su se obavljali svake pete godine u određeno vreme i na istom mestu. Prvi takav popis (koji su Rimljani nazvali „cenzus") izvršen je u vreme vladavine Servija Tulija, 550. godine.

Kako se u srednjem veku ekonomska snaga država vezivala prevashodno za stanje u poljoprivredi, zadatak statistike se u narednom periodu proširio. Pored brojčanog stanja stanovnika i njihove imovine, u popise se polako uvode i registri zemljišta i stoke. U pojedinim evropskim zemljama se od XVI veka beleži i broj rođenih, umrlih i venčanih, što će kasnije rezultirati razvojem statistike prirodnog kretanja stanovništva. Rezultati ovih popisa danas imaju veliki istorijski značaj. Na osnovu njih može se pratiti ne samo razvoj poljoprivrede već i migratorna kretanja stanovništva u XVIII i XIX veku. Jedna od značajnijih studija posvećena demografskoj i agrarnoj statistici Vojvodine od 1767. do 1867. godine, nastala je upravo na osnovu podataka iz ovih popisa (vidi Hegediš, Čobanović, 1991). Krajem XVIII i početkom XIX veka, jačanjem industrije popisi su, osim prebrojavanja stanovništva, poljoprivrednih gazdinstava i stoke, počeli da beleže i statistiku spoljne trgovine, cena, javnih finansija i ostalih ekonomskih pokazatelja. Tada nastaju prve službe koje su počele da objavljuju statističke publikacije.

U XVII veku takođe počinju da se osnivaju i prva statistička društva. Kraljevsko društvo (The Royal Society), osnovano 1660. godine u Londonu, Volter Vilkoks (Willcox, 1934) ocenjuje kao kolevku statistike u sadašnjem smislu te reči. Nakon njega, u XIX veku osnovano je 40 statističkih udruženja širom sveta.

Na prostorima bivše Jugoslavije statistika počinje da se razvija sredinom XVIII veka (vidi Zarkovich, 1990). Dok se u drugim zemljama Zapadne Evrope njen razvoj vezuje za potrebe nauke, u bivšoj Jugoslaviji razvoj statistike je bio podstaknut političkim programom Austrije, kojim je trebalo da bude unapređeno obrazovanje i modernizovana državna uprava. Kao rezultat ovakve politike, statistika je uvedena u univerzitetske programe, najpre u Hrvatskoj, a državna administracija je zauzvrat očekivala podršku u prikupljanju podataka i njihovo korišćenje pri donošenju državnih odluka. Slični događaji odvijali su se nedugo zatim i u Srbiji, ali kao rezultat kulturnog uticaja Austrije i Mađarske. Kako su mladi ljudi iz Srbije odlazili na univerzitete u Austriju i Mađarsku, tamo su imali priliku da se upoznaju sa idejama o statistici.

Tokom ovog perioda, tačnije od 1700. do 1900. godine, statistika u svetu doživljava horizontalni i vertikalni razvoj, objašnjava Stigler (Stigler, 1986). Horizontalni, jer se statistički metodi šire među naukama, od astronomije i geodezije, biologije do društvenih nauka, dok se vertikalni može objasniti promenama u shvatanju uloge verovatnoće. Naime, razumevanje verovatnoće je, u ovom periodu, napredovalo od poistovećivanja sa igrama na sreću do uvođenja inverzne verovatnoće i početaka statističkog zaključivanja.

Savremena statistika je mnogo više od alatke, „torbe trikova" ili mešavine izolovanih tehnika korisnih u pojedinim naukama. Postoji jedinstvo u pogledu upotrebe statističkih metoda. Isti kompjuterski program koji analizira podatke mogu koristiti ekonomisti, hemičari, sociolozi, psiholozi i ostali naučnici. Iako tumačenja statističke analize mogu suptilno da se razlikuju među naukama, koncepti i procedure tih interpretacija, njihove logičke posledice i ograničenja su isti za sve. Zbog toga, empirijskim naukama moderna statistika postaje logika i metodologija za merenje neizvesnosti u planiranju i tumačenju eksperimenata i posmatranja.

1.2 STATISTIČKO ISTRAŽIVANJE

Statističko istraživanje podrazumeva ispitivanje varijacija masovnih pojava sa ciljem otkrivanja njihovih pravilnosti i zakonitosti. Ukoliko ova vrsta istraživanja ne bi imala karakteristiku masovnosti bilo bi nemoguće otkriti bilo koju pravilnost ili zakonitost proučavanih pojava. Baviti se statističkim istraživanjem podrazumeva analizirati podatke koji prema izvoru mogu biti primarni ili sekundarni. Primarni podaci se prikupljaju u skladu sa određenim ciljevima istraživanja, dok se analiza bazirana na sekundarnim podacima zasniva na proučavanju već prikupljenih i sređenih podataka nekog drugog istraživanja.

Prikupljanju primarnih podataka o nekoj pojavi predstoje važni istraživački koraci koje je potrebno sprovesti. Oni se sastoje od formulisanja:

- cilja i predmeta istraživanja, kao i istraživačkih hipoteza,
- statističkog skupa i jedinica posmatranja,
- izbora obeležja i njihovih modaliteta,
- upitnika,
- načina prikupljanja podataka i
- analize i prezentacije rezultata istraživanja.

Kako su mnogi od ovih koraka predmet proučavanja metodologije naučnih istraživanja, njima se na ovom mestu nećemo baviti. Našu pažnju usmerićemo na definisanje statističkog skupa i jedinica posmatranja, izbor promenljive i njenih modaliteta, kao i načinu prikupljanja i prezentacije podataka.

Statistički skup sastavljen je od jedinica posmatranja, koje mogu biti: osobe, stvari, institucije, biljke i tako dalje. Dakle, sve ono što može biti predmet istraživanja. Prema opsegu, statistički skup može biti konačan ili beskonačan, ali i realan ili hipotetičan. U konačne statističke skupove ubrajamo sve one kod kojih je moguće prebrojati jedinice posmatranja, za razliku od beskonačnih, gde prebrojavanje nije moguće. Realni statistički skupovi su sastavljeni od jedinica koje egzistiraju, dok se hipotetičke jedinice definišu određenim pravilom ili modelom.

Definisati određeni statistički skup znači odrediti ga pojmovno, prostorno i vremenski. Na taj način se određuju opšta svojstva koje mora imati svaka jedinica posmatranja kako bi mogla biti član određenog skupa. Pojmovnim definisanjam se utvrđuje pripadnost nekom statističkom skupu s obzirom na pojam jedinice. Na primer, statistički skup punoletnih osoba podrazumeva sve osobe koje su napunile 18 godina i starije. Skup studenata čine sve osobe koje su upisane na visokoškolskim institucijama i time stekle određena prava i obaveze. Prostornim definisanjem se određuje prostor kojem pripadaju sve jedinice jednog statističkog skupa. Ukoliko posmatramo prethodno definisani skup studenata, onda ćemo ga prostornim definisanjem još jasnije odrediti. Recimo, da se radi o svim studentima koji su upisani na visokoškolskim ustanovama u Republici Srbiji. I na kraju, vremenskom definicijom statističkog skupa određujemo vreme u kome će se istraživati sve jedinice statističkog skupa. To može biti jedan vremenski trenutak ili vremenski interval. Na primer, vremenskim trenutkom su definisani popisi stanovništva, jer se sprovode svake desete godine. Vremenskim intervalom se mogu odrediti, recimo, istraživanja iz oblasti turizma zasnovana na prikupljanju podataka o turistima u letnjem periodu.

Pojmovnim, prostornim i vremenskim definisanjem statističkih skupova zadovoljava se princip homogenosti i samo takvi skupovi mogu biti predmet statističke analize. Međutim, predmet statističkog istraživanja nisu jedinice skupova same po sebi, već samo njihova svojstva, tj. obeležja (promenljive)[1]. Ukoliko se skup podataka o posmatranim promenljivama prikuplja za svaku jedinicu statističkog skupa govorimo o osnovnom skupu (populaciji). Broj prikupljenih podataka iz osnovnog skupa obeležavamo velikim slovom N. Kako ovakva istraživanja iziskuju dosta vremena i

[1] Za određena svojstva jedinice skupova koja se prikupljaju statističkim istraživanjem često se koriste nazivi obeležje, varijabla ili promenljiva, što su sve sinonimi. Zbog toga ćemo u daljem tekstu koristiti izraz promenljiva.

finansijskih sredstava, često se podaci proučavanih promenljivih prikupljaju na podskupu jedinica, tj. na osnovu uzorka (veličina uzorka obeležava se malim slovom *n*). On je osnova za zaključivanje o populaciji i po svojim karakteristikama bi trebalo verno da odražava karakteristike osnovnog skupa.

1.3 VRSTE STATISTIČKIH PODATAKA

1.3.1 Merne skale

Podaci u statistici nastaju merenjem određenih svojstava jedinica statističkih skupova ili njihovih podskupova (uzoraka). Ispitivana svojstva nazivaju se promenljive i mogu da se pojavljuju u dva ili više modaliteta. Merenje se uvek odvija prema određenom pravilu i zahteva posebnu skalu sa određenim jedinicama mere. Ukoliko se posmatraju metrička svojstva skale možemo ih podeliti na nominalne, ordinalne, intervalne i skale odnosa. Prve dve vrste skala služe za merenje kvalitativnih promenljivih, dok se intervalna i skala odnosa koriste za merenje kvantitativnih podataka.

Nominalna skala data je u obliku kategorijalnih jedinica. Ukoliko se njome opisuje izraženi varijabilitet određenog svojstva govori se o atributivnoj nominalnoj skali (bračno stanje: neoženjen-neudata, oženjen-udata, udavac-udovica, razveden-razvedena i nepoznato), dok se navođenjem prostornih jedinica stvara geografska nominalna skala (npr. okruzi u Vojvodini: severnobački, srednjebanatski, severnobanatski, južnobanatski, zapadnobački, južnobački i sremski). U pojedinim slučajevima nominalnoj skali se mogu pridružiti i brojevi, ali oni služe jedino kao indikatori (muško – 1, žensko – 2). Kako brojevi predstavljaju jedino simbole a ne kvalitet određenog modaliteta, ne postoji kriterijum prema kojem bi se vredosti mogle odrediti kao „veće od”, ili „manje od” drugih vrednosti. Ova vrsta skale upotrebljava se jedino za razvrstavanje podataka u kategorije. Ukoliko se u skali pojavi veliki broj modaliteta (na primer, vrsta delatnosti, bolesti i slično) onda se mogu koristiti nomenklature s ciljem svrstavanja obeležja u slične podgrupe. Od mera deskriptivne statistike, za analizu nominalne skale možemo koristiti jedino modus, kako se savetuje u mnogim udžbenicima statistike.

Ordinalna merna skala služi za označavanje redosleda, pa je njena svrha jedino određivanje da li je nešto veće ili manje od drugoga. Dakle, merenje se ovde svodi na rangiranje prema značaju modaliteta. Na primer, cenu nekog proizvoda možemo označiti kao izuzetno povoljnu, povoljnu i nepovoljnu. Tipičan primer ordinalne skale su školske ocene: odličan, vrlodobar, dobar, dovoljan i nedovoljan. Osim modusa, za obradu ordinalnih skala možemo koristiti i medijanu. Kako je upotreba statističkih mera za analizu nominalne i ordinalne merne skale strogo ograničena, u praktičnom radu se često prave greške pri analizi ovih podataka.

Intervalna merna skala poseduje određene jedinice mere i dogovoreno utvrđenu nulu. Za razliku od prethodnih skala, intervalna dozvoljava klasifikaciju, rangiranje i brojčano utvrđivanje razlika u jedinicama mere u kojoj je obeležje izraženo. To je omogućeno jer je definisana razlika između brojeva u skali jednaka na svakom delu skale. Tipični primeri intervalnih skala su temperature vazduha, visine vazdušnog pritiska... Nedostatak ovih skala je nepostojanje apsolutne nule.

Skala odnosa pored toga što poseduje sve osobine intervalne skale ima i apsolutnu nulu kojom se ukazuje na nepostojanje merenog svojstva. Skale odnosa su dužina, visina, godine starosti, ukupan profit, itd. Na modalitetima skale odnosa mogu se sprovoditi sve računske operacije (sabiranja, oduzimanja, množenja i delenja), te zato ona ima najbolja metrička svojstva.

1.3.2 Vrste promenljivih

S obzirom na to koja je merna skala upotrebljena za merenje, promenljive možemo podeliti na kvalitativne i kvantitativne. Promenljive merene na nominalnoj i ordinalnoj mernoj skali ocenjujemo kao kvalitativne (kategorijalne), dok se intervalna i skala odnosa koriste za merenje kvantitativnih (numeričkih) promenljivih. Klasifikaciju možemo izvršiti i prema vrednostima koje specifična promenljiva može imati. Prema ovom kriterijumu razlikujemo diskretne i kontinuirane promenljive. Diskretne su one koje mogu imati samo određeni skup vrednosti koje su međusobno jasno razdvojene (pol, nacionalna pripadnost itd.), dok kontinuirane mogu da imaju bilo koju vrednost u okviru ranga koji predstavlja limit te promenljive (visina, težina i dr.). Kako različite promenljive proizvode drugačije vrste podataka, njihovo odabiranje je od presudnog su značaja za izbor statističkog metoda kojim ćemo ih analizirati.

1.4 GRUPISANJE I PRIKAZIVANJE PODATAKA

Nakon prikupljanja podataka u prvom koraku statističke analize pristupa se njihovom sređivanju. Dobijeni skup podataka se ovim postupkom raščlanjuje na podskupove (modalitete promenljive), koji treba da su definisani tako da se međusobno ne preklapaju. Poredak modaliteta je teorijski proizvoljan, ali, radi preglednosti, istraživači ih često ređaju prema abecednom (azbučnom) redu, opadajućoj vrednosti frekvencija, već ustanovljenoj nomenklaturi i slično.

Ukoliko se razvrstavanje podataka vrši prema modalitetima jedne promenljive govori se o jednodimenzionalnom skaliranju. Pri razvrstavanju podataka istraživači moraju poštovati principe isključivosti i iscrpnosti. Već smo napomenuli da se pojedini podatak može razvrstati samo u jedan podskup i to je načelo isključivosti, dok iscrpnost znači da svaki podatak mora biti razvrstan u jedan od podskupova. Broj podataka koji pripada određenom podskupu (modalitetu) nazivamo njegovom frekvencijom, a zbir frekvencija svih modaliteta jednak je ukupnom broju prikupljenih podataka.

1.4.1 Grupisanje podataka kvalitativnih promenljivih

Primer 1

U tabeli 1.1 prikazani su konačni rezultati popisa obavljenog 2002. godine o izjašnjavanju stanovništva Republike Srbije prema maternjem jeziku.

Tabela 1.1 *Raspodela odgovora stanovništva R Srbije prema maternjem jeziku (2002)*

Maternji jezik	frekvencija
srpski jezik	6620699
albanski jezik	63835
bosanski jezik	134749
bugarski jezik	16459
vlaški jezik	54818
mađarski jezik	286508
makedonski jezik	14355
romski jezik	82242
rumunski jezik	34515
slovački jezik	57498
hrvatski jezik	27588
ostali jezici	40858
nije se izjasnio i nepoznato	63877
Ukupno	7498001

Izvor: *Popis stanovništva, domaćinstva i stanova u 2002. Stanovništvo. Veroispovest, maternji jezik i nacionalna ili etnička pripadnost prema starosti i polu* (2003). Beograd: Republički zavod za statistiku, str. 16.

Prva kolona ove tabele opisuje promenljivu „Maternji jezik" i ima 13 modaliteta, dok je druga kolona posvećena raspodeli odgovora ukupno popisanog stanovništva. Ono što je važno napomenuti je i način prikazivanja odgovora. U tabeli br. 1.1 raspodela odgovora je prikazana u apsolutnim frekvencijama (f_i). Međutim, radi jednostavnijeg uvida u strukture proučavane pojave vrlo često se podaci statističkih istraživanja prikazuju pomoću relativnih frekvencija (*fr*), od kojih se nadalje mogu izračunati i procenti (tabela 1.2). Relativne frekvencije jednog modaliteta dobijamo kada njenu apsolutnu frekvenciju podelimo sa ukupnim zbirom svih frekvencija. One pokazuju koliki deo od ukupnog broja jedinica posmatranja pripada odgovarajućem modalitetu. Pomnožene sa 100, daju procentualnu raspodelu odgovora.

$$fr = \frac{f_i}{\Sigma f_i} \qquad [1.1]$$

Zbir svih relativnih frekvencija je 1, dok je zbir svih procenata raspodeljenih po modalitetima jedne promenljive 100 (vidi tabelu 1.2). Osim toga, relativne frekvencije se koriste za upoređivanje grupisanih nizova više skupova prema modalitetima iste promenljive.

Tabela 1.2 *Apsolutna, relativna i procentualna raspodela odgovora stanovništva Republike Srbije (2002) prema maternjem jeziku*

Maternji jezik	f_i	fr[2]	%
srpski jezik	6620699	0,88	88,30
albanski jezik	63835	0,01	0,85
bosanski jezik	134749	0,02	1,80
bugarski jezik	16459	0,00	0,22
vlaški jezik	54818	0,01	0,73
mađarski jezik	286508	0,04	3,82
makedonski jezik	14355	0,00	0,19
romski jezik	82242	0,01	1,10
rumunski jezik	34515	0,00	0,46
slovački jezik	57498	0,01	0,77
hrvatski jezik	27588	0,00	0,37
ostali jezici	40858	0,01	0,54
nije se izjasnio i nepoznato	63877	0,01	0,85
Ukupno	7498001	1,00	100,00

[2] Kako su relativne frekvencije zaokruživane na drugoj decimali, kod pojedinih kategorija (kao što je: bugarski, makedonski, rumunski i hrvatski) njihova relativna učešća su prikazana 0.

1.4.2 Grupisanje podataka kvantitativnih promenljivih

Način uređivanja kvantitativnih podataka zavisi od broja podataka i od toga da li je promenljiva diskretna ili kontinuirana.

Primer 2

Naredna tabela (tabela 1.3) prikazuje nesređen niz ukupnog broja poena predispitnih obaveza koje je 50 studenata steklo na predmetu Statistika. „Broj bodova” predstavlja primer kvantitativne diskontinuirane promenljive.

Tabela 1.3 *Negrupisani podaci*

0	17	31	10	14	11	21	35	18	26
10	14	30	21	30	24	23	21	17	27
27	33	10	16	22	11	29	18	25	46
12	0	6	34	0	6	9	15	26	23
14	22	32	12	6	23	20	0	27	34

Kvantitativni podaci se mogu srediti na isti način kao i kvalitativni, to jest prebrojavanjem učestalosti (frekvencija) svakog modaliteta jedne promenljive. Kako u ovom slučaju broj poena predstavlja jedan modalitet, tabela učestalosti (tabela 1.4) će imati veoma mnogo kategorija.

Tabela 1.4 *Grupisani podaci*

Broj bodova	Apsolutna frekvencija
0	4
6	3
9	1
10	3

11	2
12	2
14	3
15	1
16	1
17	2
18	2
20	1
21	3
22	2
23	3
24	1
25	1
26	2
27	3
29	1
30	2
31	1
32	1
33	1
34	2
35	1
46	1
Ukupno	50

Kada promenljiva ima veliki broj kategorija, distribucija njihovih frekvencija može biti dosta rasuta. Osim toga, ovako sređeni statistički niz je često veoma nepregledan zbog velikog broja kategorija. Zbog toga se u statistici numerički nizovi često prikazuju pomoću intervalnih serija. One su sastavljene od određenog broja intervala koji u sebe uključuju sve vrednosti između donje i gornje granice tog intervala. Intervali se moraju odrediti na takav način da se gornja granica jednog ne sme preklapati sa donjom granicom narednog intervala. Na primer, ako prvi interval obuhvata vrednosti od 1 do 4, onda naredni treba početi sa 5, ukoliko se radi o prekidnim i diskretnim numeričkim podacima. U slučaju neprekidnih numeričkih podataka donju granicu narednog intervala bismo počeli sa 4,1. Broj grupnih intervala (c) u raspodeli frekvencija zavisi od ukupnog broja podataka. Da se broj intervala ne bi određivao proizvoljno, njegovo utvrđivanje se obično vrši pomoću Stardžesovog pravila.

$$c = 1 + 3{,}3 \cdot \log n \qquad [1.2]$$

gde je n broj podataka u seriji.

Na osnovu ovog pravila može se izračunati i širina intervala (i) pomoću sledećeg obrasca:

$$i = \frac{x_{max} - x_{min}}{c} \qquad [1.3]$$

Ukoliko se obrasci [1.2] i [1.3] primene na primer br. 2, dobijamo intervalnu seriju sa 7 intervala, širine 7.

$$1 + 3{,}3 \cdot \log 30 \approx 7$$

$$\frac{46 - 0}{7} \approx 7$$

Za numeričku analizu intervalnih serija potrebno je izračunati i sredine intervala (x'_i). Intervalna sredina jednaka je aritmetičkoj sredini njegove donje i gornje granice.

$$x'_i = \frac{x_{min} + x_{max}}{2} \qquad [1.4]$$

Na osnovu obrasca [1.4] dobijamo da je sredina prvog interval 3 (jer je $\frac{0+6}{2} = 3$), sredina drugog 10 i tako dalje (tabela 1.5). Sredine intervala su neophodne radi dalje analize ovakve numeričke serije i njihovu ulogu ćemo detaljnije objasniti u narednom poglavlju.

Tabela 1.5 *Intervalna serija*

Grupni intervali	frekvencije	x_i'
0–6	7	3
7–13	8	10
14–20	10	17
21–27	15	24
28–34	8	31
35–41	1	38
42–48	1	45
Ukupno	50	

U ovom slučaju radi se o zatvorenim intervalima jednake veličine. Oni se koriste za grupisanje približno simetričnih podataka. Međutim, grupisanje podataka često se vrši i pomoću poluotvorenih intervala (tabela 1.6).

Tabela 1.6 *Umrli usled samoubistva s obzirom na starost, 2011. godine na teritoriji Republike Srbije*

starost	do 14	15–24	25–34	35–44	45–54	55–64	65–74	75 i više
frekvencija	2	39	98	133	220	242	216	306

Izvor: *Demografska statistika u Republici Srbiji 2011* (2011). Beograd: Republički zavod za statistiku, str. 187.

Kao što primećujete iz tabele 1.6, prvi i poslednji interval nemaju jednu od granica: prvi donju, a poslednji gornju. Da bismo mogli da nastavimo analizu ovakvih podataka potrebno je da se nedostajuće granice poluotvorenih intervala procene. Njihova procena vrši se od slučaja do slučaja. U prvom intervalu donja granica će iznositi (5), dok će gornja granica poslednjeg intervala iznositi (84). Procenjene vrednosti se obično stavljaju u zagradu kako bi se skrenula pažnja na njih.

1.4.3 Grafičko prikazivanje podataka

Osim tabelarnog prikazivanja podataka u statistici se koriste i različiti oblici grafikona. Neki od njih su primereni za prikazivanje statističkih nizova kvalitativnih promenljivih, dok ostali mogu poslužiti isključivo za prikazivanje kvantitativnih podataka. Pored tabela, kvalitativni podaci se mogu grafički prikazati pitom, kartogramom, histogramom, Paretovim dijagramom i mnogim drugim grafičkim oblicima. Ukoliko se promenljiva sastoji od malog broja modaliteta tada se za njeno grafičko prikazivanje može, između ostalog, koristiti i pita (strukturni krug) jer pruža pregledan vezuelni efekat. Geografske kvalitativne promenljive se obično prikazuju kartogramima, dok se za prikazivanje frekvencija kvalitativne promenljive sa više modaliteta koristi histogram.

Pita (*pie charts*/*circle graphs*) je podeljena na delove od kojih svaki deo predstavlja frekvenciju (apsolutnu ili relativnu) ili proporciju odgovarajuće kategorije u osnovnom skupu ili uzorku.

Primer 3

Prema rezultatima popisa iz 2011. godine na teritoriji Republike Srbije popisan je ukupno 7.498.001 stanovnik, od kojih je 3.645.930 stanovnika bilo muškog a 3.852.071 ženskog pola. Ovakva raspodela stanovništva prema polu (u procentima) može se grafički predstaviti pomoću pite (grafikon 1.1).

Grafikon 1.1. *Prikaz nominalnog niza pomoću pite*

Izvor: *Opštine i regioni u Republici Srbiji 2011* (2011). Beograd: Republički zavod za statistiku str. 38

Pritom je prikazana struktura stanovništva Srbije u 2011. godini prema polu. Broj stepeni svakog dela kruga (broj stepeni sektora kruga), za apsolutne frekvencije, određuje se izrazom:

$$S_i = \frac{f_i}{N} \cdot 360 \qquad [1.5]$$

Ukoliko pitom prikazujemo seriju izraženu procentima, onda ćemo deo kruga izračunati na sledeći način:

$$s_i = P_i \cdot 3{,}60 \qquad [1.6]$$

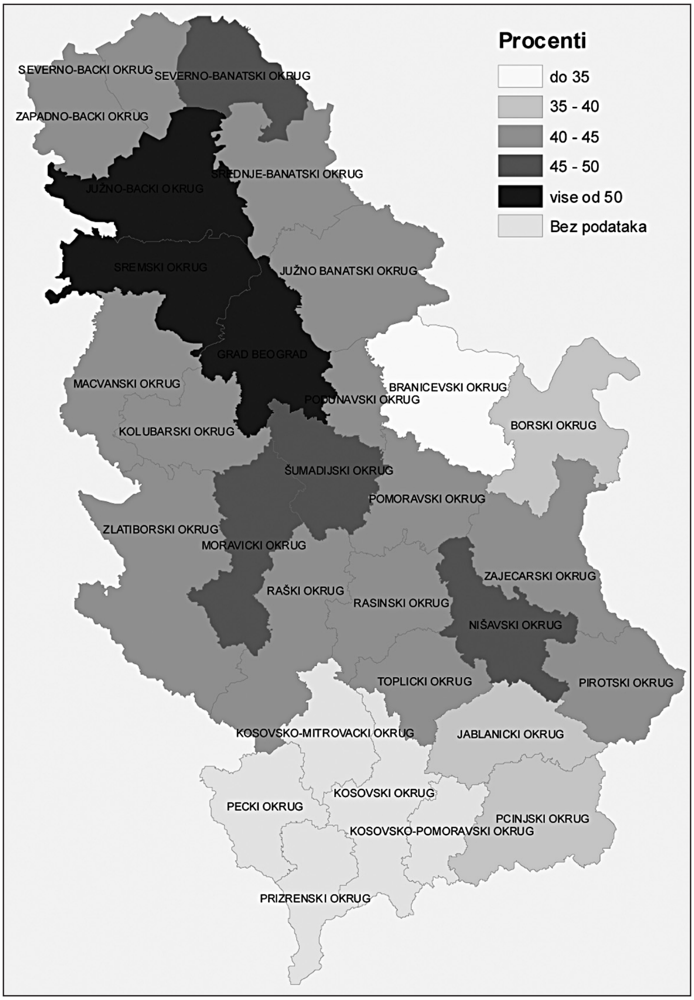
Pita se može koristiti i za poređenje struktura dvaju ili više statističkih nizova. U ovom slučaju potrebno je nacrtati krugove istih radijusa. Razlike između struktura dobijaju se poređenjem veličina sektora kruga kojima se predstavljaju isti modaliteti.

Kartogrami služe za grafičko prikazivanje frekvencija geografskih kvalitativnih promenljivih. Dakle, ukoliko istražujemo neku pojavu po geografskim oblastima, takvu promenljivu možemo grafički prikazati (osim strukturnim krugom i histogramom) i pomoću kartograma.

Primer 4

U narednom grafikonu (grafikon 1.2.), na osnovu popisa iz 2002. godine, prikazan je udeo migrantskog u ukupnom stanovništvu Republike Srbije po okruzima.

Grafikon 1.2. *Prikaz nominalnog geografskog niza*



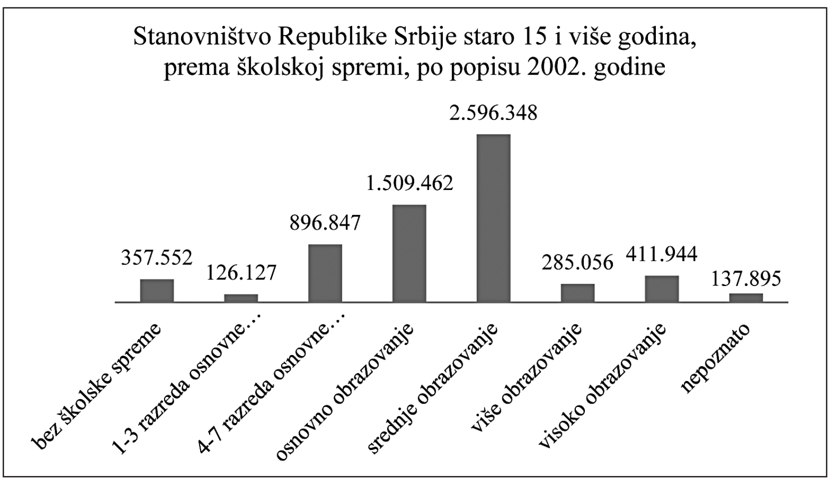
Izvor: http://webrzs.stat.gov.rs/axd/savetovanje/Spasovski9.pps

Histogramom se, zbog preglednosti, obično prikazuju nominalne promenljive sa većim brojem modaliteta. On se konstruiše na sledeći način: na x-osu (apscisu) nanose se različite kategorije kvalitativne promenljive koje su predstavljene intervalima iste širine, dok se frekvencije nanose na y-osu (ordinatu). Za svaku kategoriju ucrtava se po jedan stubac čija visina odgovara frekvenciji (apsolutnoj, relativnoj ili procentualnoj raspodeli) posmatrane kategorije.

Primer 5

U cilju proučavanja obrazovne strukture stanovništva Republike Srbije, 2002. godine popisan je ukupno 6.321.231 stanovnik, star 15 i više godina. Od ukupnog broja popisanih lica, 357.552 je bez školske spreme, 126.127 je završilo od 1 do 3 razreda osnovne škole, njih 896.847 od 4 do 7 razreda osnovne škole, dok 1.509.462 ima osnovno obrazovanje. Srednju školu je završilo 2.596.348 stanovnika, višu 285.056, visoko obrazovanje je steklo 411.944 stanovnika, dok je za njih 137.895 školska sprema nepoznata. U narednom grafikonu (grafikon 1.3) predstavljene su apsolut-

Grafikon 1.3. *Prikaz apsolutnih frekvencija pomoću histograma*



Izvor: *Popis stanovništva, domaćinstva i stanova u 2002. Stanovništvo. Školska sprema i pismenost*, (2003). Beograd: Republički zavod za statistiku, str. 14–15 ne frekvencije stanovništva starog 15 i više godina, prema školskoj spremi, po popisu iz 2002. godine.

Histogram ne služi samo za prikazivanje rezultata istraživanja kvalitativnih promenljivih. Njega možemo koristiti i za prezentovanje rezultata kvantitativnih merenja.

Paretov dijagram (*Pareto chart*) služi za prikazivanje značenja pojedinih događaja (pomoću frekvencija ili procenata) razvrstanih po veličini. Klasifikovanje slučajeva prema stepenu važnosti (udela) omogućuje da se pažnja usmeri na rešavanje najvažnijih problema (uzroka koji izazivaju neku pojavu), ne ulazeći pri tome u analizu manje važnih. Ova analiza zasnovana je na načelu „80/20”[3], prema kome je približno 80% problema izazvano sa 20% uzroka. Ili, drugačije rečeno, 80% aktivnosti daje 20% rezultata, dok 20% najvažnijih (ključnih) aktivnosti izazove 80% rezultata.

Primer 6

U narednoj tabeli (tabela 1.7) prikazana su umrla lica muškog pola, na teritoriji Republike Srbije, prema uzroku smrti u 2009. godini. Da bismo Paretovom analizom utvrdili kojih to ključnih 20% uzroka tj. bolesti izaziva smrtni ishod kod 80% muškaraca, raspodelu frekvencija smo najpre izrazili procentualno, a potom smo izračunali i kumulativni niz procentualne raspodele. Kumulativni niz izračunat je postupnim zbrajanjem procenata. Prva vrednost kumulativnog niza je 49,1% (49,1% smrti izazvano je bolestima krvotoka), druga vrednost je 72,4% (nastala je zbrajanjem prve i druge strukture (49,1 + 23,3) i označava da je 72,4% smrtnih slučajeva kod muškog pola izazvano bolestima krvotoka i tumorima). Treća vrednost izno-

[3] Paretova analiza dobila je naziv po italijanskom sociologu i ekonomisti Vilfredu Paretu (1848–1923) koji je 1896. godine otkrio da raspodela prihoda i bogatstva ne sledi normalnu raspodelu, već je uglavnom nakrivljena. U svojoj prvoj knjizi *Cours d`economie politique* on objašnjava da se 80% bogatstva nalazi u rukuma 20% populacije. Ova relacija je kasnije nazvana „pravilo 80/20” ili „Paretov princip”.

si 77,6% (nastala je zbrajanjem prve tri strukture 49,1 + 23,3 + 5,2 = 77,6), a Paretovom analizom bismo stali do četvrte vrednosti kumulativnog niza 82,8. Njome bismo objasnili da je u 82,8% slučajeva smrt kod pripadnika muškog pola na teritoriji Republike Srbije, u 2009. godini, izazvana: bolestima krvotoka; tumorima; simptomima, znacima i nenormalnim kliničkim i laboratorijskim nalazima, neklasifikovanim na drugom mestu; kao i povredama, trovanjima i posledicama delovanja spoljnih faktora. Ove četiri vrste bolesti čine 25% od ukupnih uzroka koji dovode do smrtnog ishoda.

Tabela 1.7 *Umrla lica muškog pola, na teritoriji Republike Srbije, prema uzroku smrti u 2009. godini*

Vrsta bolesti	Međunarodna klasifikacija bolesti	Frekvencija	Struktura u %	Kumulativni niz u %
Bolesti krvotoka	I00-I99	25.740	49,1	49,1
Tumori	C00-D48	12.213	23,3	72,4
Simptomi, znaci i i nenormalni...[4]	R00-R99	2.731	5,2	77,6
Povrede...[5]	S00-T98	2.709	5,2	82,8
Bolesti disajnih puteva	J00-J98	2.589	4,9	87,7
Bolesti organa za varenje	K00-K92	2.029	3,9	91,6
Bolesti žlezda...[6]	E00-E88	1.322	2,5	94,1

[4] Simptomi, znaci i nenormalni klinički i laboratorijski nalazi, neklasifikovani na drugom mestu.

[5] Povrede, trovanja i posledice delovanja spoljnih faktora

[6] Bolesti žlezda s unutrašnjim lučenjem, ishrane i metabolizma

Bolesti mokraćno-polnog sistema	N00-N98	1.184	2,3	96,4
Bolesti nervnog sistema	G00-G98	734	1,4	97,8
Duševni poremećaji	F00-F99	499	1,0	98,8
Zarazne i parazitske bolesti	A00-B99	254	0,5	99,3
Stanja nastala u perinatalnom periodu	P00-P96	175	0,3	99,6
Urođene anomalije	Q00-Q99	70	0,1	99,7
Bolesti krvi ...[7]	D50-D89	68	0,1	99,8
Bolesti mišićno...[8]	M00-M99	36	0,1	99,9
Bolesti kože i potkožnog tkiva	L00-L98	24	0,1	100
Ukupno		52.377	100,00	

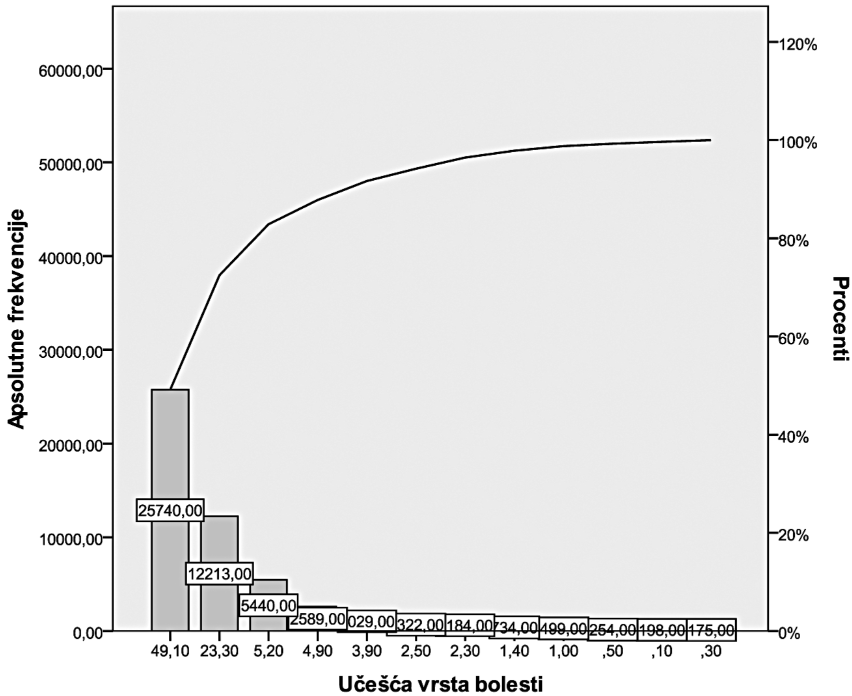
Izvor: *Demografska statistika 2009* (2010). Beograd: Republički zavod za statistiku, str. 100.

Na osnovu Paretovog dijagrama (dijagram 1.4) možemo videti da učešća pojedinih bolesti sa smrtnim ishodom nisu jednolična, već asimetrična, usled manjeg broja uzroka. Podloga za crtanje ovog dijagrama je niz čiji su elementi uređeni prema opadajućoj vrednosti frekvencija. Grafikon je kombinacija histograma (u kojem su prikazane frekvencije) i linijskog dijagrama (kojim je dat prikaz kumulativni procentualni niz).

7 Bolesti krvi i krvotvornih organa i neki poremećaji imunog sistema

8 Bolesti mišićno-koštanog sistema i vezivnog tkiva

Grafikon 1.4 *Paretov dijagram umrlog muškog stanovništva u R Srbiji, 2009. godine*



Dijagram „stabljika i list” (*Stem and Leaf Plot*) upotrebljava se za prikaz pojedinačnih vrednosti numeričke promenljive. Konstruiše se tako što se podaci najpre urede po veličini, a zatim se određuju vrednosti „stabljike” i „listova”. Ako su vrednosti serije dvocifreni brojevi, stabljiku čine desetine a listove jedinice. Kod trocifrenih brojeva, stabljiku mogu činiti stotine a listove desetice i jedinice. Dijagram se na isti način formira i kada seriju čine decimalni brojevi. Na primer, za podatak 62, stabljika je 6 a list 2.

U narednoj tabeli (tabela 1.8) prikazane su godine starosti 20 gostiju koji su odseli u jednom hotelu.[9]

[9] Podaci su preuzeti iz: Angel et al. 2005: 773.

Tabela 1.8 *Godine starosti*

27	28	29	31	32
39	43	44	44	45
47	50	56	58	59
60	62	68	71	72

Na osnovu njih konstruisaćemo dijagram „stabljika-list" (grafikon 1.5).

Grafikon 1.5. *Godine starosti pomoću stabljika-list dijagrama*

Frekvencije	Stabljika & List
3	2 | 789
3	3 | 129
5	4 | 34457
4	5 | 0689
3	6 | 028
2	7 | 12

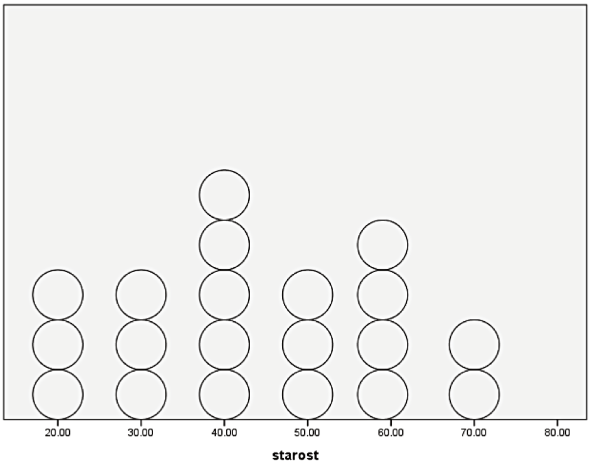
Širina stabljike: 10
Svaki list: 1 podatak

Iz ovog dijagrama možemo videti da su najzastupljeniji gosti 40-ih godina starosti, potom 50-ih, kao i da samo dva gosta imaju preko 70 godina. Dijagram „stabljika-list" ima isti vizuelni efekat kao i histogram.

Tačkasti dijagram (*Scatter Dot*) crta se tako što se na x osi označe modaliteti promenljive, a zatim se tačkama predstavljaju vrednosti članova numeričkog niza. I ovaj grafikon ima isti vizuelni efekat kao i histogram. Ako se u datom nizu nalaze dve ili više jednakih vrednosti, onda se tačke ucrtavaju jedna iznad druge. Na grafikonu 1.6 prikazane su vrednosti prethodnog dijagrama stabljika-list (grafikon 1.5)

U praksi se veoma često susrećemo sa potrebom istovremenog prikazivanja podataka prema modalitetima dve ili više promenljivih. Oni se obično prikazuju kombinovanom tabelom, koja se još naziva i tabelom konti-

Grafikon 1.6 *Prikaz tačkastog dijagrama*



gencije. Tabela se satoji od zaglavlja, pretkolone, polja tabele marginalnog reda i marginalne kolone. U marginalnoj koloni se prikazuju frekvencije jedne promenljive, dok su u marginalnom redu prikazane frekvencije druge promenljive.

Primer 7

U tabeli 1.9 prikazano je nepismeno stanovništvo staro 15–24 godine, prema polu, popis 2002.

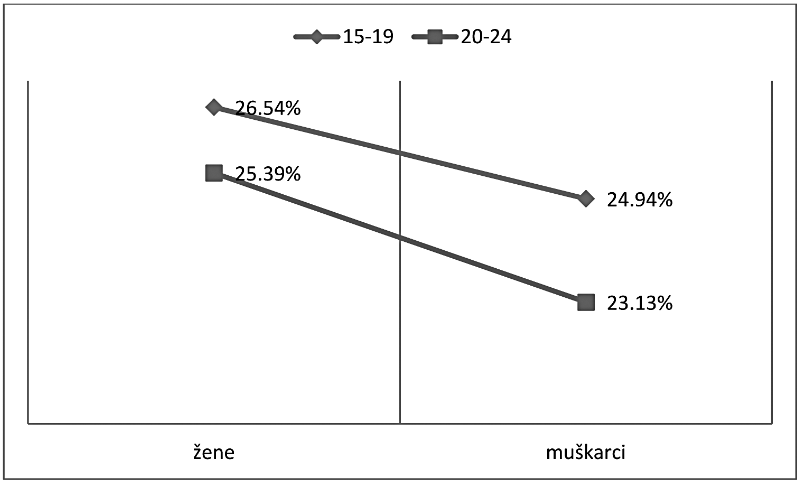
Tabela 1.9 *Nepismeno stanovništvo staro 15–24 godine, prema polu, 2002. godine*

	Žene		Muškarci		Ukupno	
Starost	f_i	%	f_i	%		%
15–19	1.721	26,54	1.617	24,94	3.338	51,48
20–24	1.646	25,39	1.500	23,13	3.146	48,52
Ukupno	3.367	51,93	3.117	48,07	6.484	100,00

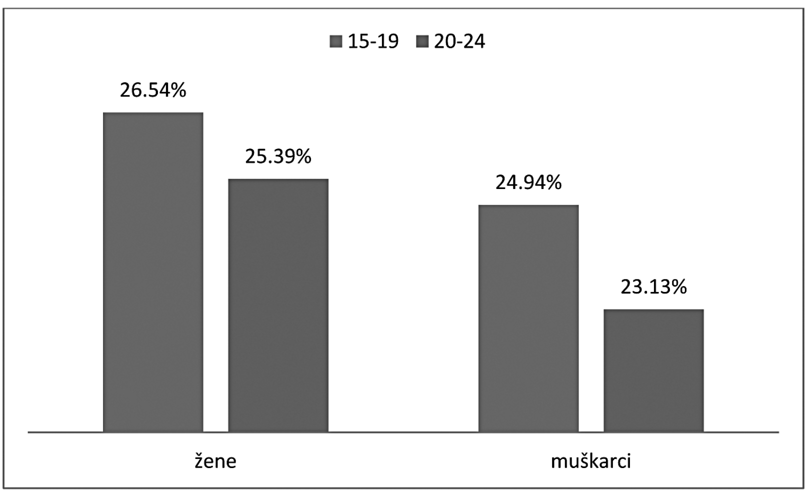
Izvor: *Žene i muškarci u Srbiji*, (2008). Beograd: Republički zavod za statistiku, str. 32.

Podatke iz tabele možemo grafički predstaviti pomoću linijskog dijagrama (grafikon 1.7) i histograma sa podeljenim stupcima (grafikon 1.8).

Grafikon 1.7 *Nepismeno stanovništvo staro 15–24 godine, prema polu, 2002. godine*



Grafikon 1.8 *Nepismeno stanovništvo staro 15–24 godine, prema polu, 2002. godine*



Za prikazivanje strukture stanovništva prema starosti i polu razvijen je još jedan grafički prikaz. Veoma je korišćen u demografskoj analizi i pomoću njega su čak formulisani tipovi starosnih struktura stanovništva. Naziva se piramida starosti (*Population Pyramid*).

U tabeli 1.10 prikazana je raspodela stanovništva Republike Srbije prema starosti i polu, prema popisu iz 2002. godine.

Tabela 1.10 *Stanovništvo Republike Srbije prema starosti i polu, po popisu 2002.*[10]

Godine starosti	Pol		Ukupno
	Muško	Žensko	
0–4	175.753	166.591	342.344
5–9	203.059	191.537	394.596
10–14	225.451	214.379	439.830
15–19	253.353	242.298	495.651
20–24	261.232	251.197	512.429
25–29	253.636	250.930	504.566
30–34	237.553	238.894	476.447
35–39	240.552	245.457	486.009
40–44	263.071	268.757	531.828
45–49	309.769	311.784	621.553
50–54	281.306	290.047	571.353
55–59	186.571	202.614	389.185
60–64	207.676	236.108	443.784
65–69	209.580	250.826	460.406
70–74	167.178	220.106	387.284
75–79	94.552	152.786	247.338

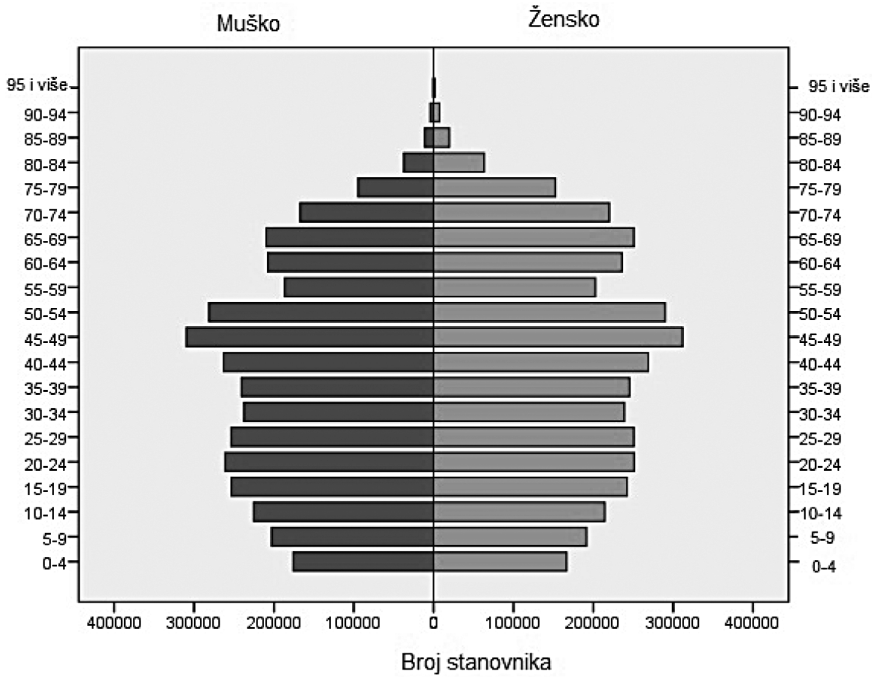
[10] U tabeli je prikazana raspodela stanovništva prema starosti i polu bez kategorije „nepoznato”.

80–84	37317	63.361	100.678
85–89	11.092	19.842	30.934
90–94	4.204	7.542	11.746
95 i više	707	1.412	2.119
Ukupno	3.645.930	3.852.071	7.498.001

Izvor: *Opštine i regioni u Republici Srbiji 2011* (2011). Beograd: Republički zavod za statistiku, str. 38–39.

Piramida starosti se crta na taj način što se sa leve strane dijagrama unose frekvencije starosnih intervala muškog, a sa desne ženskog stanovništva (grafikon 1.9).

Grafikon 1.9 *Piramida starosti stanovništva Republike Srbije prema popisu iz 2002. godine*



ZADACI[11]

1. Kojoj mernoj skali pripadaju merenja sledećih promenljivih.
   Pol: muški
   ženski

   Starost: 15–24
   25–34
   35–44
   45–54
   55–64
   65 i više

   Završena škola ispitanika: nepotpuna osnovna škola
   osnovna škola
   srednja škola
   viša škola
   fakultet
   magistratura (master), doktorat.

2. Prikazane su cene (u dinarima) 120 proizvoda u jednoj trgovini.

100	600	100	500	100	150	400	170	2.000	100	400	300
1.500	100	1.300	500	100	200	250	400	500	800	700	1.400
220	510	180	800	250	100	1.500	380	2.600	1.000	800	250
500	250	1.000	1.500	250	500	700	100	100	100	1.500	200
100	500	600	100	250	150	1.000	500	1.600	2.000	350	100
200	200	100	100	150	500	100	2.000	150	1500	100	200
800	1.100	400	700	300	200	2.400	100	1.500	600	200	200
200	400	300	300	500	200	600	500	800	100	200	300
300	500	800	200	300	300	800	1.000	1.500	1.800	200	250
100	200	2.000	100	200	100	260	500	500	150	1.000	1.250

   a) Prikažite navedene podatke tačkastim dijagramom.
   b) Formirajte distribuciju frekvencija s intervalima koristeći Stardžesovo pravilo.
   c) Prikažite u tabeli distribuciju frekvencija.

[11] Zadaci se mogu preuzeti sa sajta Filozofskog fakulteta u Novom Sadu: http://ff.uns.ac.rs/studenti/oglasne%20table/studenti_oglasne_table_sociologija.html

d) Izračunajte relativne frekvencije i frekvencije kumulativnog niza.
e) Prikažite podatke dijagramom stabljika-list.

3. Tabela 1 pokazuje raspodelu frekvencija za starosnu strukturu 50 zaposlenih u jednom preduzeću.

Tabela 1 *Starosna struktura zaposlenih*

Starost	Broj zaposlenih
18–30	12
31–43	19
44–56	14
57–69	5
Ukupno	50

a) Odredite sredine grupnih intervala.
b) Koliko iznosi širina grupnih intervala?
c) Prikažite tabelarno raspodelu relativnih frekvencija.
d) Koliki je procenat zaposlenih mlađih od 43 godina?

4. Popisom stanovništva iz 2002. godine praćena su, između ostalog, domaćinstva u Republici Srbiji prema broju članova. Rezultati ovog istraživanja prikazani su u tabeli 2.

Tabela 2 *Domaćinstva, prema broju članova, po popisu 2002.*

Broj članova domaćinstva	Republika Srbija – po regionima				Ukupno
	Beogradski region	Region Vojvodine	Region Šumadije i Zapadne Srbije	Region Južne i Istočne Srbije	
1	129.836	149.867	122.234	102.838	504.775
2	141.306	180.858	159.806	143.331	625.301
3	117.628	139.843	121.814	100.896	480.181
4	118.975	153.886	147.612	115.490	535.963

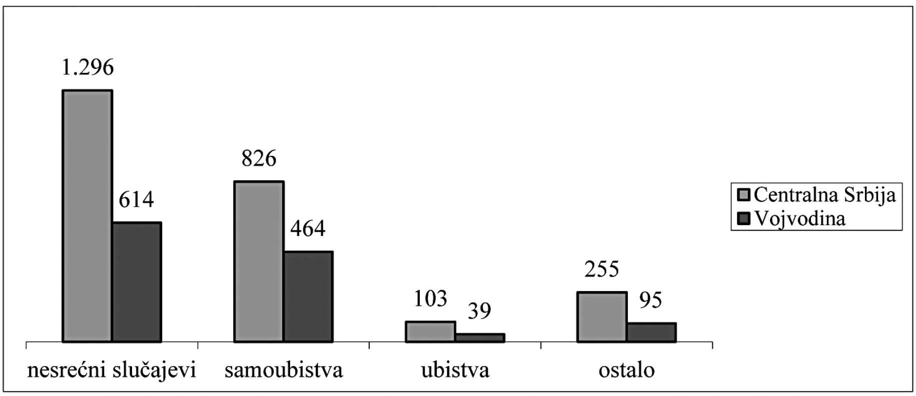
5	37.433	52.766	65.007	50.773	205.979
6	15.424	22.779	40.612	32.874	111.689
7	4.415	6.591	13.951	11.860	36.817
8 i više	2.308	3.367	7.898	6.912	20.485
Ukupno	567.325	709.957	678.934	564.974	2.521.190

Izvor: *Statistički godišnjak Republike Srbije 2011* (2011). Beograd: Republički zavod za statistiku, str. 37.

a) Objasnite kojoj vrsti pripadaju promenljive „Broj članova domaćinstva" i „Republika Srbija – po regionima".
b) Date podatke prikažite pomoću relativnih frekvencija i procenata.
c) Izračunate procente prikažite pomoću linijskog dijagrama.

5. Na grafikonu 1 prikazan je broj nasilnih smrti po oblastima Centralna Srbija i Vojvodina za 2008. godinu.

Grafikon 1 *Nasilne smrti u centralnoj Srbiji i Vojvodini, 2008. godina*



Izvor: *Demografska statistika 2008* (2009). Beograd: Republički zavod za statistiku, str. 243–244

a) Kojoj vrsti mernih skala pripada promenljiva „Vrste nasilne smrti"?
b) Podatke iz histograma prikažite tabelarno. Na osnovu apsolutnih frekvencija izračunajte strukture i kumulativni niz struktura.
c) Paretovom analizom dokažite pravilo „80–20", posebno za podatke za centralnu Srbiju, a posebno za Vojvodinu.

MERE CENTRIRANOSTI

Potpune mere centriranosti

Položajne mere centriranosti

2 MERE CENTRIRANOSTI

Centriranost je jedna od najznačajnijih i najčešće korišćenih pokazatelja numeričkih karakteristika serije podataka. Koristi se gotovo u svim oblastima statističke analize, opisuje čitav statistički skup i omogućuje poređenje različitih skupova. Služi za opisivanje strukture skupa, praćenje njegove dinamike, ali i kao osnova za proučavanje varijacija i oblika raspoređivanja. Centrirana vrednost je pokazatelj centralne tendencije i uslovljena je varijabilnošću vrednosti obeležja skupa.

Ukoliko grupu pokazatelja centriranih (srednjih) ocena posmatramo u odnosu na način njihovog određivanja, možemo govoriti o potpunim (ili izračunatim) centriranim vrednostima, jer se izračunavaju na osnovu svih vrednosti promenljive posmatranog skupa. S druge strane, imamo srednje ocene koje se određuju na osnovu položaja koji zauzimaju u seriji i njih nazivamo položajnim (pozicionim) srednjim ocenama. U grupu potpunih mera centriranosti ubrajamo aritmetičku i geometrijsku sredinu (kao i harmonijsku, kvadratnu i kubnu), dok se u grupi položajnih mera centriranosti nalaze modus, medijana i kvartili.

Izbor mere centriranosti kojom ćemo opisivati seriju zavisi od vrste promenljive. Kod numeričkog niza možemo upotrebljavati sve gore pomenute potpune i položajne srednje ocene, dok se nominalna promenljiva može, pod određenim uslovima, opisivati modusom, a rangirana medusom, medijanom i kvartilima.

2.1 POTPUNE MERE CENTRIRANOSTI

2.1.1 Aritmetička sredina

Aritmetička sredina je najvažnija i najčešće korišćena mera centriranosti. Spada u grupu potpunih ili izračunatih mera. Određuje se tako što se saberu vrednosti numeričke promenljive i podele sa ukupnim brojem elemenata osnovnog skupa ili uzorka. Aritmetička sredina se izražava u meri u kojoj je iskazana i statistička serija iz koje se izračunava. Uobičajeno je da se obeležava grčkim slovom *mi* (μ) ukoliko se govori o aritmetičkoj sredini osnovnog skupa, ili sa iks nadvučenim – *iks barom* ($\bar{x}$), ukoliko je izračunavamo za uzorak.

Aritmetička sredina se iz negrupisanih podataka, izračunava na osnovu obrasca

$$\bar{x} = \frac{1}{n} \Sigma_{i=1}^{n} x_i, \qquad [2.1]$$

dok se za distribuciju frekvencija koristi sledeći izraz:

$$\bar{x} = \frac{\Sigma_{i=1}^{k} f_i \, x_i}{\Sigma_{i=1}^{k} f_i} \qquad [2.1a]$$

Ponderi mogu biti apsolutne frekvencije, relativne frekvencije ili na osnovu njih izračunati procenti. Kako su relativne frekvencije upravno proporcionalne apsolutnim frekvencijama, izraz za ponderisanu aritmetičku sredinu distribucije sa proporcijama glasi:

$$\bar{x} = \Sigma_{i=1}^{k} fr \cdot x_i \qquad [2.2]$$

Postupak izračunavanja aritmetičke sredine ilustrovaćemo na primeru ukupnog broja zaposlenih u Republici Srbiji za period od tri godine (primer 1).

Primer 1

U narednoj tabeli (tabela 2.1) prikazan je ukupan broj zaposlenih u Republici Srbiji, za 2009, 2010. i 2011. godinu.

Tabela 2.1 *Broj zaposlenih u Republici Srbiji*

godina	Ukupno
2009.	1.889.085
2010.	1.795.775
2011.	1.746.138

Izvor: Republički zavod za statistiku, sajt www.stat.gov.rs

Kako se radi o negrupisanim numeričkim podacima i o osnovnom skupu (jer je prikazan ukupan broj zaposlenih u Republici Srbiji), za izračunavanje aritmetičke sredine koristićemo obrazac [2.1], a za njeno obeležavanje koristićemo grčko slovo μ.

$$\mu = \frac{1.889.085 + 1.795.775 + 1.746.138}{3} = 1.810.332{,}667 \text{ zaposlenih}$$

Prosečan broj zaposlenih u Republici Srbiji za period od 2009. do 2011. godine iznosio je 1.810.332, 667 zaposlenih.

Raspodelu domaćinstava na teritoriji Vojvodine prema podacima popisa iz 2002. godine (primer 2) poslužiće nam za objašnjenje postupka izračunavanja ponderisane aritmetičke sredine.

Primer 2

U tabeli 2.2 prikazana je raspodela domaćinstava na teritoriji Vojvodine prema broju članova u 2002. godine.

Tabela 2.2 *Domaćinstva Vojvodine prema broju članova, 2002. godine*

Broj članova domaćinstva	f_i
1	149.867
2	180.858
3	139.843
4	153.886
5	52.766
6	22.779
7	6.591
8	3.367
Ukupno	709.957

Izvor: Republički zavod za statistiku, sajt www.stat.gov.rs

Za izračunavanje aritmetičke sredine iz ove serije podataka, napravićemo radnu tabelu (tabela 2.2.1) koja će nam pomoći da to na brži i precizniji način uradimo.

Tabela 2.2.1 *Radna tabela za izračunavanje aritmetičke sredine*

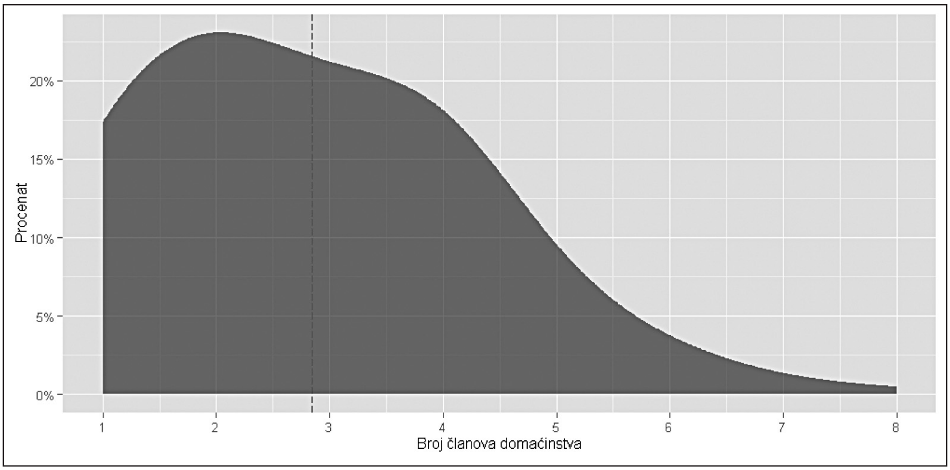
x_i	f_i	$x_i f_i$	fr	$fr \cdot x_i$
1	149.867	149.867	0,21	0,21
2	180.858	361.716	0,25	0,50
3	139.843	419.529	0,20	0,60
4	153.886	615.544	0,22	0,88
5	52.766	263.830	0,07	0,35
6	22.779	136.674	0,03	0,18
7	6.591	46.137	0,01	0,07
8	3.367	26.936	0,01	0,08
Ukupno	709.957	2.020.233	1,00	2,87

Iz obrasca [2.1a] vidimo da su nam za izračunavanje aritmetičke sredine potrebne sume proizvoda obeležja i njihovih frekvencija (treća kolona), kao i ukupan broj frekvencija (druga kolona). Kada ta dva zbira podelimo dobićemo aritmetičku sredinu za ovu statističku seriju.

$$\mu = \frac{2020233}{709957} = 2{,}85 \approx 3 \text{ } \textit{člana domaćinstva}$$

Dakle, prosečan broj članova domačinstva u Vojvodini za 2002. godinu iznosio je približno 3 člana po domaćinstvu. Raspodela frekvencija iz tabele 2.2 prikazana je u narednom grafikonu (grafikon 2.1), gde je isprekidanom linijom obeležena vrednost artitmetičke sredine.

Grafikon 2.1 *Domaćinstva Vojvodine prema broju članova, 2002. godine*



Podatke iz tabele raspodele domaćinstava prema broju članova iskoristićemo i za objašnjenje izračunavanja aritmetičke sredine na osnovu relativnih frekvencija i procenata. U četvrtoj koloni radne tabele 2.2.1 date su relativne frekvencije, dok su u petoj koloni izračunati proizvodi obeležja serije i njihovih relativnih frekvencija. Suma pete kolone predstavlja aritmetičku sredinu na osnovu obrasca [2.2]:

$$\mu = 2{,}87 \approx 3 \text{ } \textit{člana domaćinstva}$$

Razlika između vrednosti 2,85 i 2,87 pojavila se usled zaokruživanja decimalnih brojeva u petoj koloni.

U mnogim slučajevima numerička statistička serija se može prikazati i pomoću intervala. Njihovo određivanje i grupisanje frekvencija u odgovarajuće intervale objasnili smo u uvodnom delu. Na sledećem primeru prikazaćemo izračunavanje aritmetičke sredine pomoću intervalne numeričke serije. Za izračunavanje proseka u intervalnoj seriji potrebno je da odredimo sredine grupnih intervala (x'_i) i da nakon toga svaku sredinu pomnožimo sa njenom frekvencijom (f_i). Aritmetička sredina će se dobiti kada sumu proizvoda ($\Sigma_{i=1}^{k} x'_i \cdot f_i$) podelimo sa ukupnim brojem frekvencija ($\Sigma_{i=1}^{k} f_i$), tj. pomoću sledeće formule:

$$\mu = \frac{\Sigma_{i=1}^{k} x'_i \cdot f_i}{\Sigma_{i=1}^{k} f_i} \qquad [2.3]$$

Primer 3

U narednoj tabeli (tabela 2.3) data je raspodela nepismenog stanovništva prema starosti u radnom kontigentu stanovništva Republike Srbije, po rezultatima popisa iz 2002. godine.

Tabela 2.3 *Nepismeno stanovništvo prema starosti u radnom kontigentu na teritoriji Republike Srbije, 2002. godine*

Starost (x_i)	Frekvencije (f_i)	Sredine intervala (x'_i)	$x'_i f_i$
15–19	3.338	17,5	58.415
20–24	3.146	22,5	70.785
25–29	2.822	27,5	77.605
30–34	2.865	32,5	93.112,5
35–39	3.287	37,5	123.262,5
40–44	4.282	42,5	181.985
45–49	5.481	47,5	260.347,5

50–54	5.719	52,5	300.247,5
55–59	7.103	57,5	408.422,5
60–64	15.231	62,5	951.937,5
Ukupno	53.274		2.526.120

Izvor: *Popis stanovništva, domaćinstva i stanova u 2002. Stanovništvo. Školska sprema i pismenost*. (2003). Beograd: Republički zavod za statistiku, str. 16–17

Prosečna starost nepismenog stanovništva u radnom kontigentu odrediće se pomoću obrasca [2.3] i iznosi:

$$\mu = \frac{2526120}{53274} = 47{,}42 \textit{ godina}$$

Na istom primeru objasnićemo i metod izračunavanja aritmetičke sredine pomoću kumulacije frekvencija (treća kolona u tabeli 2.3.1).

Tabela 2.3.1 *Radna tabela za izračunavanje aritmetičke sredine metodom kumulacije frekvencija*

Sredine intervala (x_i')	Frekvencije (f_i)	Kumulacija iznad
17,5	3.338	53.274
22,5	3.146	49.936
27,5	2.822	46.790
32,5	2.865	43.968
37,5	3.287	41.103
42,5	4.282	37.816
47,5	5.481	33.534
52,5	5.719	28.053
57,5	7.103	22.334
62,5	15.231	15.231
Ukupno	53.274	372.039

Važno je zapamtiti da se kod ovog metoda koristi isključivo kumulacija iznad i njena suma, a ne suma kumulacija ispod.

Ako sa $\overline{K}$ označimo kumulativnu sredinu i izračunamo je tako što sumu kumulacija podelimo sa sumom frekvencija, tada će ona iznositi

$$\overline{K} = \frac{372.039}{53.274} = 6{,}98$$

Kumulativna sredina se unosi u formulu za izračunavanje aritmetičke sredine na osnovu kumulacija koja glasi:

$$\mu = x'_0 + \overline{K} \cdot i \qquad [2.4]$$

gde je: x'_0 vrednost sredine intervala koja prethodi sredini prvog grupnog intervala i ona u ovom slučaju iznosi 12,5 (17,5 – 5 = 12,5), dok je i širina grupnog intervala. Tako dobijamo aritmetičku sredinu metodom kumulacija frekvencija i ona iznosi:

$$\mu = 12{,}5 + 6{,}98 \cdot 5 = 47{,}4 \textit{ godina}$$

U slučajevima kada intervalna serija ima veći broj grupa i kada su njihove sredine intervala izuzetno mali ili veliki brojevi, upotrebljavaju se skraćeni načini izračunavanja aritmetičke sredine: metod linearne transformacije obeležja i metod provizorne (radne) sredine.

Za ilustraciju ovih metoda uzećemo istu tabelu raspodele frekvencija (tabela 2.3). Metod linearne transformacije obeležja zahteva uvođenje novog obeležja Z_i.

$$Z_i = \frac{x'_i - K}{i}$$

gde je x'_i sredina grupnog intervala, K predstavlja jednu od grupnih sredina od koje se počinje (obično je to sredina grupnog intervala koji ima najveću frekvenciju), dok je i i u ovom slučaju širina grupnog intervala. Sredina transformisanih obeležja dobija se pomoću formule:

$$\overline{Z} = \frac{\sum_{i=1}^{k} f_i Z_i}{N} \qquad \text{ili} \qquad \overline{Z} = \frac{\mu - K}{i}$$

Dakle, aritmetička sredina metodom linearne transformacije obeležja dobija se na osnovu formule:

$$\mu = \overline{Z}_i + K \qquad [2.5]$$

Tabela 2.3.2 *Radna tabela za izračunavanje aritmetičke sredine metodom linearne transformacije obeležja*

Sredine intervala (x'_i)	Frekvencije (f_i)	Z	Zf
17,5	3.338	–9	–30.042
22,5	3.146	–8	–25.168
27,5	2.822	–7	–19.754
32,5	2.865	–6	–17.190
37,5	3.287	–5	–16.435
42,5	4.282	–4	–17.128
47,5	5481	–3	–16.443
52,5	5.719	–2	–11.438
57,5	7.103	–1	–7.103
62,5	15.231	0	0
Ukupno	53.274		–160.701

Na osnovu radne tabele 2.3.2 vidimo da je sredina grupnog intervala koji ima najveću frekvenciju 62,5 i ona će u našem primeru biti vrednost K. Zatim ćemo u treću kolonu radne tabele uneti odnos razlike grupnih sredina intervala i broja K sa širinom grupnog intervala. Suma ove kolone podeljena sa sumom frekvencija daje sredinu transformisanog obeležja ($\bar{Z}$)

$$\bar{Z} = \frac{-160.701}{53.274} = -3{,}02,$$

a ono nam je potrebno radi izračunavanja aritmetičke sredine na osnovu obrasca [2.5].

$$\mu = -3{,}02 \cdot 5 + 62{,}5 = 47{,}4 \textit{ godina}$$

I u ovom slučaju vrednost aritmetičke sredine je 47,4 godina jer je ovaj metod, kao i metod kumulacije frekvencija računat za iste vrednosti.

U metodu provizorne sredine kreće se od nekog proizvoljno izabranog broja K, zatim se izračunavaju odstupanja sredina grupnih intervala od broja K i dobijaju nove vrednosti (tabela 2.3.3)

$$y_i = x_i - K.$$

Aritmetička sredina ovih novih vrednosti ($\bar{y}$) dobija se pomoću formule:

$$\bar{y} = \frac{\Sigma_{i=1}^{k}(x'_i - K) f_i}{\Sigma_{i=1}^{k} f_i} \qquad \text{ili} \qquad \bar{y} = \bar{x} - K$$

Iz ovih obrazaca dobijamo aritmetičku sredinu:

$$\mu = \bar{y} + K$$

Tabela 2.3.3 *Radna tabela za izračunavanje aritmetičke sredine metodom provizorne sredine*

(x'_i)	(f_i)	$x'_i - K$	$(x'_i - K)\,f_i$
17,5	3.338	–45	–150.210
22,5	3.146	–40	–125.840
27,5	2.822	–35	–98.770
32,5	2.865	–30	–85.950
37,5	3.287	–25	–82.175
42,5	4.282	–20	–85.640
47,5	5.481	–15	–82.215
52,5	5.719	–10	–57.190
57,5	7.103	–5	–35.515
62,5	15.231	0	0
Ukupno	53.274		–803.505

$$\bar{y} = \frac{-803.505}{53.274} = -15{,}08$$

$$\mu = -15{,}08 + 62{,}5 = 47{,}42 \text{ godina}$$

Kako potrebu merenja svih mera centralne tendencije uslovljava varijabilitet vrednosti promenljive jednog statističkog skupa, ona ima veliki uticaj i na same vrednosti mera centriranosti. Primetićemo da, ako je serija sastavljena od istih numeričkih obeležja, i sama aritmetička sredina će biti isti taj broj. Međutim, ako se u jednoj seriji podataka pojave vrednosti koje puno odstupaju od proseka ostalih vrednosti, one će umnogome uticati na njenu vrednost. Aritmetička sredina je osetljiva na ekstremne vrednosti (nestandardne opservacije). Zamislite ovakav primer. U jednom odeljenju 7. razreda neke osnovne škole razredni starešina je slučajnim putem izabrao 5 učenika i pogledao njihove ocene iz matematike. Zamislite takođe da su svi izabrani đaci dobili petice iz matematike. Prosečna ocena ovih đaka je 5. Međutim, da se među njima našao đak koji je dobio 1, prosečna ocena bi bila nešto drugačija i iznosila bi 4,20. Da bi se ublažio uticaj ekstremnih podataka na vrednost aritmetičke sredine, možemo se poslužiti „odsečenom aritmetičkom sredinom”. Radi se o tome da kod serija sa velikim varijabilitetom možete izuzeti ekstremne vrednosti i na tako „skraćenoj” statističkoj seriji izračunati prosek. U jednoj sređenoj seriji podataka (gde su podaci poređani u rastući niz), ekstremne vrednosti će se uvek pojaviti na samom početku (kao ekstremno male vrednosti) ili na kraju serije (kao ekstremno velike). Da biste korektno primenili odsečenu aritmetičku sredinu, ovakvu seriju morate umanjiti („odseći”) za isti procentualni deo (najčešće od 5 do 25% podataka), kako s početka, tako i s kraja raspodele.

Primer 4

Sledeći podaci odnose se na godine starosti 10 zaposlenih u jednom preduzeću: 47, 53, 38, 26, 39, 49, 18, 65, 31, 33. Da bismo izračunali 20% odsečenu aritmetičku sredinu, podatke najpre moramo poređati u rastući niz:

18]   26   31   33   38   39   47   49   53   [65.

Kako je 20% od 10 podataka 2, to znači da seriju moramo umanjiti za 1 podatak s početka i 1 s kraja. Tako će naša serija imati ukupno 8 podataka, a kako se radi o negrupisanom nizu za izračunavanje aritmetičke sredine upotrebićemo obrazac [2.1]:

$$\bar{x} = \frac{26 + 31 + 33 + 38 + 39 + 47 + 49 + 53}{8} = \frac{316}{8} = 39{,}5 \text{ godina}$$

U praksi se često dešava da smo jednu pojavu merili nekoliko puta (npr. uspeh učenika) i izračunali nekoliko aritmetičkih sredina (prosečan uspeh učenika). Ukoliko želimo da izračunamo aritmetičku sredinu ovih aritmetičkih sredina onda njihove vrednosti možemo da saberemo i podelimo brojem aritmetičkih sredina, ali jedino pod uslovom da su one izračunate iz jednakog broja rezultata. Dakle, pod uslovom da se radi o istovrsnim merenjima i da važi da je $n_1 = n_2 = n_3 = \cdots = n_n$, kombinovanu aritmetičku sredinu ($\bar{\bar{x}}$) možemo dobiti pomoću formule:

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \cdots + \bar{x}_n}{n_{\bar{x}}} \qquad [2.6]$$

gde je $n_{\bar{x}}$ broj aritmetičkih sredina.

Međutim, ako pojedinačne aritmetičke sredine nismo dobili na osnovu jednakog broja elemenata u uzorku (ili u osnovnom skupu) tada njihov zbir ne možemo podeliti njihovim brojem, već moramo koristiti sledeći obrazac:

$$\bar{\bar{x}} = \frac{n_1 \cdot \bar{x}_1 + n_2 \cdot \bar{x}_2 + \cdots + n_n \cdot \bar{x}_n}{n_1 + n_2 + n_n} \qquad [2.6a]$$

gde n_1 predstavlja veličinu prvog uzorka, a n_2 veličinu drugog uzorka.

Primer 5

Na osnovu uzorka od 10 različitih udžbenika iz statistike izračunata je prosečna cena od 2.569 dinara, a na osnovu uzorka 8 različitih udžbenika iz matematike dobijena je prosečna cena od 1.568 dinara. Na osnovu kombinovane aritmetičke sredine možemo izračunati prosečnu cenu ovih knjiga.

$$\bar{\bar{x}} = \frac{10 \cdot 2.569 + 8 \cdot 1.568}{10 + 8} = \frac{25.690 + 12.544}{18} = \frac{38.234}{18} = 2.124{,}11 \textit{ dinara}$$

Na osnovu svega što smo do sada rekli, možemo istaći sledeće osobine aritmetičke sredine:

1. veća je od najmanje i manja od najveće vrednosti posmatranog obeležja ($x_1 \leq \mu \leq x_k$);
2. izjednačuje se sa vrednostima obeležja, kada su one međusobno jednake ($x_1 = x_2 = \cdots = x_k = \mu$);

3. zbir odstupanja aritmetičke sredine od pojedinačnih vrednosti obeležja jednak je nuli $\Sigma_{i=1}^{N} (x_i - \mu) = 0$; $\Sigma_{i=1}^{N} f_i (x_i - \mu) = 0$;

4. zbir kvadrata odstupanja aritmetičke sredine od pojedinačnih vrednosti obeležja manji je od zbira kvadrata odstupanja bilo koje vrednosti obeležja (pa i drugih srednjih vrednosti ako nisu jednake aritmetičkoj sredini) od ostalih vrednosti obeležja: $\Sigma_{i=1}^{N} (x_i - \bar{x})^2 < \Sigma_{i=1}^{N} (x_i - a)^2$;

5. ako svakoj vrednosti x dodamo ili oduzmemo jednu *konstantu* (c), tada će nova aritmetička sredina biti uvećana ili umanjena za tu konstantu; ako svaku vrednost x pomnožimo nekom konstantom tada je nova aritmetička sredina proizvod aritmetičke sredine i konstante;

6. uspešno reprezentuje one rasporede kod kojih najveću frekvenciju imaju one vrednosti obeležja koje su najbliže aritmetičkoj sredini. Ona se ne može izračunati za rasporede sa otvorenim grupnim intervalima ako nisu poznate sredine tih intervala;

7. iskazuje se u jedinici mere u kojoj je izražena jedinica u seriji.

Navedene osobine ilustrovaćemo pomoću tabele 2.4.

Tabela 2.4 *Odstupanja od aritmetičke sredine*

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$x_i - a$	$(x_i - a)^2$	$x_i + c$	$x_i - c$	$x_i \cdot c$
5	–5	25	–3	9	10	0	25
8	–2	4	0	0	13	3	40
10	0	0	2	4	15	5	50
12	2	4	4	16	17	7	60
15	5	25	7	49	20	10	75
ukupno 50	0	58		78	75	25	250

Aritmetička sredina za ovaj niz podataka iznosi 10, što dokazuje prvu tvrdnju jer je $5 < 10 < 15$. Na osnovu druge kolone pokazali smo da je suma odstupanja svakog elementa serije od aritmetičke sredine jednaka 0.

Četvrta osobina aritmetičke sredine govori da je zbir kvadrata odstupanja pojedinačnih vrednosti od aritmetičke sredine, u našem slučaju to je

58 (zbir vrednosti u trećoj koloni), manji od zbira odstupanja pojedinačnih vrednosti od bilo koje vrednosti obeležja (zbir vrednosti u petoj koloni). Njega smo dobili tako što smo odabrali vrednost 8 (a) i svaki podatak iz tabele umanjili za njegovu vrednost. Kako je zbir kvadrata ovih odstupanja 78, onda važi da je 58 < 78.

Ako je prema 5. osobini $\bar{x}' = \bar{x} + c$, a broj 5 uzmemo kao konstantu, onda $\bar{x}' = \frac{75}{5}$ 15, odnosno 15 = 10 + 5. Ili, $\bar{x}' = \bar{x} - c$ onda $\bar{x}' = \frac{25}{5} = 5$ odnosno 5 = 10 – 5. Isto važi i za $\bar{x}' = \bar{x} \cdot c$ jer je $\bar{x}' = \frac{250}{5} = 50$ odnosno $50 = 10 \cdot 5$.

2.1.2 Geometrijska sredina

Geometrijska sredina pripada grupi potpunih (izračunatih) mera centriranosti što znači da su svi elementi uzorka uključeni u njeno izračunavanje. Izražava se u jedinicama mere u kojima je iskazana statistička serija. Iako se upotrebljava mnogo ređe od aritmetičke sredine, njenu primenu nalazimo u dinamičkoj analizi vremenskih serija kada je važnije utvrditi razlike u odnosima nego u apsolutnim vrednostima. U demografskim istraživanjima, istraživači je najčešće koriste da utvrde stopu promene broja stanovnika u godinama između dva popisa, u ekonomskim, za izračunavanje stope rasta na osnovu lančanih indeksa, itd. (vidi: Žižić, Lovrić, Pavličić, 1992: 45; Savić, 2005: 39.)

Za n vrednosti neke numeričke promenljive X geometrijska sredina je *n-ti* koren iz proizvoda njenih vrednosti. Za negrupisane vrednosti jedne numeričke promenljive, upotrebićemo sledeći izraz za izračunavanje geometrijske sredine:

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \cdots \cdot x_i \cdot \cdots \cdot x_n}, \; x_i > 0, \forall i \qquad [2.7]$$

Kraći zapis je

$$G = \sqrt[n]{\prod_{i=1}^{n} x_i},$$

Geometrijsku sredinu iz negrupisanih podataka možemo dobiti logaritmovanjem vrednosti podataka i tada koristimo sledeći izraz

$$log\ G = \frac{1}{n}\,\Sigma_{i=1}^{n}\ log\ x_i \qquad [2.7a]$$

da bi se potom antilogaritmom dobila prava vrednost geometrijske sredine. Logaritmovanje podataka se često koristi ukoliko se radi o visokim vrednostima numeričke serije, te se time olakšava izračunavanje.

Za grupisane podatke, geometrijska sredina data je izrazom:

$$G = \sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdot x_i^{f_i} \cdot \cdots \cdot x_k^{f_k}} \qquad [2.8]$$

ili u skraćenom zapisu

$$G = \sqrt[n]{\prod_{i=1}^{k} x_i^{f_i}}, \qquad \text{gde je} \qquad n = \Sigma_{i=1}^{k} f_i$$

ili logaritmovanjem podataka:

$$log\ G = \frac{1}{n}\prod_{i=1}^{k} f_i\ log\ x_i \qquad [2.8a]$$

čime se dobija logaritmovana geometrijska sredinu koju, isto kao i u prethodnom slučaju, treba antilogaritmovati.

Zbog toga što se geometrijska sredina dobija iz proizvoda vrednosti promenljive, nju ne možemo izračunati ukoliko je neka vrednost u seriji jednaka 0.

Primer 6

Slučajnim uzorkom izabrano je 7 studenata koji su polagali pismeni ispit iz Statistike i zabeleženi su njihovi bodovi.

8, 14, 24, 25, 32, 33, 45.

Za izračunavanje geometrijske sredine upotrebićemo izraz [2.7], jer se radi o negrupisanoj seriji podataka.

$$G = \sqrt[7]{8 \cdot 14 \cdot 24 \cdot 25 \cdot 32 \cdot 33 \cdot 45} = 22{,}79$$

Ukoliko bismo na ovu seriju podataka primenili skraćeni izraz [2.7a], onda bismo se za njeno izračunvanje poslužili tabelom 2.4.

Tabela 2.4 *Radna tabela za izračunavanje geometrijske sredine*

x_i	log x_i
8	0,90
14	1,15
24	1,38
25	1,40
32	1,51
33	1,52
45	1,65
ukupno	9,51

$$log\ G = \frac{1}{7}\sum_{i=1}^{n} 9{,}51 = 1{,}36$$

Ovakvu logaritmovanu geometrijsku sredinu je potrebno antilogaritmovati i na taj način dobiti pravu vrednost geometrijske sredine.

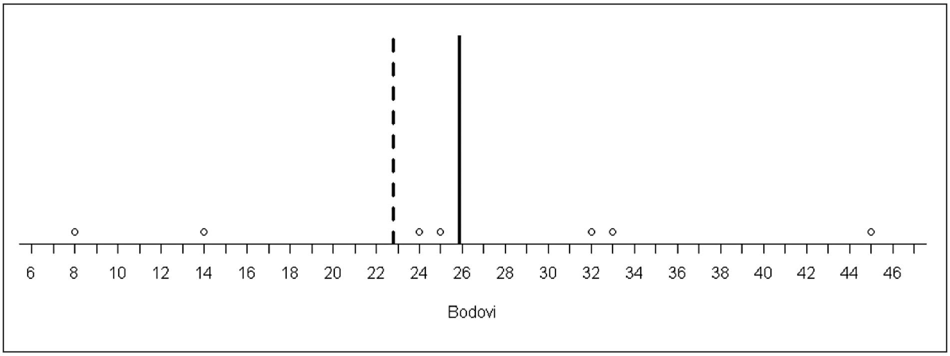
$$G = 10^{log\ n} = 10^{1{,}36} = 22{,}91$$

Za istu seriju podataka aritmetičku sredinu (dobijena izrazom [2.1]) iznosila bi:

$$\bar{x} = \frac{1}{7}\sum_{i=1}^{n} 8 + 14 + 24 + 25 + 32 + 33 + 45 = \frac{181}{7} = 25{,}86$$

Primetićete da je geometrijska sredina manja vrednost od aritmetičke sredine računate za istu seriju podataka, što pokazuje i grafikon 2.2 (punom linijom označen je položaj aritmetičke sredine, isprekidanom položaj geometrijske, dok su tačkama označeni podaci o prikupljenim bodovima. Pošto se radi o meri centriranosti, geometrijska sredina je, takođe, veća od najmanje i manja od najveće vrednosti u posmatranoj seriji, a u slučaju da su sve vrednosti jednake i ona uzima istu tu vrednost.

Grafikon 2.2 *Aritmetička i geometrijska sredina ispitnih bodova*



Primer 7

U narednoj tabeli prikazana je distribucija anketiranih (f_i) prema broju članova domaćinstva (x_i):

Tabela 2.5 Distribucija anketiranih prema broju članova domaćinstva

x_i	1	2	3	4	5	6
f_i	3	12	15	10	5	5

Za distribuciju frekvencija koristimo geometrijsku sredinu datu izrazom [2.8]:

$$G = \sqrt[50]{1^3 \cdot 2^{12} \cdot 3^{15} \cdot 4^{10} \cdot 5^5 \cdot 6^5} = 3{,}04 \text{ članova.}$$

Za primenu izraza [2.8a] potrebno je napraviti sledeću radnu tabelu (2.5.1):

Tabela 2.5.1 *Radna tabela za izračunavanje ponderisane geometrijske sredine*

x_i	$log\ x_i$	f_i	$f_i\ log\ x_i$
1	0	3	0
2	0,30	12	3,61
3	0,48	15	7,16
4	0,60	10	6,02
5	0,70	5	3,49
6	0,78	5	3,89
ukupno		50	24,17

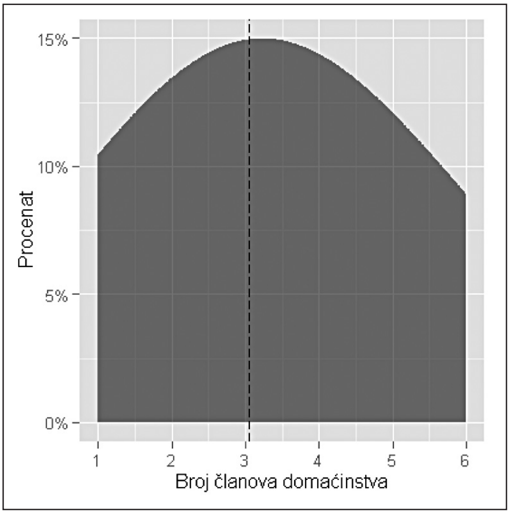
$$log\ G = \frac{24{,}17}{50} = 0{,}48$$

$$G = 10^{0{,}48} = 3{,}02 \text{ članova}$$[12]

Položaj geometrijske sredine prikazan grafikonom 2.2.

[12] Razlika između dobijenih vrednosti geometrijske sredine primenom izraza 2.8 i 2.8a nastala je zaokruživanjem decimalnih brojeva u tabeli 2.5a.

Grafikon 2.2 *Raspodela anketiranih prema broju članova domaćinstva*



Primer 8

Geometrijsku sredinu izračunatu na osnovu intervalne numeričke serije (tabela 2.6) prikazaćemo na primeru nepismenog stanovništva Republike Srbije 2002. godine. U ovom slučaju koristićemo izraz [2.8a].

Tabela 2.6 *Nepismeno stanovništvo prema starosti u radnom kontigentu na teritoriji Republike Srbije, 2002. godine*

Starost (x_i)	Frekvencije (f_i)	Sredine intervala (x'_i)	$log\ x'_i$	$f_i\, x'_i$
15–19	3.338	17,5	1,24	4.149,26
20–24	3.146	22,5	1,35	4.253,97
25–29	2.822	27,5	1,44	4.061,80
30–34	2.865	32,5	1,51	4.331,55

35–39	3.287	37,5	1,57	5.173,85
40–44	4.282	42,5	1,63	6.972,77
45–49	5.481	47,5	1,68	9.189,98
50–54	5.719	52,5	1,72	9.837,60
55–59	7.103	57,5	1,76	12.498,92
60–64	15.231	62,5	1,80	27.353,05
Ukupno	53.274			87.822,69

$$log\ G = \frac{87.822{,}69}{53.274} = 1{,}65$$

$$G = 10^{1{,}65} = 44{,}67 \text{ godina}$$

Aritmetička sredina za ovu seriju podataka iznosila je 47,42 godine, dok je prosečna starost nepismenog stanovništva izračunata na osnovu geometrijske sredine 44,67 godina.

Primer 9

Broj stanovnika u Srbiji, prema popisu iz 2002. godine, iznosio je 7.498.001, a 2011. godine 7.186.862 stanovnika. Naći prosečnu godišnju stopu promene broja stanovnika. Proceniti broj stanovnika 2015. i 2017. godine.

Geometrijska stopa rasta (r_g) na osnovu empirijskih podataka dobija se pomoću izraza:

$$r_g = \left(\sqrt[N-1]{\frac{Y_N}{Y_{N-1}}} - 1\right) \cdot 100,$$

gde Y_{N-1} označava prvu vrednost u vremenskoj seriji, Y_N poslednju vrednost, a N broj podataka. Kako se naš primer zasniva na podacima osnovnog skupa koristimo veliko slovo N (u slučaju podataka dobijenih na osnovu uzorka koristili bismo malo n).

$$r_g = \left(\sqrt[10-1]{\frac{7.186.862}{7.498.001}} - 1\right) \cdot 100 = (0{,}99 - 1) \cdot 100 = -1\%$$

Na osnovu dobijene geometrijske stope rasta možemo videti da se stanovništvo Republike Srbije svake godine u posmatranom periodu u proseku smanjivalo za 1%.

Procenu broja stanovnika u narednom periodu (za 2015. i 2017. godinu) možemo dobiti pomoću sledeće formule:

$$Y_{N+1} = Y_N \cdot \sqrt[N-1]{\frac{Y_N}{Y_{N-1}}}$$

$$Y_{2012} = 7.186.862 \cdot 0{,}99 = 7.114.993{,}38$$

$$Y_{2013} = 7.114.993{,}38 \cdot 0{,}99 = 7.043.843{,}446$$

$$Y_{2014} = 7.043.843{,}446 \cdot 0{,}99 = 6.973.405{,}012$$

$$Y_{2015} = 6.973.405{,}012 \cdot 0{,}99 = 6.903.670{,}962$$

$$Y_{2016} = 6.903.670{,}962 \cdot 0{,}99 = 6.834.634{,}252$$

$$Y_{2017} = 6.834.634{,}252 \cdot 0{,}99 = 6.766.287{,}909$$

Na osnovu procene o kretanju broja stanovnika primenom geometrijske stope rasta vidimo da će broj stanovnika Republike Srbije u 2015. godini iznositi 6.903.671, dok će 2017. Srbija imati 6.766.288 stanovnika. Međutim, potrebno je skrenuti pažnju da su rezultati aproksimacije pomoću geometrijske stope rasta dosta neprecizni i da je demografija za procenu broja stanovnika razvila mnogo sofisticiranije i složenije metode.

2.2 POLOŽAJNE MERE CENTRIRANOSTI

2.2.1 Modus

Modus je srednja vrednost koja pripada grupi položajnih vrednosti. Po definiciji, modus je modalitet promenljive koja ima najveću frekvenciju. Njega možemo određivati u nominalnim i rang promenljivama ili u numeričkim. U statističkoj seriji moguće je pronaći jedan modalitet sa najvećom frekvencijom i tada kažemo da je ona unimodalna. Međutim, serija može imati dva ili više modaliteta sa istim (najvećim) frekvencijama i u tom slučaju se ona opisuje kao bimodalna ili multimodalna. Jedna statistička serija može imati podjednaku učestalost javljanja svih modaliteta i tada modus nije moguće odrediti. Modus nije moguće odrediti i kod negrupisanih podataka.

Primer 10

Kada su podaci grupisani, modus se jednostavno pronalazi. Potrebno je jedino pronaći modalitet koji ima najveću frekvenciju. Prema tabeli 2.7 u kojoj je prikazan raspored domaćinstva Vojvodine prema broju članova, modus 2 člana domaćinstva jer se u seriji pojavljuje 180.858 puta.

Tabela 2.7 *Domaćinstva Vojvodine prema broju članova, 2002. godine*

Broj članova domaćinstva	f_i
1	149.867
2	180.858
3	139.843
4	153.886
5	52.766
6	22.779
7	6.591
8	3.367
Ukupno	709.957

U slučaju intervalnih serija, modus ćemo izračunati na osnovu sledeće formule:

$$M_o = L_1 + \frac{(b-a)}{(b-a)+(b-c)} \cdot i \qquad [2.9]$$

gde je L_1 donja granica modalnog intervala, b najveća frekvencija, a frekvencija iznad nje, c frekvencija iza najveće, dok je i širina grupnog intervala.

Primer 11

U narednoj tabeli (tabela 2.8) prikazan je raspored stanovništva Republike Srbije prema starosti na osnovu popisa iz 2011. godine. Modalni interval u ovoj statističkoj seriji je interval 55–59 godina starosti jer je stanovništvo ove starosne grupe najzastupljenije, tako da se modus mora naći u okviru ovog intervala. Na osnovu formule [2.9] dobijamo

$$M_o = 55 + \frac{(596.279 - 520.344)}{(596.279 - 520.344) + (596.279 - 528.414)} \cdot 5 = 57{,}65 \approx 58 \, godina$$

Tabela 2.8 *Stanovništvo Republike Srbije prema starosti, popis 2011. godine*

Starost	Frekvencije
0–4	328.255
5–9	350.154
10–14	346.869
15–19	401.994
20–24	439.741
25–29	480.286
30–34	496.362

35–39	493.934
40–44	469.928
45–49	483.986
50–54	520.344
55–59	596.279
60–64	528.414
65–69	339.444
70–74	354.142
75–79	298.612
80–84	176.568

Izvor: *Popis stanovništva, domaćinstava i stanova 2011. u Republici Srbiji. Starost i pol. Podaci po naseljima*. (2012). Beograd: Republički zavod za statistiku, str. 26–27

Dakle, na osnovu popisa iz 2011. godine, u Republici Srbiji je najzastupljenije stanovništvo staro 58 godina.

2.2.2 Medijana i kvartili

Medijana je položajna vrednost koja numerički niz uređen po veličini deli na dva jednaka dela. Osim kod numeričkih serija, medijanu možemo koristiti i za opisivanje rang promenljivih. One se takođe mogu urediti prema intenzitetu merenog svojstva i ta osobina sređivanja nam daje mogućnost za primenu medijane. Izračunavanje medijane se u ovom slučaju svodi na pronalaženje vrednosti numeričke ili rang promenljive koja zadovoljava definiciju medijane.

Medijana se, kao i sve druge mere centralne tendencije, izražava u jedinicama mere u kojima je iskazana serija podataka za koju se računa; u svakoj statističkoj seriji (ukoliko je ona numerička ili rang serija) postoji samo jedna vrednost medijane; nalazi se između najmanje i najveće vrednosti niza; na njenu vrednost ne utiču ekstremne vrednosti i zbir apsolutnih odstupanja vrednosti promenljive od medijane je minimalan.

Medijana kod neparnog broja podataka predstavlja vrednost promenljive srednjeg člana uređenog niza.

Primer 12

Ukoliko uzmemo niz od 7 članova sa vrednostima 7 13 32 45 59 77 98, medijana je četvrti član u seriji, vrednost 45, jer upravo ona deli seriju na dva jednaka dela.

Kod grupisanih neparnih podataka medijanu možemo odrediti pomoću izraza:

$$M_e = x_{\frac{n+1}{2}} \qquad [2.10]$$

Primer 13

Podatke o raspodeli domaćinstva Vojvodine prema broju članova u 2002. godini koristili smo za izračunavanje modusa i videli smo da su domaćinstva sa 2 člana najzastupljenija u Srbiji za navedeni period (tabela 2.9).

Tabela 2.9 *Domaćinstva Vojvodine prema broju članova, 2002. godine*

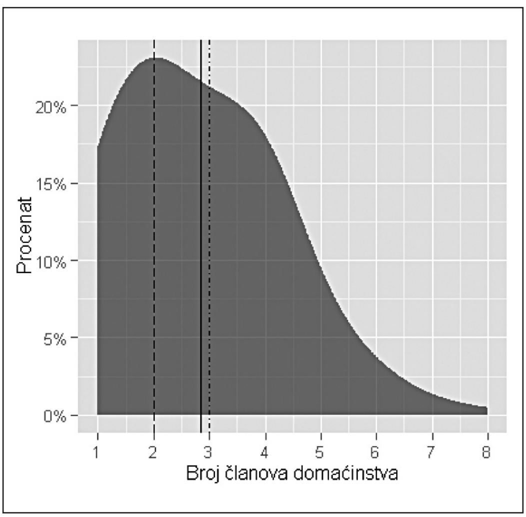
Broj članova domaćinstva	f_i	kumulativ
1	149.867	149.867
2	180.858	330.725
3	139.843	470.568
4	153.886	624.454
5	52.766	677.220

6	22.779	699.999
7	6.591	706.590
8	3.367	709.957
Ukupno	709.957	

$$M_e = x_{\frac{n+1}{2}} = x_{\frac{709.957+1}{2}} = x_{354.979} = 3 \textit{ člana domaćinstva} \qquad [2.10]$$

Na osnovu obrasca [2.11] vidimo da je medijana 354.979. član u seriji i to je broj 3. Njega smo pronašli pomoću kumulativa jer se 354.979. član našao u trećem kumulativu u kojem su prvih 470.568 podataka serije. Vrednost serije koja odgovara ovom kumulativu je upravo broj 3 i to je podatak koji deli ovaj statistički niz na dva jednaka dela. Dakle, polovina domaćinstava u Republici Srbiji, 2002. godine, imala je do 3 člana, a druga polovina više od 3 člana. Podsetimo se da je prosečan broj članova domaćinstva iznosio približno 3 (2,85). Položaj modusa (linija krajnje levo), aritmetičke sredine (linija u sredini) i medijane (linija krajnje desno) prikazali smo grafikonom 2.3.

Grafikon 2.3 *Domaćinstva prema broju članova, 2002. godine*



Primer 14

Kod intervalnih numeričkih serija proces izračunavanja medijane se odvija na sličan način. Za prikaz ove serije koristićemo podatke Nacionalne službe za zapošljavanje o broju nezaposlenih lica u Republici Srbiji, krajem februara 2013. godine.

Tabela 2.10 *Nezaposlena lica prema starosti u regionu južne i istočne Srbije, na kraju februara 2013. godine*

starost	frekvencije	kumulativ
15–19	6.061	6.061
20–24	24.114	30.175
25–29	28.064	58.239
30–34	25.570	83.809
35–39	26.163	109.972
40–44	24.910	134.882
45–49	25.308	160.190
50–54	23.984	184.174
55–59	19.921	204.095
60–64	7.768	211.863
ukupno	211.863	

Izvor: služba za zapošljavanje. *Mesečni statistički bilten*. Februar 2013, str. 20–21 (sajt www.nsz.gov.rs)

Za ovako prikazane podatke najpre je potrebno pronaći medijalni interval i on se određuje na osnovu formule [2.10], koju smo do sada koristili za određivanje medijane i uz pomoć kumulacija.

$$x_{\frac{n+1}{2}} = x_{\frac{211.863+1}{2}} = x_{105.932} = \textit{interval } 35\text{–}39 \textit{ godina}$$

Medijalni interval je starosna grupa od 35 do 39 godina starosti jer se 105.932. podatak u seriji našao u petom kumulativu. Sada je potrebno pronaći medijanu u okviru ovog intervala. Nju ćemo odrediti na osnovu izraza:

$$M_e = L_1 + \frac{L_2 - L_1}{F_{m2} - F_{m1}} \cdot \left(\frac{\Sigma f_i}{2} - F_{m1}\right) \qquad [2.11]$$

gde je:

L_1 – donja granica medijalnog intervala
L_2 – gornja granica medijalnog intervala
F_{m1} – kumulacija intervala iznad medijalnog
F_{m2} – kumulacija medijalnog intervala

Odatle dobijamo:

$$M_e = 35 + \frac{39 - 35}{109.972 - 83.809} \cdot \left(\frac{211.863}{2} - 83.809\right) = 38{,}38 \approx 38 \textit{ godina}$$

Dakle, u regionu južne i istočne Srbije polovina nezaposlenog stanovništva je bila mlađa od 38 godina, a polovina starija. Najveći broj nezaposlenog stanovništva je starosti 28 godina (modus je 28), dok je prosečna starost nezaposlenih iznosila 39 godina (aritmetička sredina je 39,07).

Za parne negrupisane i grupisane statističke serije medijana se dobija izrazom:

$$M_e = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} \qquad [2.12]$$

Primer 15

Za naredni niz podataka potrebno je izračunati medijanu na osnovu izraza [2.12]:

7 13 32 45 59 77

$$M_e = \frac{x_{\frac{6}{2}} + x_{\frac{6}{2}+1}}{2} = \frac{x_3 = x_4}{2} = \frac{32 + 45}{2} = 38{,}5$$

Primer 16

U narednoj tabeli prikazani su rezultati popisa obavljenog u Republici Srbiji 2011. godine, a koji se odnose na raspodelu ženskog stanovništva starog 15 i više godina prema broju živorođene dece.

Tabela 2.11 *Žensko stanovništvo staro 15 i više godina prema broju živorođene dece, popis 2011*

Broj živorođene dece	frekvencija	$x_i \cdot f_i$	kumulativ
1	629.117	629.117	629.117
2	1.361.775	2.723.550	3.352.667
3	283.259	849.777	4.202.444
4	62.397	249.588	4.452.032
5 i više	33.346	166.730	4.618.762
ukupno		4.618.762	

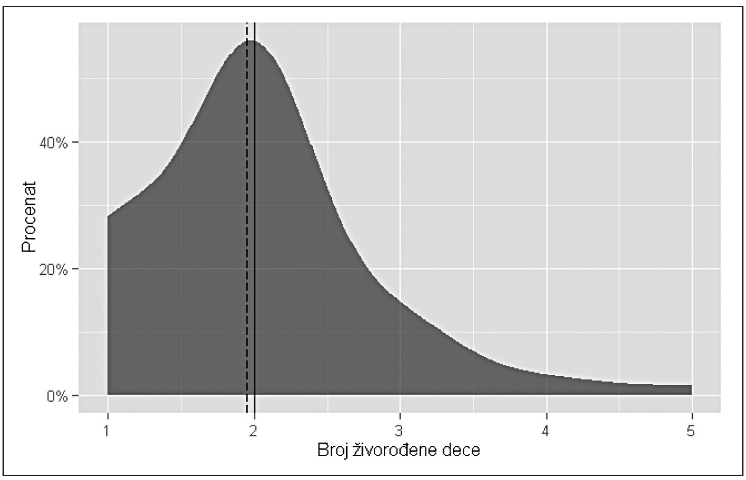
Izvor: *Popis stanovništva, domaćinstva i stanova u 2011. u Republici Srbiji. Fertilitet ženskog stanovništva. Podaci po opštinama i gradovima* (2013). Beograd: Republički zavod za statistiku, str. 240

Na osnovu izraza [2.12] dobijamo:

$$M_e = \frac{x_{\frac{4.618.762}{2}} + x_{\frac{4.618.762}{2}+1}}{2} = \frac{2+2}{2} = 2\ deteta$$

Medijana se nalazi između 2.309.381. i 2.309.382. podatka. Pomoću kumulacije frekvencija vidimo da se oni nalaze u drugoj kumulaciji i da je obeležje druge kumulacije 2 živorođena deteta. Dakle, na osnovu popisa 2011. godine u Republici Srbiji, polovina ženskog stanovništva je rodila jedno ili dvoje dece, a druga polovina više od dva deteta (grafikon 2.4).

Grafikon 2.4 *Raspodela ženskog stanovništva starog 15 i više godina prema broju živorođene dece, popis 2011*[13]



Međutim, za intervalni statistički niz sa parnim brojem podataka koristićemo sledeći izraz:

$$M_e = L_1 + \frac{\frac{N}{2} - \Sigma f_1}{f_2} \cdot i \qquad [2.13]$$

[13] Punom linijom obeležen je položaj modusa i medijane, dok je isprekidanom označena vrednost aritmetičke sredine.

Primer 17

Za prikaz izračunavanja medijane kod intervalne parne serije koristićemo, takođe, podatke Nacionalne službe za zapošljavanje. Ovoga puta medijanu ćemo izračunati na osnovu podataka prikupljenih za nezaposleno stanovništvo u Vojvodini, krajem februara 2013. godine (tabela 2.12).

Tabela 2.12 *Nezaposlena lica prema starosti u regionu Vojvodine, na kraju februara 2013. godine*

starost	frekvencije	kumulativ
15–19	6.840	6.840
20–24	23.163	30.003
25–29	27.508	57.511
30–34	25.349	82.860
35–39	25.010	107.870
40–44	24.046	131.916
45–49	24.851	156.767
50–54	24.310	181.077
55–59	20.174	201.251
60–64	7.041	208.292
ukupno	208.292	

Izvor: Nacionalna služba za zapošljavanje. *Mesečni statistički bilten*. Februar 2013, str. 20–21 (sajt www.nsz.gov.rs)

Izračunavanje započinje pronalaženjem medijalnog intervala. Kako je polovina od ukupnog broja podataka $\frac{208292}{2} = 104.146$ podatak, a on se nalazi u petoj kumulaciji, to je medijalni interval i u ovom slučaju starosna grupa od 35 do 39 godina i u okviru nje se nalazi medijana. Primenom izraza [2.13] dobijamo:

$$M_e = 35 + \frac{\frac{208.292}{2} - (6.840 + 23.163 + 27.508 + 25.349)}{25.010} \cdot 5 = 39{,}35 \approx 39 \text{ godina}$$

Medijana je mera centralne tendencije koja se ubraja među kvantile. Najčešće upotrebljavani kvantili su kvartili. Po definiciji, kvartili su vrednosti numeričke promenljive ili modaliteti rang promenljive koji jedan uređen statistički niz dele na četiri jednaka dela. Tako prvi kvartil predstavlja vrednost obeležja od kojih prvih 25% podataka uređenog statističkog skupa ima manju ili jednaku vrednost tog obeležja, drugi kvartil je medijana, i on kao što smo videli, ima vrednost od koje je 50% elemenata skupa manje ili jednako toj vrednosti, dok se treći kvartil definiše kao ona vrednost od koje 75% elemenata sređenog skupa ima manju ili jednaku vrednost.

Način određivanja kvartila analogan je metodu izračunavanja medijane.

Na osnovu serije neparnih negrupisanih podataka iz prvog primera, rekli smo da je medijana 45.

7 13 32 45 59 77 98

Prvi kvartil je broj 13, jer je:

$$Q_1 = x_{\frac{n+1}{4}} = y_{\frac{7+1}{4}} = x_2 = 13 \qquad [2.14]$$

dok treći dobijamo na osnovu izraza:

$$Q_3 = x_{\frac{3(n+1)}{4}} = y_{\frac{3\cdot 8}{4}} = x_6 = 77 \qquad [2.14a]$$

Dakle, 25% podataka ove serije je manje ili jednako broju 13, 50% podataka je manje ili jednako od 45, dok je 75% podataka manje ili jednako od 77.

Za izračunavanje kvartila kod negrupisanih parnih podataka koristićemo seriju iz četvrtog primera.

7	13	32	45	59	77

Za ove podatke medijana je iznosila 38,5. Prvi kvartil dobijamo na sledeći način:

$$Q_1 = \frac{x_{\frac{n}{4}} + x_{\frac{n}{4}+1}}{2} = \frac{x_{\frac{6}{4}} + x_{\frac{6}{4}+1}}{2} = \frac{x_{1,5} + x_{2,5}}{2} = \frac{10 + 22,5}{2} = 16,5 \quad [2.15]$$

U ovom slučaju vidimo da se prvi kvartil nalazi između $x_{1,5}$ i $x_{2,5}$ podatka, odnosno između prvog i drugog podatka i drugog i trećeg. Aproksimacijom pronalazimo da su to sledeći podaci: $\frac{7+13}{2} = 10$ i $\frac{13+32}{2} = 22,5$.

Treći kvartil je:

$$Q_3 = \frac{x_{\frac{3n}{4}} + x_{\frac{3n+1}{4}}}{2} = \frac{x_{\frac{3\cdot 6}{4}} + x_{\frac{3\cdot 6}{4}+1}}{2} = \frac{x_{4,5} + x_{5,5}}{2} = \frac{52 + 68}{2} = 60 \quad [2.15a]$$

Prikaz izračunavanja kvartila kod grupisanih podataka ilustrovaćemo na drugom i petom primeru. Za neparne i parne grupisane podatke koriste se izrazi [2.14], [2.14a], [2.15] i [2.15a].

Na osnovu podataka o raspodeli domaćinstava Vojvodine prema broju članova u 2002. godini, videli smo da je drugi kvartil (medijana) 3 člana domaćinstva, tj. da 50% članova domaćinstva u Vojvodini brojalo do 3 člana. Za izračunavanje kvartila koristićemo izračunate vrednosti iz tabele 2.13.

Tabela 2.13 *Domaćinstva Vojvodine prema broju članova, 2002. godine*

Broj članova domaćinstva	f_i	$x_i f_i$	kumulativ
1	149.867	149.867	149.867
2	180.858	361.716	511.583
3	139.843	419.529	931.112
4	153.886	615.544	1.546.656
5	52.766	263.830	1.810.486
6	22.779	136.674	1.947.160
7	6.591	46.137	1.993.297
8	3.367	26.936	2.020.233
Ukupno		2.020.233	

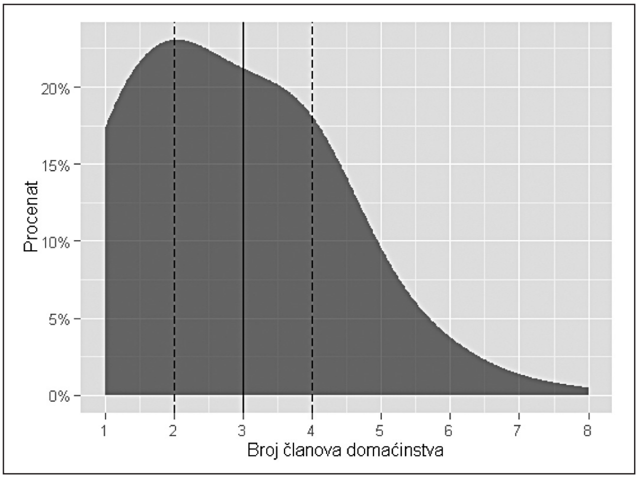
Prvi kvartil je 2, jer je:

$$Q_1 = x_{\frac{2.020.233+1}{4}} = x_{\frac{2.020.233}{4}} = x_{505.058,5} = \frac{x_{505.058} + x_{505.059}}{2} = \frac{2+2}{2} = 2$$

$$Q_3 = x_{\frac{3(2.020.233+1)}{4}} = x_{\frac{3\cdot 2.020.234}{4}} = x_{1.515.175,5} = \frac{x_{1.515.175} + x_{1.515.176}}{2} = \frac{4+4}{2} = 4$$

Odnosno, 25% domaćinstva u Vojvodini, na osnovu popisa iz 2002. godine, imalo je 1 ili 2 člana. Do 3 člana je imalo 50% domaćinstva (vrednost medijane), dok je 75% domaćinstva imalo manje od 4 člana (grafikon 2.5).

Grafikon 2.5 *Domaćinstva Vojvodine prema broju članova, 2002. godine*



Rezultati popisa iz 2011. godine pokazuju da je 50% ženskog stanovništva starog 15 i više godina imalo jedno ili dvoje dece. U ovom slučaju napravili smo novu radnu tabelu (tabela 2.14) koja će nam pomoći prilikom izračunavanja kvartila.

Tabela 2.14 *Žensko stanovništvo staro 15 i više godina prema broju živorođene dece, popis 2011*

Broj živorođene dece	frekvencija	$x_i \cdot f_i$	kumulativ
1	629.117	629.117	629.117
2	1.361.775	2.723.550	3.352.667
3	283.259	849.777	4.202.444
4	62.397	249.588	4.452.032
5 i više	33.346	166.730	4.618.762
ukupno		4.618.762	

Međutim, prvi kvartil takođe iznosi 2, jer je:

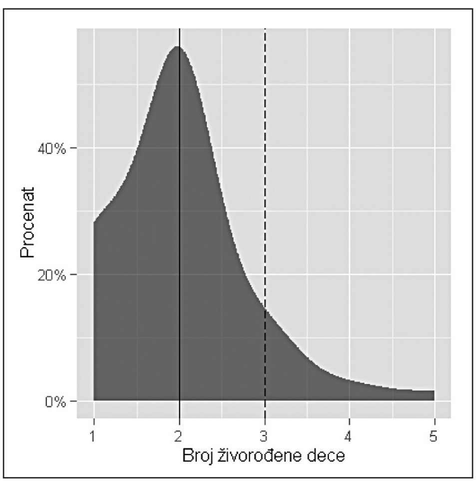
$$Q_1 = \frac{x_{\frac{4.618.762}{4}} + x_{\frac{4.618.762}{4}+1}}{2} = \frac{x_{1.154.690,5} + x_{1.154.691,5}}{2} = \frac{2+2}{2} = 2$$

Podatak da je 25% žena imalo jedno ili dvoje dece, isto kao i 50%, ne treba da zbunjuje jer je frekvencija ženskog stanovništva koje su rodile 2 dece najveća u seriji. Zbog toga je moguće da prvi i drugi kvartil pripadnu istom modalitetu (grafikon 2.6).

Treći kvartil je 3 živorođene dece. Što znači da je 75% ženskog stanovništva u Republici Srbiji, na osnovu popisa iz 2011. godine, rodilo do troje živorođene dece.

$$Q_1 = \frac{x_{\frac{3\cdot 4.618.762}{4}} + x_{\frac{3\cdot 4.618.762+1}{4}}}{2} = \frac{x_{3.464.071,5} + x_{3.464.071,75}}{2} = \frac{3+3}{2} = 3$$

Grafikon 2.6 *Žensko stanovništvo staro 15 i više godina prema broju živorođene dece, popis 2011*



Ako je formirana intervalna numerička serija, u kojoj je zbir frekvencija neparan broj, prvi kvartil dobijamo pomoću formule [2.16], a treći pomoću [2.16a]:

$$Q_1 = L_1 + \frac{\frac{N}{4} - \Sigma f_1}{f_{kvar}} \cdot i \qquad [2.16]$$

$$Q_3 = L_1 + \frac{\frac{3N}{4} - \Sigma f_1}{f_{kvar}} \cdot i \qquad [2.16a]$$

gde je:

L_1 – donja granica kvartilnog intervala
N – ukupan broj članova niza
Σf_1 – zbir frekvencija do kvartlnog intervala
f_{kvar} – frekvencija kvartilnog interval
i – širina kvartilnog intervala.

Da bismo lakše razumeli kvartile i ovom prilikom koristićemo podatke iz kojih smo već izračunali medijanu. Intervalne neparne i parne statističke serije preuzećemo iz trećeg i šestog primera, a za izračunavanje kvartila koristićemo tabelu 2.15.

Tabela 2.15 *Nezaposlena lica prema starosti u regionu južne i istočne Srbije, na kraju februara 2013. godine*

starost	frekvencije	kumulativ
15–19	6.061	6.061
20–24	24.114	30.175
25–29	28.064	58.239
30–34	25.570	83.809

35–39	26.163	109.972
40–44	24.910	134.882
45–49	25.308	160.190
50–54	23.984	184.174
55–59	19.921	204.095
60–64	7.768	211.863
ukupno	211.863	

U ovom primeru, na osnovu tabele 2.15, zbir frekvencija je neparan broj (211.863), pa se prvi kvartil izračunava uz korišćenje kumulacije ispod. Kako prvi kvartil, po svojoj definiciji, obuhvata četvrtinu serije, to jedna četvrtina iznosi 211.863 : 4 = 52.965,5. Ovaj podatak se nalazi u trećoj kumulaciji kojoj odgovara interval 25 – 29. Dakle, donja granica grupnog intervala u kojoj se nalazi prvi kvartil je 25. Ukupan zbir frekvencija je 211.863, dok je zbir frekvencija do kvartilnog intervala 30.175. Frekvencija kvartilnog intervala iznosi 28.064, a njegova širina 5. Sada možemo ove podatke uvrstiti u formulu [2.16] i izračunati prvi kvartil.

$$Q_1 = 25 + \frac{\frac{211.863}{4} - 30.175}{28.064} \cdot 5 \approx 29$$

Za određivanje trećeg kvartilnog intervala treba izračunati tri četvrtine od zbira frekvencija. To je 3 · 211.863 : 4 = 158.897,25. Ovaj broj nalazi se u sedmoj kumulaciji kojoj pripada interval 45–49, pa je donja granica interval trećeg kvartila 45. Zbir frekvencija do ovog kvartilnog intervala iznosi 134882, a njegova frekvencija je 25308. Ako ove podatke uvrstimo u formulu [2.16a], dobijamo:

$$Q_3 = 45 + \frac{\frac{3 \cdot 211.863}{4} - 134.882}{25.308} \cdot 5 = 49.74$$

Jedna trećina nezaposlenih lica u regionu južne i istočne Srbije, krajem februara 2013. godine, bila je mlađa od 29 godina, polovina nezaposlenih je do 39 godina starosti, a tri četvrtine nezaposlenih lica u ovom region je mlađe od 50 godina.

Na sličan način dobijamo kvartile iz parnih intervalnih serija. Za ovu vrstu numeričke serije prvi i treći kvartil ćemo dobiti pomoću sledećih formula:

$$Q_1 = L_1 + \frac{\frac{N+1}{4} - \Sigma f_1}{f_{kvar}} \cdot i \qquad [2.168]$$

$$Q_3 = L_1 + \frac{\frac{3N+1}{4} - \Sigma f_1}{f_{kvar}} \cdot i \qquad [2.18a]$$

Na osnovu tabele 2.7 vidimo da je četvrtina ukupnog broja frekvencija 208.292 : 4 = 52.073 i da se ovaj podatak nalazi u trećoj kumulaciji kojoj pripada interval 25–29. Dakle, donja granica intervala prvog kvartila je 25. Zbir frekvencija do kvartilnog intervala iznosi 30.003, a frekvencija samog kvartilnog intervala je 27.508. Iz ovih podataka, a na osnovu obrasca [2.18], dobijamo da je prvi kvartil 29,01.

$$Q_1 = 25 + \frac{\frac{208.292 + 1}{4} - 30.003}{27.508} \cdot 5 = 29{,}01$$

Za treći kvartil potrebno je pronaći odgovarajuće vrednosti kao u prethodnom primeru. Tri četvrtine od ukupnog zbira frekvencija iznosi 156.219. Ovaj podatak je smešten u sedmoj kumulaciji kojoj pripada interval 45–49, pa je donja granica intervala trećeg kvartila 45 godina. Zbir frekvencija do ovog kvartilnog intervala iznosi 131.916, a frekvencija kvartilnog intervala 24.851. Sada ove podatke možemo uvrstiti u obrazac i dobiti vrednost trećeg kvartila.

$$Q_3 = 45 + \frac{\frac{3 \cdot 208.292 + 1}{4} - 131.916}{24.851} \cdot 5 = 49{,}89$$

Tabela 2.16 *Nezaposlena lica prema starosti u regionu Vojvodine, na kraju februara 2013. godine*

starost	frekvencije	kumulativ
15–19	6.840	6.840
20–24	23.163	30.003
25–29	27.508	57.511
30–34	25.349	82.860
35–39	25.010	107.870
40–44	24.046	131.916
45–49	24.851	156.767
50–54	24.310	181.077
55–59	20.174	201.251
60–64	7.041	208.292
ukupno	208.292	

Iz ove analize možemo zaključiti da je, u februaru 2013. godine, jedna četvrtina nezaposlenih Vojvođana mlađa od 29 godina (ili sa navršenim 29 godina), polovina ima do 39 godina, dok je tri četvrtine nezaposlenih lica mlađe od 50 godina. Takođe možemo primetiti da je raspodela nezaposlenog stanovništva između Vojvodine i istočne i južne Srbije (s izuzetkom medijane), krajem februara 2013. godine bila ista.

Na kraju ovog dela treba dodati da ovde obrađene mere centriranosti nisu jedine koje se upotrebljavaju za deskripciju statističkih serija. Tako se u grupi potpunih srednjih vrednosti, pored aritmetičke i geometrijske sredine, mogu koristiti harmonijska, kvadratna, kubna i logaritamska sredina. Poređenjem ovih brojevnih sredina prvi se bavio istaknuti francuski matematičar Koši (Augustin Louis Couchy, 1789–1857), pa se na osnovu njegovog rada razvila Košijeva teorema koja se bavi dokazivanjem reda velična srednjih vrednosti. Izračunavanjem sredina na istim numeričkim serijama zadovoljava se ova teorema po kojoj je harmonijska sredina manja ili jednaka geometrijskoj, geometrijska manja ili jednaka aritmetičkoj, ova pak manja ili jednaka kvadratnoj, a kvadratna manja ili jednaka kubnoj sredini.

S druge strane, kvantili se upotpunjuju kvintalima, decilima i percentilima. Tako kvintali raščlanjuju sređeni niz na pet jednakih delova, decili na deset jednakih delova, a percentili na sto delova. Ono što je značajno pomenuti je da broj kvantila za jedan manji od njihovog reda. Tako kvartila ima 3, kvintala 4, decila 9, a percentila 99. Međutim, ove mere se slabije koriste u opisivanju serija, pa ih ovom prilikom nećemo objašnjavati.

ZADACI[14]

1. Navedite po čemu se razlikuju potpune od pozicionih srednjih vrednosti.
2. Koje merne skale, i zbog čega, možete opisivati aritmetičkom i geometrijskom sredinom, modusom, medijanom i kvartilima?
3. Objasnite aritmetičku sredinu i navedite njene osobine.
4. U kojim slučajevima se koriste skraćeni metodi za izračunavanje aritmetičke sredine?
5. Šta je to geometrijska sredina?
6. Gde se geometrijska sredina najčešće upotrebljava?
7. Kako glasi definicija modusa i koliko jedna serija može da ih ima?
8. Šta je medijana po svojoj definiciji?
9. Na koji način kvartili dele jednu seriju i koliko ih ima?

[14] Zadaci se mogu preuzeti sa sajta Filozofskog fakulteta u Novom Sadu: http://ff.uns.ac.rs/studenti/oglasne%20table/studenti_oglasne_table_sociologija.html

10. Procenjeni broj stanovnika prema polu, tokom perioda od 1990. do 2000. godine, dat je sledećom tabelom:

Tabela 1 *Procenjeni broj stanovnika R Srbije prema polu*

Godina	Centralna Srbija		Vojvodina	
	muško	žensko	muško	žensko
1990.	2.890.281	2.958.836	997.835	1.050.985
1991.	2.861.988	2.949.229	980.283	1.033.089
1992.	2.860.227	2.950.705	976.632	1.030.855
1993.	2.858.394	2.951.617	973.418	1.028.190
1994.	2.857.490	2.951.755	970.791	1.025.387
1995.	2.855.539	2.950.752	968.005	1.022.562
1996.	2.852.666	2.948.309	964.967	1.019.152
1997.	2.847.454	2.944.189	961.423	1.015.513
1998.	2.701.756	2.833.280	984.762	1.047.947
1999.	2.684.348	2.822.588	984.839	1.048.626
2000.	2.671.946	2.812.974	983.831	1.047.595

Izvor: *Demografska statistika 2002. i 2003.* (2006). Beograd: Republički zavod za statistiku, str. 36.

Uporedite i protumačite geometrijske stope rasta muškog i ženskog stanovništva centralne Srbije i Vojvodine u posmatranom periodu.

11. U narednoj tabeli[15] prikazana je projekcija stanovništva Srbije, prema grupama starosti i polu, za 2022. godinu, koju je Republički zavod za statistiku Srbije odredio na osnovu podataka popisa stanovništva 2002. godine i pretpostavkama o budućem kretanju migracija, fertiliteta i mortaliteta.

[15] U tabeli su izostavljeni podaci o procenjenom broju stanovnika za poslednji interval „80 i više”.

Tabela 2 *Projekcija stanovništva R Srbije prema starosti i polu*

Godine starosti	2022.		
	ukupno	muški	ženski
0–4	347.018	177.082	169.936
5–9	366.986	187.155	179.831
10–14	383.018	195.370	187.649
15–19	395054	200.374	194.682
20–24	370.301	188.758	181.541
25–29	410.150	210.349	199.802
30–34	448.976	228.736	220.241
35–39	504.890	256.340	248.549
40–44	518.838	262.084	256.752
45–49	506.357	251.939	254.418
50–54	465.709	228.121	237.589
55–59	455.841	219.766	236.076
60–64	465.861	221.299	244.563
65–69	486.914	228.140	258.774
70–74	394.987	178.450	216.537
75–79	210.741	89.984	120.756

Izvor: *Demografska statistika 2002. i 2003.* (2006). Beograd: Republički zavod za statistiku, str. 49.

Jednim od metoda za skraćeno izračunavanje aritmetičke sredine odredite prosečnu starost ukupnog, muškog i ženskog stanovništva Srbije u 2022. godini. Potom odredite medijalnu i modalnu prosečnu starost, kao i prvi i treći kvartil. Protumačite dobijene podatke.

MERE VARIJABILITETA

Merenje varijabilnosti kvantitativnih promenljivih

Merenje varijabilnosti kvalitativnih promenljivih

3 MERE VARIJABILITETA

U prethodnom poglavlju smo pokazali da aritmetička sredina predstavlja prosek izračunat na osnovu svih podataka jedne kvantitativne promenljive. Međutim, ukoliko uporedimo različite numeričke serije možemo doći do zaključka da one ponekad imaju jednake aritmetičke sredine, pa čak i moduse i medijane, iako se međusobno veoma razlikuju (vidi tabele 3.1 i 3.1a). Različitost može biti u broju podataka, njihovoj vrednosti, ali i u veličini odstupanja pojedinačnih vrednosti od aritmetičke sredine. Upravo ova odstupanja predstavljaju varijabilitet (disperziju, raspršenost ili rastureno) posmatrane serije.

Primer 1

U tabeli 3.1 prikazana je raspodela 35 domaćinstava prema broju članova u tri grada.

Tabela 3.1 *Raspodela domaćinstava prema broju članova u gradovima*

Grad A		Grad B		Grad C	
modalitet	frekvencija	modalitet	frekvencija	modalitet	frekvencija
1	2	1	10	1	5
2	9	2	2	2	9
3	15	3	14	3	10

4	4	4	0	4	3
5	3	5	0	5	3
6	2	6	9	6	5

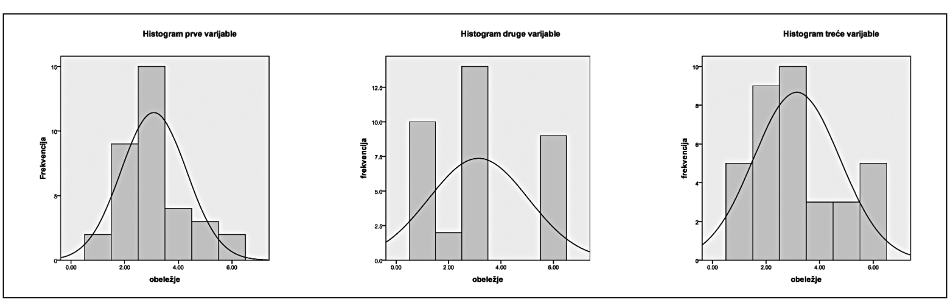
Na osnovu podataka iz tabele 3.1 možemo primetiti da su vrednosti aritmetičke sredine, modusa i medijane iste za sve tri promenljive (tabela 3.1.1). Međutim, one se među sobom ipak razlikuju, a to se najbolje može videti po drugačijim raspodelama frekvencija oko njihovih modaliteta. Upravo ove raspodele frekvencija utiču i na veličinu odstupanja vrednosti obeležja od aritmetičke sredine i stvaraju varijabilitet.

Tabela 3.1.1 *Deskriptivna statistika podataka iz tabele 3.1*

	Grad A (x)	Grad B (y)	Grad C (z)
Aritmetička sredina	≈3	≈3	≈3
Modus	3	3	3
Medijana	3	3	3

Pošto govorimo o različitim statističkim serijama možda je najbolje prikazati ih grafičkim putem. Uobičajeno je da se varijabilitet serija grafički prikazuje histogramom[16] (grafikon 3.1) ili box-plotom (grafikon 3.2).

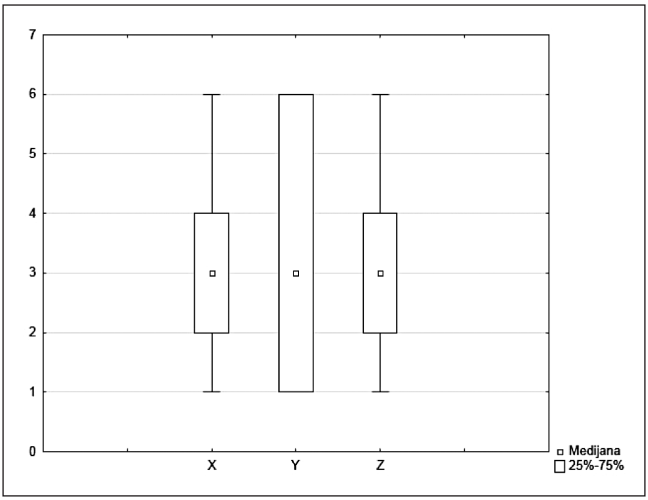
Grafikon 3.1 *Histogrami frekvencija promenljivih iz tabele 3.1*



[16] Način crtanja i objašnjenje histograma dali smo u uvodnom delu.

Box-plotom (B-P) se analizira varijabilitet dve ili više numeričkih serija izraženih u istim mernim jedinicama. Kako ovaj način prikazivanja pruža važne informacije o prirodi proučavanih podataka (kao što su interval varijacije, interkvartilna razlika (raspon varijacije središnjih 50% podataka koji su uređeni po veličini), moguća asimetrija ili postojanje ekstremnih vrednosti, njime dobijamo nešto drugačiju sliku o varijabilitetu nego histogramom. Tako se prva i treća promenljiva na osnovu box-plota uopšte ne razlikuju dok njihovi histogrami frekvencija pokazuju značajna odstupanja. Zbog toga je poželjno kombinovati ove dve vrste grafičkog prikaza ukoliko želimo uvid u varijabilitet promenljivih koje proučavamo.

Grafikon 3.2 *Box-plot promenljvih iz tabele 3.1*



B-P konstruiše se na osnovu nekoliko koraka:

– rangiramo podatke u rastući poredak i izračunamo vrednost medijane, prvog i trećeg kvartila, te interkvartilne razlike (obrazac [3.2]);
– potom izračunamo donju $\left(Q_1 - (1{,}5 \cdot I_q)\right)$ i gornju $\left(Q_3 + (1{,}5 \cdot I_q)\right)$ unutrašnju granicu;
– odredimo najmanju i najveću vrednost u zadatoj seriji podataka između dve unutrašnje granice;

– na vertikalnoj liniji prikažemo vrednosti serije, a oko nje ucrtamo pravougaonik čija donja strana počinje pozicijom prvog kvartila, a desna pozicijom trećeg kvartila. Unutar pravougaonika se povuče vertikalna linija (ili ucrta tačka) na poziciji medijane. U slučaju simetrične serije podataka[17] medijana će se nalaziti na sredini pravougaonika, a raspršenost će biti skoro jednaka sa obe strane pravougaonika;
– povučemo dve linije kojima spajamo pozicije najmanje i najveće vrednosti između dve unutrašnje granice sa pravougaonikom. Vrednosti koje se nađu izvan ove dve unutrašnje granice obeležavaju se zvezdicom i nazivaju se nestandardne opservacije. One mogu biti umerene ili ekstremne;
– na kraju izračunamo donju $\left(Q_1 - (3{,}0 \cdot I_q)\right)$ i gornju $\left(Q_3 - (3{,}0 \cdot I_q)\right)$ spoljašnju granicu. U slučaju da se vrednosti nestandardnih opservacija nađu izvan bilo koje od dve unutrašnje granice, ali unutar bilo koje od dve spoljašnje granice, nazivaju se umerene nestandardne opservacije. Vrednost/i koje se nađu izvan bilo koje od dve spoljašnje granice nazivaju se ekstremne nestandardne opservacije.

3.1 MERENJE VARIJABILNOSTI KVANTITATIVNIH PROMENLJIVIH

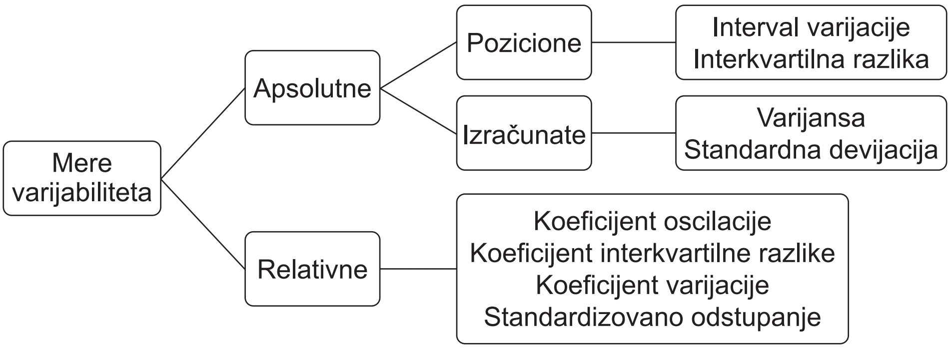
U statistici razlikujemo dve grupe mera varijabiliteta. Prva se odnosi na apsolutne mere, dok se u drugu svrstavaju relativni pokazatelji. Razlika je u tome što mere prve grupe izražavaju varijabilitet u apsolutnim iznosima onih mernih jedinica u kojima se iskazuje serija podataka za koju se on izračunava, dok se relativne mere iskazuju procentualno i mogu se koristiti za poređenja varijabiliteta različito iskazanih serija. Ova različitost može da se ogleda u rangu jedinica u kojima se serije iskazuju (npr. hiljade dinara naspram miliona dinara) ili u različitim mernim jedinicama (npr. visina naspram težine).

Definisane su mnoge mere varijabilnosti, ali mi ćemo ovom prilikom predstaviti samo one najznačajnije i to sledećim redosledom: krenućemo

[17] Simetrijom serija bavićemo se u narednom poglavlju.

od apsolutnih mera, vodeći računa o njihovoj preciznosti i ukazujući na najznačajnije prednosti i nedostatke svake od opisivanih mera, a potom ćemo objasniti i odabrane relativne mere varijabilnosti, njihovo izračunavanje i tumačenje.

Shematski prikaz najznačajnijih mera varijabiliteta kvantitativnih promenljivih



3.1.1 Apsolutne mere varijabilnosti kvantitativnih promenljivih

Interval varijacije (rang) je najjednostavniji pokazatelj varijabiliteta i predstavlja algebarsku razliku između maksimalne i minimalne vrednosti obeležja jedne numeričke serije, bila ona sa negrupisanim ili grupisanim podacima. Izračunava se pomoću izraza:

$$I = x_{max} - x_{min} \quad [3.1]$$

Na osnovu primera iz tabele 3.1 interval varijacije naše prve promenljive iznosio bi:

$$I = 6 - 1 = 5$$

Isti varijabilitet na osnovu ovog pokazatelja imale bi i ostale dve promenljive.

Rezultat ove mere varijacije je samo približna informacija o varijabilnosti, jer uzima samo dva obeležja iz serije, najveću i najmanju. Kako ekstremne vrednosti ponekad mogu biti i veoma udaljene od najvećeg dela vrednosti obeležja, to je interval varijacije krajnje nepodesan za statističku analizu. Ponekad se možemo susresti sa numeričkim serijama koje imaju isti interval varijacije a da se međusobno značajno razlikuju upravo po grupisanju ostalih vrednosti proučavanog obeležja oko aritmetičke sredine. Dakle, njegov osnovni nedostatak je taj što ne zavisi od svih vrednosti obeležja kvantitativne promenljive, dok je njegova prednost u brzom i jednostavnom računanju.

Nedostatak intervala varijacije donekle je ublažen interkvartilnom razlikom, tj. razlikom između trećeg i prvog kvartila. Kako kvartili dele jedan statistički niz na četiri jednaka dela, interkvartilnom razlikom [3.2] se varijabilitet podataka jedne numeričke serije meri polovinom podataka.

$$I_q = Q_3 - Q_1 \tag{3.2}$$

Kako se u našem primeru radi o parnim grupisanim podacima za izračunavanje prvog i trećeg kvartila koristićemo formule [2.16] i [2.16a]. Tako interkvartilna razlika prve promenljive (grad A) iznosi $I_{q1} = 4{,}5 - 3 = 1{,}5$, druge (grad B) $I_{q2} = 5 - 3 = 2$, dok je za treću (grad C) ona $I_{q3} = 6 - 3 = 3$. Vidimo da se interkvartilnom razlikom već uočavaju razlike u varijabilnosti između promenljivih, dok je interval varijacije bio neosetljiv na njih.

Već i na osnovu poređenja vrednosti intervala varijacije i interkvartilne razlike mogu se donositi neki zaključci o seriji koju posmatramo. Ako ih izračunamo na istoj seriji i dobijemo veliku vrednost intervala varijacije, a malu interkvartilnu razliku, možemo doneti zaključak o postojanju ekstremnih vrednosti i o maloj varijaciji ostalih vrednosti obeležja. Drugim rečima, u posmatranoj seriji se uočavaju ekstremne vrednosti, dok se ostali podaci u većoj meri grupišu oko aritmetičke sredine.

Preciznost merenja varijabilnosti zavisi od izračunavanja odstupanja svih pojedinačnih vrednosti obeležja od aritmetičke sredine. Već smo videli da je algebarski zbir svih odstupanja vrednosti obeležja jedne promenljive od aritmetičke sredine jednak nuli i zbog toga sledeća mera varijabiliteta, umesto algebarskih, koristi apsolutna odstupanja. Ova mera naziva se srednje apsolutno odstupanje ([3.3] za negrupisane podatke i [3.3a] za grupisane) i pokazuje prosek apsolutnih odstupanja pojedinačnih vrednosti obeležja od aritmetičke sredine:

$$S_o = \frac{\Sigma_{i=1}^{n} \left| x - \bar{x} \right|}{n} \tag{3.3}$$

$$S_o = \frac{\Sigma_{i=1}^{n} f_i \left| x - \bar{x} \right|}{\Sigma_{i=1}^{n} f_i} \tag{3.3a}$$

Srednje apsolutno odstupanje se može izračunati i na osnovu odstupanja vrednosti obeležja od medijane ([3.3b] za negrupisane i [3.3c] za grupisane podatke):

$$S_o = \frac{\Sigma_{i=1}^{n} \left| x - M_e \right|}{n} \tag{3.3b}$$

$$S_o = \frac{\Sigma_{i=1}^{n} f_i \left| x - M_e \right|}{\Sigma_{i=1}^{n} f_i} \tag{3.3c}$$

Ukoliko je $S_o = 0$ odstupanja ne postoje. Iako na srednje apsolutno odstupanje utiču svi članovi posmatrane serije srazmerno svojoj veličini, njegov osnovni nedostatak je upravo taj što uzima apsolutna, a ne realna odstupanja.

Pogledajmo opet naše promenljive iz primera 3.1. Priroda njihovih podataka zahteva od nas korišćenje formula [3.3a] i [3.3c] i zbog toga ćemo formirati sledeću radnu tabelu (tabela 3.1.2), koja će nam olakšati izračunavanje srednjeg apsolutnog odstupanja. Aritmetička sredina prve promenljive iznosi 3,09, druge 3,14 i treće 3,14.

3.1.2 *Radna tabela za izračunavanje srednjeg apsolutnog odstupanja*

Grad A				Grad B				Grad C			
x_i	f_i	$f_i(x_i-\bar{x})$	$f_i(x_i-\bar{x})$	x_i	f_i	$f_i(x_i-\bar{x})$	$f_i(x_i-\bar{x})$	x_i	f_i	$f_i(x_i-\bar{x})$	$f_i(x_i-\bar{x})$
1	2	–4,17	4,17	1	10	–21,43	21,43	1	5	–10,71	10,71
2	9	–9,77	9,77	2	2	–2,29	2,29	2	9	–10,29	9
3	15	–1,29	1,29	3	14	–2,00	2,00	3	10	–1,43	1,43
4	4	3,66	3,66	4	0	0	0	4	3	2,57	2,57
5	3	5,74	5,74	5	0	0	0	5	3	5,57	5,57
6	2	5,83	5,83	6	9	25,71	25,71	6	5	14,29	14,29
Ukupno		–	30,46	Ukupno		–	51,43	Ukupno		–	43,57

$$S_{o1}=\frac{30,46}{35}=0,87 \qquad S_{o2}=\frac{51,43}{35}=1,47 \qquad S_{o3}=\frac{43,57}{35}=1,24$$

Na osnovu izračunatih srednjih apsolutnih odstupanja možemo da zaključimo da je druga promenljiva sa najvećim odstupanjima njenih elemenata od aritmetičke sredine. Na isti način bismo formirali radnu tabelu za izračunavanje srednjeg apsolutnog odstupanja na osnovu formule [3.3c]. Jedina razlika je što bismo izračunavali odstupanje elemenata od medijane, a ne od aritmetičke sredine.

Varijansa ili srednje kvadratno odstupanje je prosek kvadrata odstupanja pojedinačnih vrednosti od aritmetičke sredine (ili od medijane). Ona se kreće u granicama $[0, +\infty)$.

Kako se varijansa može izračunavati iz svih vrsta numeričkih serija, u narednoj tabeli navešćemo formule za izračunavanje varijanse za odgovarajuće serije podataka u odnosu na ceo osnovni skup ili uzorak.

Tabela 3.2 *Obrasci za izračunavanje varijanse* [3.4] *na osnovu vrste numeričke serije i u odnosu na osnovni skup ili uzorak*

Vrsta numeričke serije	Osnovni skup	Uzorak
Negrupisani podaci	$\sigma^2 = \frac{\Sigma_{i=1}^{N}(x_i - \mu)^2}{N}$, $\sigma^2 = \frac{\Sigma_{i=1}^{N} x_i^2}{N} - \mu^2$ ili $\sigma^2 = \frac{\Sigma_{i=1}^{N} x_i^2 - \frac{(\Sigma_{i=1}^{N} x_i)^2}{N}}{N}$	$\sigma^2 = \frac{\Sigma_{i=1}^{N}(x_i - \bar{x})^2}{n-1}$ [18] $\sigma^2 = \frac{\Sigma_{i=1}^{N}(x_i^2 - \mathrm{n}\bar{x}^2}{n-1}$ $\sigma^2 = \frac{\Sigma_{i=1}^{n} x_i^2 - \frac{(\Sigma_{i=1}^{n} x_i)^2}{n}}{n-1}$
Grupisani podaci	$\sigma^2 = \frac{\Sigma_{i=1}^{k} f_i (x_i - \mu)^2}{\Sigma_{i=1}^{k} f_i}$ ili $\sigma^2 = \frac{\Sigma_{i=1}^{k} f_i x_i^2}{\Sigma_{i=1}^{k} f_i} - \mu^2$	$\sigma^2 = \frac{\Sigma_{i=1}^{k} f_i (x_i - \bar{x})^2}{\Sigma_{i=1}^{k} f_i - 1}$ ili $\sigma^2 = \frac{\Sigma_{i=1}^{k} f_i x_i^2}{\Sigma_{i=1}^{k} f_i} - \bar{x}^2$
Metod proizvoljnog početka	$\sigma^2 = \frac{\Sigma_{i=1}^{k} f_i y_i^2}{\Sigma_{i=1}^{k} f_i} - \bar{y}^2$	$\sigma^2 = \frac{\Sigma_{i=1}^{k} f_i y_i^2 - \bar{y}^2 \Sigma_{i=1}^{k} f_i}{\Sigma_{i=1}^{k} f_i - 1} - \bar{y}^2$
Metod linearne transformacije obeležja	$\sigma^2 = i^2 \left(\frac{\Sigma_{i=1}^{k} z_i^2 f_i}{\Sigma_{i=1}^{k} f_i}\right)$	$\sigma^2 = i^2 \left(\frac{\Sigma_{i=1}^{k} z_i^2 f_i - \bar{z}^2 \Sigma_{i=1}^{k} f_i}{\Sigma_{i=1}^{k} f_i}\right)$
Relativne frekvencije	$\sigma^2 = \sum_{i=1}^{k} (f_r)_i (x_i - \mu)^2$	$\sigma^2 = \sum_{i=1}^{k} (f_r)_i (x_i - \bar{x})^2 \frac{\Sigma_{i=1}^{k} f_i}{\Sigma_{i=1}^{k} f_i - 1}$

Varijansa se ne izražava u jedinicama mere u kojoj je izražena statistička serija, jer je to mera drugog stepena. Ukoliko bismo je tako izraža-

[18] Imenilac $n - 1$ predstavlja broj stepeni slobode.

vali, dobili bismo, na primer, dinara na kvadrat, broj članova domaćinstva na kvadrat itd. Zbog toga se izračunava mera varijabilnosti prvog stepena, koja se naziva standardna devijacija ili standardno odstupanje. Pokazuje srednju meru odstupanja pojedinačnih vrednosti obeležja od aritmetičke sredine i izražava se u jedinicama mere u kojoj je iskazana serija. Vrednost standardne varijacije se može kretati u granicama od $[0, +\infty)$. Dobija se izračunavanjem kvadratnog korena iz varijanse.

$$\sigma = \sqrt{\sigma^2}. \tag{3.5}$$

Kako su podaci iz tabele 3.1 grupisani za izračunavanje varijanse, možemo koristiti jednu od dve formule navedene u tabeli 3.2, a koje se odnose na grupisane podatke i na uzorak.

Tabela 3.1.3 *Radna tabela za izračunavanje varijanse i standardne devijacije iz grupisanih podataka (primer prve promenljive)*

x_i	f_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i\,(x_i - \bar{x})^2$
1	2	–2,09	4,35	8,70
2	9	–1,09	1,18	10,61
3	15	–0,09	0,01	0,11
4	4	0,91	0,84	3,34
5	3	1,91	3,67	10,99
6	2	2,91	8,49	16,99
Ukupno	/	/	/	41,94

$$\sigma^2 = \frac{\Sigma_{i=1}^{k} f_i\,(x_i - \bar{x})^2}{\Sigma_{i=1}^{k} f_i - 1} = \frac{41,94}{34} = 1,23$$

$$\sigma = \sqrt{1,23} = 1,1$$

Varijabilitet određen intervalom varijacije, interkvartilnom razlikom i standardnom devijacijom (odnosno varijansom), prikazan je u tabeli 3.1.4. Ovi rezultati jasno pokazuju razlike u iskazivanju varijabiliteta. U ovom slučaju, interval varijacije ne pokazuje osetljivost na različitu raspršenost podataka u navedenim promenljivima, dok su interkvartilnom razlikom one primetne. Razlog tome je, kako smo već naveli, što interkvartilna razlika koristi 50% od ukupnog broja podataka u seriji. Međutim, velika razlika između ova dva pokazatelja nam govori da u sve tri promenljive postoje ekstremne vrednosti i da su one, u odnosu na ostala grupisanja podataka, najuočljivije u prvoj promenljivoj. Na kraju tabele date su izračunate vrednosti standardnih devijacija, koje su dosta manje od vrednosti prve dve mere. Tako su u prvoj promenljvoj odstupanja podataka od proseka za 1,1 jedinica mere, u drugoj su ona najveća i iznose 1,9, dok za treću ona iznose 1,6. Primetićete da su sada i tumačenja varijabiliteta nešto drugačija. Dok je interkvartilna razlika pokazala najveći varijabilitet u trećoj promenljivoj, standardnom devijacijom vidimo da je odstupanje od proseka ipak najveće u drugoj promenljivoj.

Tabela 3.1.4 *Varijabiliteti izračunat na osnovu različitih apsolutnih mera*

Mere varijabiliteta	Grad A	Grad B	Grad C
Interval varijacije	5	5	5
Interkvartilna razlika	1,5	2	3
Varijansa	1,2	3,6	2,6
Standardna devijacija	1,1	1,9	1,6

Kako smo na osnovu prvog primera hteli da objasnimo šta je varijabilitet, faktore koji na njega utiču, kao i preciznost različitih apsolutnih mera, sledeći primer imaće za cilj konkretnije tumačenje varijabiliteta u ispitivanju društvenih pojava, ali i objašnjenje komparativne analize varijabilnosti pomoću apsolutnih mera.

Primer 2

U narednoj tabeli prikazani su podaci rezultata popisa iz 2002. i 2011. godine, a koji se odnose na žensko stanovništvo staro 15 i više godina prema starosti i broju živorođene dece. Kako se u odabranim knjigama popisa rezultati popisa različito prikazuju, izvođenje komparativne analize zahteva prikazivanje podataka na jednoobrazan način. Stoga smo se opredelili da analiziramo žensko stanovništvo staro od 20 do 29 godina prema jednom, dva, tri i četiri živorođena deteta. Zadatak, koji nam predstoji, je da pomoću izračunatih standardnih varijacija uporedimo odstupanja koje ove serije imaju od proseka.

Primer dozvoljava korišćenje standardne devijacije za poređenje varijabiliteta upravo zbog jednakosti aritmetičkih sredina. Zbog toga što u praksi postoje serije sa jednakim prosecima, standardnu devijaciju ne smemo apriori odbaciti kada je poređenje odstupanja od proseka u pitanju.

Tabela 3.3 *Žensko stanovništvo staro od 20 do 29 godina prema broju živorođene dece, popis 2002. i 2011. godine*

Broj živorođene dece	Žene stare od 20 do 29 godina koje su rađale	
	2002.	2011.
1	105.993	78.112
2	84.999	52.956
3	11.435	9.591
4	1.840	2.168
Ukupno	204.267	142.827

Izvor: *Popis stanovništva, domaćinstva i stanova u 2002. Žensko stanovništvo staro 15 i više godina. Podaci po opštinama* (2004). Beograd: Republički zavod za statistiku, str. 9.

Popis stanovništva, domaćinstava i stanova 2011. u Republici Srbiji. Fertilitet ženskog stanovništva. Podaci po opštinama i gradovima (2013). Beograd: Republički zavod za statistiku, str. 22

U ovom primeru koristićemo sledeći obrazac varijanse za izračunavanje varijabiliteta:

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i x_i^2}{\sum_{i=1}^{k} f_i} - \mu^2$$

jer se radi o grupisanim podacima i o osnovnom skupu[19]. Za njegovo izračunavanje poslužićemo se sledećom radnom tabelom:

Tabela 3.3.1 *Radna tabela za izračunavanje varijanse i standardne devijacije*

x_i	Žene stare od 20 do 29 godina koje su rađale			
	2002.			
	f_i	$x_i f_i$	x_i^2	$f_i x_i^2$
1	105.993	105.993	1	105.993
2	84.999	169.998	4	339.996
3	11.435	34.305	9	102.915
4	1.840	7.360	16	29.440
Suma	204.267	317.656		578.344

$$\mu = \frac{317.656}{204.267} = 1{,}55 = \textit{deteta}$$

$$\sigma^2_{2002} = \frac{578.344}{204.267} - \left(\frac{317.656}{204.267}\right)^2 = 2{,}83 - 2{,}42 = 0{,}41$$

$$\sigma^2_{20-29} = \sqrt{0{,}41} = 0{,}64 \textit{ deteta}$$

[19] Prikazani su rezultati popisa, a oni obuhvataju svu živorođenu decu od žena određene starosti na teritoriji Republike srbije.

Tabela 3.3.2 *Radna tabela za izračunavanje varijanse i standardne devijacije*

x_i	Žene stare od 20 do 29 godina koje su rađale			
	2011.			
	f_i	$x_i f_i$	x_i^2	$f_i x_i^2$
1	78.112	78.112	1	78.112
2	52.956	105.912	4	221.824
3	9.591	28.773	9	86.319
4	2.168	8.672	16	34.688
Suma	142.827	221.469		420.943

$$\mu = \frac{221.469}{142.827} = 1,55 = \textit{deteta}$$

$$\sigma^2{}_{2011} = \frac{420.943}{142.827} - \left(\frac{221.469}{142.827}\right)^2 = 2,95 - 2,40 = 0,55$$

$$\sigma^2{}_{20\text{–}29} = \sqrt{0,55} = 0,74 \; \textit{deteta}$$

U ovom slučaju možemo zaključiti da je odstupanje od prosečnog broja živorođene dece (koji iznosi 1,55 deteta) nešto veće kod žena starih 20–29 godina na osnovu popisa 2011. godine nego što pokazuje popis iz 2002. godine.

Sva navedena ograničenja u poređenju disperzije podataka od proseka kod više statističkih serija prevladana su relativnim merama varijabilnosti.

3.1.2 Relativne mere varijabilnosti kvantitativnih promenljivih

Ono što je zajedničko svim relativnim merama je da pomoću njih možemo da poredimo varijabilitet različitih numeričkih serija podataka.

Koeficijent oscilacije (ili relativni razmak varijacije) je odnos intervala varijacije i aritmetičke sredine. Izračunavamo ga pomoću sledeće formule:

$$K_{os} = \frac{I}{\mu} \qquad [3.6]$$

Pandan interkvartilnoj razlici je koeficijent interkvartilne razlike (ili kako se još naziva koeficijent diferencijacije). Kao relativna mera disperzije središnjih podataka, koeficijent interkvartilne razlike predstavlja odnos interkvartilne razlike i zbira kvartila. Drugim rečima, „pokazuje koliko procenata iznosi interkvartilna varijacija od zbira vrednosti prvog i trećeg kvartila" (Stojković, 2001: 215). Definiše se pomoću izraza:

$$V_Q = \frac{Q_3 - Q_1}{Q_3 + Q_1} \cdot 100 \qquad [3.7]$$

Koristi se za poređenje stepena disperzije raznorodnih statističkih nizova.

Koeficijent varijacije pokazuje procentualno odstupanje standardne devijacije od aritmetičke sredine. Izračunava se na osnovu formule:

$$V = \frac{\sigma}{\mu} \cdot 100 \qquad [3.8]$$

Primer 4

Za poređenje serija izraženih u različitim jedinicama mere, pomoću koeficijenta varijacije, poslužićemo se primerom datim u knjizi autora Borisa Petza (Petz, 2007: 65–66) u kojem se navode merenja visine i težine 10-godišnjih dečaka i devojčica u školama na teritoriji Zagreba. Rezultati ovih merenja pokazuju da je izmereno ukupno 612 dečaka i 684 devojčica. Prosečna visina dečaka bila je $\bar{x}$ = 134,4 *cm*, a standardna devijacija σ = 6,06 *cm*, dok je prosečna težina dečaka iznosila $\bar{x}$ = 29,2 *kg* sa standardnom devijacijom od σ = 6,89 *kg*. Zadatak koji autor postavlja je da se ispita da li izmereni dečaci variraju više u visini ili u težini? Odgovor na ovo pitanje daće nam rezultati koeficijenta varijacije.

$$V_{visina-dečaci} = \frac{6,06}{134,4} \cdot 100 = 4,51\%$$

$$V_{težina-dečaci} = \frac{3,89}{29,2} \cdot 100 = 13,32\%$$

Prema njima, možemo zaključiti da dečaci u ovom istraživanju variraju više u težini (jer je koeficijent varijacije 13,32%) nego u visini (koeficijent iznosi 4,51%).

Kod merenja devojčica dobijeni su sledeći rezultati: prosečna visina iznosila je $\bar{x}$ = 134,9 *cm* sa odstupanjem od σ = 6,43 *cm*, dok je prosečna težina bila $\bar{x}$ = 29,7 *kg* sa odstupanjem od σ = 4,78 *kg*. Kada koeficijent varijacije izračunamo i na ovim podacima, dobijamo:

$$V_{visina-devojčice} = \frac{6,43}{134,9} \cdot 100 = 4,77\%$$

$$V_{težina-devojčice} = \frac{4,78}{29,7} \cdot 100 = 16,09\%$$

Dakle, devojčice u ovom istraživanju imaju veća odstupanja od njihove prosečne visine i težine nego dečaci.

Normalizovano (standardizovano) odstupanje je mera varijacije koji beleži algebarska odstupanja pojedinačnih vrednosti obeležja od aritmetičke sredine izražene u jedinicama standardne devijacije.

$$X : N(\mu, \sigma^2)$$

$$Z = \frac{x_i - \mu}{\sigma} N(0,1) \qquad [3.9]$$

Primer 5

Primenu normalizovanog odstupanje prikazaćemo na primeru razmatranja studenata na testovima iz statistike (Šišić, 2006: 108). Autor navodi da je prosečan broj bodova na prvom testu iz statistike iznosio 50, dok je prosečno odstupanje bilo 10. Na drugom testu studenti su u proseku postigli 90 boda, a standardna devijacija je iznosila 20. Jedan od studenata koji je polagao ove testove na prvom je postigao 62 boda, a na drugom 105 bodova. Šta možemo da zaključimo o uspehu ovog studenta na testovima primenom standardizovanog odstupanja?

Kako imamo dva testa, izračunaćemo dva standardizovana odstupanja koristeći obrazac [3.9].

$$\overline{X}_1 : N(50, 10^2)$$

$$z_1 = \frac{62 - 50}{10} = 1{,}2$$

$$\overline{X}_2 : N(90, 20^2)$$

$$z_2 = \frac{105 - 90}{20} = 0{,}75$$

Dobijeni rezultati nam pokazuju da je student u oba slučaja postigao iznadprosečan rezultat i da je bio uspešniji na prvom testu jer je odstupanje od proseka naviše 1,2 standardne devijacije, dok je na drugom ono iznosilo 0,75 standardnih devijacija.

3.2 MERENJE VARIJABILNOSTI KVALITATIVNIH PROMENLJIVIH

Kod nominalnih mernih skala nemoguće je odrediti aritmetičku sredinu ili medijanu, te se za merenje varijabilnosti ovih promenljivih koriste drugačiji pokazatelji. Neki od njih za izračunavanje varijabiliteta koriste odstupanja od modusa (Wilcox, 1973), dok su drugi bazirani na proporcijama (npr. Lieberson 1969). Razvijanje ove oblasti omogućilo je potpuniju deskripciju nominalnih podataka koji su u značajnoj meri prisutni u društvenim istraživanjima.

Indekse zasnovane na proporcijama možemo podeliti u dve grupe. Prva grupa služi za izračunavanje i poređenje varijabiliteta između grupa, dok se druga koristi za opis varijabiliteta unutar grupa. U svrhu poređenja varijabiliteta između grupa prikazaćemo Šenonov indeks i Libersonov indeks varijabiliteta između promenljivih, dok je za merenje varijabiliteta između modaliteta jedne promenljive objašnjen Libersonov indeks za varijabilitet unutar grupe.

Libersonov indeks varijabiliteta unutar grupe izračunava se na osnovu sledećeg obrasca:

$$A_u = 1 - \Sigma_{i=1}^{n} (p_i)^2 \qquad [3.10]$$

Primer 6

Prema rezultatima popisa iz 2011. godine zabeleženi su sledeći rezultati raspodele stanovništva prema veroispovesti (tabela 3.4) u tri grada Republike Srbije: Beogradu, Novom Sadu i Nišu. Na osnovu Libersonovog indeksa varijabiliteta izračunaćemo varijabilitet stanovništva prema veroispovesti u sva tri grada (tabela 3.4.1).[20]

Tabela 3.4 *Stanovništvo prema veroispovesti po gradovima (Beograd, Novi Sad i Niš), popis 2011.*

veroispovest	Grad Beograd	Grad Novi Sad	Grad Niš
	p_i	p_i	p_i
hrišćanska	0,9041	0,8849	0,9403
islamska	0,0192	0,0139	0,0096
judaistička	0,0002	0,0002	0,0000

[20] Zbog veoma malih proporcija, za njihovo prikazivanje koristićemo više decimalnih mesta nego u ranijim primerima u kojima smo brojeve zaokruživali na drugom decimalnom mestu.

istočnjačke veroispovesti	0,0003	0,0004	0,0003
ostale veroispovesti	0,0004	0,0005	0,0002
agnostici	0,0015	0,0014	0,0004
nisu vernici	0,0245	0,0270	0,0102
nisu se izjasnili	0,0331	0,0661	0,0253
nepoznato	0,0167	0,0056	0,0137
Ukupno	1,0000	1,0000	1,0000

Izvor: *Popis stanovništva, domaćinstva i stanova 2011. u Republici Srbiji. Stanovništvo. Veroispovest, maternji jezik i nacionalna pripadnost. Podaci po opštinama i gradovima* (2013). Beograd: Republički zavod za statistiku, 46–53

Tabela 3.4.1 *Radna tabela za izračunavanje Libersonovog indeksa varijabiliteta unutar grupa*

veroispovest	Grad Beograd		Grad Novi Sad		Grad Niš	
	p_i	$(p_i)^2$	p_i	$(p_i)^2$	p_i	$(p_i)^2$
hrišćanska	0,9041	0,817	0,8849	0,78304801	0,9403	0,88416409
islamska	0,0192	0,00037	0,0139	0,00019321	0,0096	0,00009216
judaistička	0,0002	0,00000004	0,0002	0,00000004	0,0000	0
istočnjačke veroispovesti	0,0003	0,00000009	0,0004	0,00000016	0,0003	0,00000009
ostale veroispovesti	0,0004	0,00000016	0,0005	0,00000025	0,0002	0,00000004
agnostici	0,0015	0,00000225	0,0014	0,00000196	0,0004	0,00000016
nisu vernici	0,0245	0,00060025	0,0270	0,000729	0,0102	0,00010404
nisu se izjasnili	0,0331	0,00109561	0,0661	0,00436921	0,0253	0,00064009
nepoznato	0,0167	0,00027889	0,0056	0,00003136	0,0137	0,00018769
Ukupno	1,0000	0,81934729	1,0000	0,7883732	1,0000	0,88518836

$$A_{Beograd} = 1 - 0{,}81934729 = 0{,}18$$

$$A_{Novi\ Sad} = 1 - 0{,}7883732 = 0{,}21$$

$$A_{Niš} = 1 - 0{,}88518836 = 0{,}11$$

Libersonov indeks varijabiliteta između grupa koristićemo za poređenje gradova.

$$A_i = 1 - \Sigma_{i=1}^{n} (p_{xi} \cdot p_{yi}) \qquad [3.11]$$

Tabela 3.4.2 *Radna tabela za izračunavanje Libersonovog indeksa varijabiliteta između grupa (Beograd – Novi Sad)*

veroispovest	Grad Beograd	Grad Novi Sad	
	p_{xi}	p_{yi}	$p_{xi} \cdot p_{yi}$
hrišćanska	0,9041	0,8849	0,80003809
islamska	0,0192	0,0139	0,00026688
judaistička	0,0002	0,0002	0,00000004
istočnjačke veroispovesti	0,0003	0,0004	0,00000012
ostale veroispovesti	0,0004	0,0005	0,00000020
agnostici	0,0015	0,0014	0,0000021
nisu vernici	0,0245	0,0270	0,0006615
nisu se izjasnili	0,0331	0,0661	0,00218791
nepoznato	0,0167	0,0056	0,00009352
Ukupno	1,0000	1,0000	0.80325036

$$A_{Beograd - Novi\ Sad} = 1 - 0{,}80325036 \approx 0{,}2$$

Na isti način dobijamo varijabilitet između Beograda i Niša i između Novog Sada i Niša.

$$A_{Beograd - Niš} = 1 - 0{,}85162644 \approx 0{,}15$$

$$A_{Novi\ Sad - Niš} = 1 - 0{,}83423014 \approx 0{,}17$$

Iako se radi o veoma sličnoj raspodeli stanovništva u sva tri navedena grada, na osnovu ovog pokazatelja možemo zaključiti na najveće razlike u raspodeli popisanih građana prema veroispovesti postoje između Beograda i Novog Sada, potom između Novog Sada i Niša, a da se stanovništvo najsličnije izjašnjavalo o religijskoj pripadnosti u Beogradu i Nišu.

Šenonov indeks (Tepavčević, Lužanin, 2006: 65) predstavlja kvantitativnu meru različitosti nominalnih obeležja. Može se izračunati pomoću relativnih (tj. proporcija) i apsolutnih frekvencija. Kada se za njegovo izračunavanje koriste proporcije, upotrebljavamo sledeći izraz:

$$H = -\Sigma_{i=1}^{s} p_i \log p_i \qquad [3.12]$$

gde je:

s – broj kategorija posmatranog obeležja
p_i – proporcija uzorka koja se nalazi u i-toj kategoriji.

Pomoću apsolutnih frekvencija izračunavamo ga na sledeći način:

$$\mathrm{H} = \frac{n \log n - \Sigma_{i=1}^{s} f_i \log f_i}{n} \qquad [3.12a]$$

Šenonov indeks dostiže svoju maksimalnu vrednost u slučaju da su sve proporcije kategorija posmatranog obeležja međusobno jednake, odnosno:

$$H_{max} = \log s$$

Međutim, približavanje maksimalnoj vrednosti ne znači i veću različitost varijabiliteta između promenljivih. Tumačenje različitosti je upravo suprotno, što je manja vrednost Šenonovog indeksa, to je veća različitost varijabiliteta uzoraka.

Primer 7

Na osnovu rezultata jednog empirijskog istraživanja[21] koje je sprovedeno 2009. godine na teritoriji AP Vojvodine, u narednoj tabeli prikazani su rezultati nacionalne i konfesionalne pripadnosti anketiranih građana. Na osnovu Šenonovog indeksa potrebno je izračunati i uporediti varijabilitet obe statističke serije.

Tabela 3.5 *Radna tabela za izračunavanje Šenonovog indeksa*

Nacionalna pripadnost ispitanika	f_i	$f_i \log f_i$	Konfesionalna pripadnost ispitanika	f_i	$f_i \log f_i$
srpska	394	1.022,63	pravoslavna	413	1.080,39
crnogorska	13	14,48	katolička	116	239,48
hrvatska	22	29,53	luteransko- -reformatorska	1	0
mađarska	90	175,88	slovačko- -evangelistička	11	11,46
slovačka	16	19,27	grko-katolička	4	2,41
rumunska	12	12,95	adventistička	1	0
rusinska	4	2,41	islamska	10	10
romska	10	10	judejska	2	0,60
albanska	2	0,60	ateista	25	34,95
jugoslovenska	17	20,92	neverujući	21	27,77
nešto drugo	30	44,31	nešto drugo	6	4,67
Ukupno	610	1.352,98	Ukupno	610	1.411,73

[21] Radi se o empirijskom istraživanju koje je Odsek za sociologiju Filozofskog fakulteta u Novom Sadu realizovao u okviru rada na višegodišnjem projektu (2005–2010. godine) „Multikulturalnost AP Vojvodine kao činilac regionalnog povezivanja u Jugoistočnoj i Centralnoj Evropi”, finansiranog od strane Pokrajinskog sekretarijata za nauku i tehnološki razvoj.

Varijabilitet promenljive „Nacionalna pripadnost stanovnika” iznosi:

$$H = \frac{1.699{,}05 - 1.352{,}92}{610} = 0{,}57$$

Varijabilitet promenljive „Konfesionalna pripadnost ispitanika” je:

$$H = \frac{1.699{,}05 - 1.411{,}73}{610} = 0{,}47$$

Prema rezultatima ovog indeksa raspored frekvencija u drugoj promenljivoj, konfesionalna pripadnost ispitanika, pokazuje veći varijabilitet nego raspored frekvencija u prvoj promenljivoj. Maksimalna vrednost indeksa u slučaju ovih podataka iznosi 1,04, jer je:

$$H_{max} = \log 11 = 1{,}04$$

Vrednost *s* iznosi 11 jer je upravo toliko modaliteta u svakoj promenljivoj.

ZADACI[22]

1. Prema podacima elektroprivrede, distribucija jednogodišnje potrošnje električne energije po potrošačima iznosila je:

Godišnja potrošnja u kWh	Broj potrošača
0–500	300
501–1000	248
1001–1500	480

[22] Zadaci se mogu preuzeti sa sajta Filozofskog fakulteta u Novom Sadu: http://ff.uns.ac.rs/studenti/oglasne%20table/studenti_oglasne_table_sociologija.html

1501–2000	355
2001–2500	123
2501–3000	98
Ukupno	1.604

a) Izračunajte prosečnu godišnju potrošnju električne energije po potrošaču.
b) Koliko je prosečno odstupanje od proseka (σ) i prosečno postotno odstupanje od proseka (V)?
c) Odredite vrednosti kvartila i medijane te interkvartilnu razliku i koeficijent interkvartilne razlike?

2. Na jednom ispitu mereno je vreme (u minutama) koje je bilo potrebno studentima da urade zadatke iz statistike. Dobijeni su sledeći rezultati:

Vreme	Broj studenata
15	3
20	5
25	8
30	12
35	15
40	18
45	17
50	7
60	2
Ukupno	87

a) Odredite prosečno vreme rešavanja zadataka.
b) Koliki su kvartili i medijana?
c) Izračunajte vrednost interval varijacije, interkvartilne razlike i koeficijenta interkvartilne razlike?

3. U naredne tri tabele prikazane su zarade (u dinarima) zaposlenih u tri preduzeća.

Preduzeće A

44.090	28.930	39.620	57.650	49.880
52.350	32.470	34.110	34.460	45.860
32.420	47.360	72.340	31.730	52.650
69.830	51.820	52.300	57.980	51.730
83.240	35.560	55.210	46.720	43.070

Preduzeće B

27.520	35.160	48.040	42.480	44.250	30.050	31.220
31.470	39.620	40.680	60.090	29.520	44.460	56.730
43.880	44.220	48.270	43.810	43.210	27.650	40.810

Preduzeće C

39.380	23.080	19.200	18.340	34.560	32.680	19.140	41.490	33.310
41.350	15.100	39.820	43.890	33.470	19.430	37.120	18.480	25.290
13.810	13.780	15.140	37.780	36.970	42.240	43.230	35.030	45.180
32.920	41.520	40.120	28.030	22.670	34.000	4.340	24.220	27.380

a) Izračunajte aritmetičku sredinu, modus i medijanu prikazanih serija. Odredite prvi i treći kvartil.
b) Odredite interval varijacije i interkvartilnu razliku.
c) Na osnovu box-plota prokomentarišite varijabilitet serija. Odredite primanja prve četvrtine zaposlenih i prve tri četvrtine radnika u svakom preduzeću.
d) Izračunajte koeficijent varijacije i prokomentarišite ga.

4. Od 5.000 otvorenih tekućih računa u jednoj banci, uzorkom od 25 računa zabeležena su sledeća prekoračenja raspoloživih sredstava (u hiljadama dinara):

0	34	22	25	145
5	56	15	15	33
8	7	54	9	48
13	19	28	76	19
4	105	0	65	8

a) Koliki je interval varijacije?
b) Odredite vrednost prvog i trećeg kvartila. Kolika je interkvartilna razlika i koeficijent interkvartilne razlike?
c) Odredite vrednost aritmetičke sredine i medijane.
d) Nacrtajte box-plot dijagram. Prokomentarišite varijabilitet i uočene ekstremne vrednosti.

5. U narednoj tabeli dati su podaci o umrlima usled samoubistva s obzirom na starost, 2011. godine, na teritoriji Republike Srbije.

starost	Ispod 15	15–24	25–34	35–44	45–54	55–64	65–74	75 i više
frekvencija	2	39	98	133	220	242	216	306

Izvor: *Demografska statistika u R. Srbiji, 2011*. Beograd: Republički zavod za statistiku, str. 187.

a) Odredite vrednost kvartila i medijane. Iskoristivši izračunate veličine, utvrdite vrednosti odgovarajućih mera disperzije.
b) Kolika je prosečna starost umrlih usled samoubistva 2011. godine?
c) Histogramom i box-plotom prikažite seriju podataka i prokomentarišite ih.

6. U jednom sociološkom istraživanju, anketiranim građanima je postavljeno pitanje „Šta je, po vašem mišljenju, regionalizacija?” i dobijeni su sledeći odgovori:

Oblik ekonomskog organizovanja države	509
Oblik teritorijalno-administrativnog organizovanja države	456
Oblik izražavanja i zaštite kulturnih posebnosti	100
Nešto drugo	45
Ne znam	142
Ukupno	1.253

Koristeći Šenonov indeks ispitajte varijabilitet dobijenih odgovora.

7. Na osnovu prethodnog istraživanja, raspodela ispitanika prema mestu u kojem žive prikazana je u narednoj tabeli:

Šta je, po vašem mišljenju, regionalizacija?	Mesto boravka		
	Subotica	Pančevo	Novi Sad
Oblik ekonomskog organizovanja države	52	56	100
Oblik teritorijalno-administrativnog organizovanja države	60	70	93
Oblik izražavanja i zaštite kulturnih posebnosti	13	11	19
Nešto drugo	10	6	10
Ne znam	28	8	30
Ukupno	163	151	252

Koristeći Libersonove indekse ispitajte varijabilitet unutar i između gradova.

8. Izračunajte standardne devijacije navedenih distribucija frekvencija i pomoću koeficijenta varijacije uporedite njihov varijabilitet.

interval	Frekvencije promenljive X	Frekvencije promenljive Y
10,1–20,0	23	32
21,1–30,0	36	39
31,1–40,0	54	61
41,1–50,0	43	56
51,1–60	36	40

9. Godišnja plata zaposlenog sa visokom stručnom spremom u jednom preduzeću iznosi 684.000,00 dinara. Standardna devijacija iznosi 5.800,00 dinara. Koristeći standardizovano odstupanje pokažite da li je opravdano tvrđenje da je osoba sa istom kvalifikacijom koja ima godišnju platu od 524.000,00 dinara diskriminisana?

10. U narednoj tabeli prikazani su rezultati testova 10 slučajno izabranih studenata iz tri predmeta koja su polagali u zimskom semestru tekuće školske godine.

Student	Predmet A	Predmet B	Predmet C
1	54	65	100
2	67	44	60
3	87	89	53
4	34	91	88
5	28	34	37
6	66	25	49
7	86	11	23
8	19	34	16
9	97	22	62
10	45	71	38

a) Rangirajte kandidate prema uspehu na testovima i to prema ukupnom zbiru rezultata i standardizovanim vrednostima.
b) Kako izgleda rangiranje ako se uzme u obzir prosečni rezultat kandidata na svakom testu i prosečno odstupanje od proseka. Za predmet A uzet je ponder 0,34, za predmet B 0,24 dok ponder za predmet C iznosi 0,10?

MERE OBLIKA RASPOREDA

Mere asimetrije

Mere zaobljenosti

4. MERE OBLIKA RASPOREDA

4.1 MERE ASIMETRIJE

Deskriptivna analiza statističkih serija obuhvata i merenje načina rasporeda članova osnovnog skupa prema nekoj vrednosti (obično je to aritmetička sredina, ali može biti i medijana ili modus), odnosno prema osi simetrije. S obzirom na ovakva posmatranja, statističke serije mogu imati simetričan, pozitivan ili negativno asimetričan razmeštaj. U simetričnom razmeštaju svakom odstupanju vrednosti numeričke promenljive od aritmetičke sredine negativnog predznaka odgovara jednako odstupanje pozitivnog predznaka. U pozitivno asimetričnom rasporedu pozitivna odstupanja će prevladati negativna i obrnuto, u negativno asimetričnom slučaju pojavljuje se više odstupanja sa negativnim predznakom.

4.1.1 Raspored članova skupa prema aritmetičkoj sredini

Kada smo objašnjavali osobine aritmetičke sredine, pokazali smo da je suma odstupanja svih elemenata serije od aritmetičke sredine jednaka nuli. Zbog toga se ona ne može upotrebljavati za merenje asimetrije. S druge strane, aritmetička sredina kvadrata odstupanja vrednosti serije od sredine je varijansa (tj. mera disperzije), uvek je pozitivna veličina, te ni ona nije prikladna za ovu vrstu merenja. Stoga se za merenje asimetrije uzima treći centralni momenat ili aritmetička sredina odstupanja vrednosti od sredine podignute na treći stepen.

Shodno prethodnom određenju, za niz od *n* negrupisanih vrednosti numeričke promenljive *x*, treći centralni momenat dobijamo na osnovu formule:

$$\mu^3 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{n} \quad [4.1]$$

Za grupisane podatke, treći centralni momenat izračunavamo na sledeći način:

$$\mu^3 = \frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^3}{n} \quad [4.1a]$$

Pri simetričnom rasporedu, treći centralni momenat oko sredine jednak je nuli, za pozitivno asimetrične rasporede je veći od nule, dok na negativno asimetričan raspored upućuje treći centralni momenat oko sredine koji je manji od nule.

Potrebno je naglasiti da je ova mera simetrije apsolutni pokazatelj, tj. zavisi od mernih jedinica promenljive *x*, te je na osnovu njega otežano direktno zaključivanje o smeru i jačini asimetrije.

Da bi se uklonio uticaj mernih jedinica kod trećeg centralnog momenta oko sredine definisana je standardizovana mera smera i veličine asimetrije, tj. koeficijent asimetrije. Pored toga što je to relativna mera, on uzima u obzir sva odstupanja vrednosti numeričke promenljive od aritmetičke sredine pa možemo reći da je to i potpuna mera simetrije. Koeficijent asimetrije određen je izrazom:

$$\alpha_3 = \frac{\mu_3}{\sigma^3} \quad [4.2]$$

Dobija vrednost 0 ukoliko je reč o simetričnim serijama. Na pozitivno asimetričnu distribuciju upućuje vrednost koeficijenta asimetrije veći od nule, dok serija sa negativno asimetričnim rasporedom ima negativnu vrednost ovog koeficijenta.

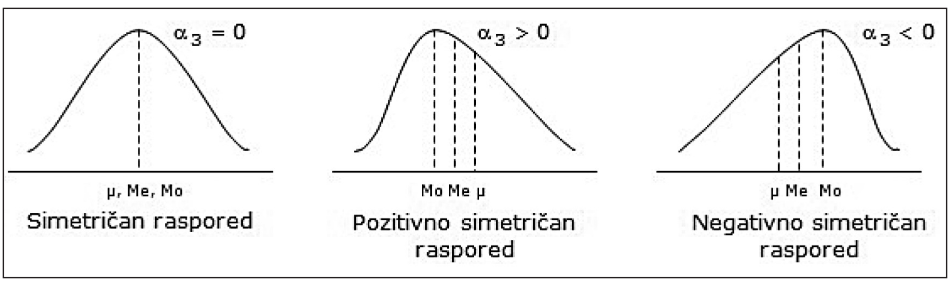
Tumačenje jačine asimetrije izraženo u apsolutnim vrednostima koeficijenta odvija se prema dole utvrđenom pravilu:

$$|\alpha_3| \leq 0{,}25; \textit{ mala asimetrija}$$
$$0{,}25 < |\alpha_3| \leq 0{,}50; \textit{ srednja asimetrija}$$
$$|\alpha_3| > 0{,}50; \textit{ jaka asimetrija}$$

4.1.2 Raspored članova skupa prema modusu i medijani

U svrhu merenja rasporeda članova osnovnog skupa pomoću modusa i medijane, definisane su Pirsonova i Bovlijeva mera asimetrije. Pirsonova mera asimetrije temelji se na odnosima aritmetičke sredine, modusa i medijane. U unimodalnoj seriji, simetričan raspored zahteva jednake vrednosti ova tri pokazatelja centralnosti ($\mu = M_e = M_o$). Kod pozitivne asimetrije, vrednost aritmetičke sredine je najveća, zatim dolazi medijana, dok je modus najmanja veličina ($M_o < M_e < \mu$). Na negativnu asimetriju jedne serije upućuje najveća vrednost modusa, manja vrednost medijane i najmanja aritmetičku sredinu ($M_o > M_e > \mu$).

Grafikon 4.1 Prikaz simetrične, pozitivno i negativno asimetrične distribucije



Polazeći od ovako definisanih pravila za merenje asimetrije, u istu svrhu možemo iskoristiti i razliku aritmetičke sredine i modusa, ali i aritmetičke sredine i medijane. Za simetrične serije ove razlike su jednake nuli, dok su kod negativno asimetričnih rasporeda one manje od nule.

Pirsonova mera asimetrije određena je aritmetičkom sredinom i modusom i definiše se izrazom:

$$S_k = \frac{\bar{x} - M_o}{\sigma} \qquad [4.3]$$

Ukoliko koristimo razliku medijane od aritmetičke sredine, Pirsonov koeficijent asimetrije odredićemo na sledeći način:

$$S_k = \frac{3\,(\bar{x} - M_e)}{\sigma} \qquad [4.3a]$$

Iz navedenih izraza možemo videti da je Pirsonova mera standardizovano odstupanje modusa (ili medijane) od aritmetičke sredine. Ova mera se kreće u intervalu ±3 u slučaju simetričnih serija.

Kod umereno asimetričnih serija udaljenost modusa od aritmetičke sredine približno je jednaka trostrukoj udaljenosti medijane od aritmetičke sredine, tj.:

$$\left| \bar{x} - M_o \right| \approx 3 \left| \bar{x} - M_e \right| \qquad [4.4]$$

Tako je na osnovu prethodna dva izraza, definisana alternativna Pirsonova mera asimetrije koja glasi:

$$S_k = \frac{3\,(\bar{x} - M_e)}{\sigma} \qquad [4.5]$$

Međutim, za razliku od koeficijenta asimetrije, Pirsonova mera asimetrije je nepotpuna mera. Kao što se može primetiti iz formula koje ga definišu, ova mera je određena ograničenim brojem informacija iz posmatrane promenljive. Osim što je nepotpuna, može biti i netačna. Postoje serije kod kojih se modus ne može odrediti, ponekad je numerički nestabilan, a i vrednost medijane se u pojedinim slučajevima mora aproksimirati interpolacijom. Zbog različitih temelja na osnovu kojih su definisani, koeficijent asimetrije i Pirsonova mera asimetrije nisu uporedive veličine.

Bovlijeva mera je određena na osnovu odnosa kvartila i medijane. Na osnovu nje, simetričnu distribuciju ćemo prepoznati po jednakoj udaljenosti donjeg i gornjeg kvartila od medijane:

$$M_e - Q_1 = Q_3 - M_e \tag{4.6}$$

Na osnovu izraza [4.6], simetričnost distribucije možemo opisati i na sledeći način:

$$Q_1 + Q_3 - 2M_e = 0$$

Kada su gore naznačene razlike veće od nule, govorimo o pozitivno asimetričnim distribucijama, tj.:

$$Q_1 + Q_3 - 2M_e > 0$$

Pozitivnu asimetriju možemo odrediti i ako važi sledeći odnos:

$$M_e - Q_1 < Q_3 - M_e$$

Za negativno asimetrične serije važi:

$$Q_1 + Q_3 - 2M_e < 0$$

ili

$$M_e - Q_1 > Q_3 - M_e$$

Kako su razlike kvartila i medijane izražene u mernim jedinicama, njihovo eliminisanje postižemo deljenjem sa interkvartilnom razlikom i tako dolazimo do izraza za Bovlijevu meru asimetrije:

$$B_o = \frac{Q_1 + Q_3 - 2M_e}{Q_3 - Q_1} \qquad [4.7]$$

Ovaj izraz može da poprimi vrednost iz zatvorenog intervala od ±1. Ako pažnju usmerimo na njegovu ocenu, ona je slična kao i ocena Pirsonove mere asimetrije. U ovom slučaju, nepreciznost Bovlijeve mere proizilazi iz kvartila jer se njima eliminiše uticaj prve i poslednje četvrtine podataka.

Primer 1

U narednoj tabeli (4.1) prikazana su merenja 20 slučajno izabranih proizvoda radi kontrole deklarisane težine. Pomoću koeficijenta asimetrije i Bovlijeve mere prikazaćemo simetričnost/asimetričnost dole navedene distribucije podataka.

Tabela 4.1 *Težine (u gr) slučajno odabranih proizvoda*

102	104	99	100	101
103	97	103	102	106
94	99	95	97	102
98	98	94	105	103

Tabela 4.1a *Radna tabela za izračunavanje koeficijenta asimetrije i Bovlijeve mere*

x_i	f_i	kumu-lativ	$x_i f_i$	$(x_i-\bar{x})^2$	$(x_i-\bar{x})^3$	$f_i(x_i-\bar{x})^2$	$f_i(x_i-\bar{x})^3$
94	2	2	188	37,21	–226,981	74,42	–453,962
95	1	3	95	26,01	–132,651	26,01	–132,651
97	2	5	194	9,61	–29,791	19,22	–59,582
98	2	7	196	4,41	–9,261	8,82	–18,522
99	2	9	198	1,21	–1,331	2,42	–2,662
100	1	10	100	0,01	–0,001	0,01	–0,001
101	1	11	101	0,81	0,729	0,81	0,729
102	3	14	306	3,61	6,859	10,83	20,577
103	3	17	309	8,41	24,389	25,23	73,167
104	1	18	104	15,21	59,319	15,21	59,319
105	1	19	105	24,01	117,649	24,01	117,649
106	1	20	106	34,81	205,379	34,81	205,379
Ukupno	20		2.002	–	–	241,80	–191,288

Na osnovu obrasca za izračunavanje ponderisane aritmetičke sredine, prosečna težina 20 slučajno izabranih proizvoda iznosila je 100,1 grama.

$$\bar{x} = \frac{2.002}{20} = 100{,}1 \; grama$$

Informacija o prosečnoj težini potrebna nam je da bismo mogli da izračunamo treći centralni momenat, ali i standardnu devijaciju podignutu na treći stepen, dakle, osnovne elemente obrasca koeficijenta asimetrije.

$$\mu_3 = \frac{\Sigma_{i=1}^{n} f_i\,(x_i-\bar{x})^3}{n} = \frac{-191{,}288}{20} = -9{,}5644$$

Na osnovu vrednosti trećeg centralnog momenta, koja je u ovom slučaju manja od nule, vidimo da se radi o negativno asimetričnoj seriji.

$$\sigma^2 = \frac{\Sigma_{i=1}^{k} f_i (x_i - \bar{x})^3}{\Sigma_{i=1}^{k} f_i - 1} = \frac{241{,}80}{19} = 12{,}73$$

$$\sigma = \sqrt{12{,}73} = 3{,}57$$

$$\sigma^3 = 3{,}57^3 = 45{,}50$$

Vrednosti trećeg centralnog momenta i standardne devijacije podignute na treći stepen uvrstićemo u obrazac [4.2].

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{-9{,}5644}{45{,}50} = -0{,}21$$

Dobijena vrednost koeficijenta asimetrije iznosi –0,21 pa možemo zaključiti da je gore navedena distribucija malo asimetrična, jer je $|0{,}21| \leq 0{,}25$, kao i da je reč o negativnoj asimetriji $-0{,}21 < 0$.

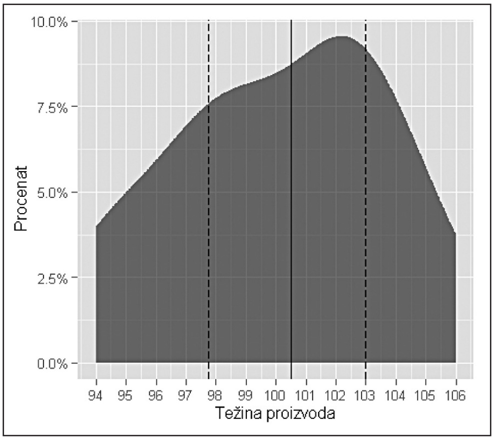
Kako možemo videti iz tabele 4.1a posmatrana serija nije unimodalna, te se za nju ne može izračunati Pirsonova mera asimetrije. Postoje dva obeležja (102 i 103) sa najvećom frekvencijom (3), pa govorimo o bimodalnoj seriji podataka.

Kao što smo videli iz izraza [4.7], za izračunavanje Bovlijeve mere asimetrije potrebno je najpre odrediti kvartile i medijanu. Prvi kvartil iznosi 97,5, drugi (medijana) 100,5, dok se treći kvartil u posmatranoj seriji nalazi kod obeležja 103. Prvi kvartil dobijen je pomoću formule [2.17], treći smo izračunali na osnovu obrasca [2.17a], dok smo za izračunavanje medijane koristili izraz [2.14].

$$B_o = \frac{97{,}5 + 103 - 2 \cdot 100{,}5}{103 - 97{,}5} = \frac{-0{,}5}{5{,}5} = -0{,}09$$

Izračunata vrednost Bovlijeve mere asimetrije manja je od nule, i u ovom slučaju vidimo da je serija blago negativno asimetrična. Njen grafički prikaz prikazan je grafikonom 4.2 u kojem isprekidane linije označavaju položaje prvog i trećeg kvartila, dok je punom linijom označena medijana.

Grafikom 4.2 *Težine (u gr) slučajno odabranih proizvoda – asimetrija*



Primer 2

U narednoj tabeli prikazana je raspodela živorođenih prema starosti majke u Republici Srbiji, 2010. godine. Koristeći prezentovane mere asimetrije analiziraćemo navedenu statističku seriju.

Tabela 4.2 *Živorođeni prema starosti majke u Republici Srbiji, 2010. godine*[23]

starost majke	živorođeni
15–19	4.109
20–24	14.765
25–29	22.821

[23] U navedenoj seriji izostavljeni su podaci o broju živorođenih od majki starih ispod 15 i 50 i više godina.

30–34	17.728
35–39	7.236
40–44	1.210
45–49	88
Ukupno	67.957

Izvor: *Demografska statistika u Republici Srbiji 2011* (2011). Beograd: Republički zavod za statistiku, str. 59.

I ovog puta konstruisaćemo radnu tabelu da bismo lakše odredili mere asimetrije.

Tabela 4.2a *Radna tabela za izračunavanje mera asimetrije*

x_i	f_i	x'_i	$x'_i f_i$	kumulacija	$f_i\,(x'_i - \bar{x})^2$	$f_i\,(x'_i - \bar{x})^3$
15–19	4.109	17	69.853	4.109	494.480,76	–5.424.453,92
20–24	14.765	22	324.830	18.874	526.237,89	–3.141.640,192
25–29	22.821	27	616.167	41.695	21.472,28	–20.828,11
30–34	17.728	32	567.296	59.423	287.918,68	1.160.312,26
35–39	7.236	37	267.732	66.659	590.029,958	5.327.970,47
40–44	1.210	42	50.820	67.869	238.177,49	3.341.630,17
45–49	88	47	4.136	67.957	31.868,40	606.455,64
Ukupno	67.957	–	1.900.834	–	2.190.185,44	1.849.446,32

Prosečna starost majke pri rađanju u Republici Srbiji za 2010. godinu iznosila je 27,97 godina. Ona nam je potrebna zbog izračunavanja trećeg centralnog momenta i standardne devijacije podignute na treći stepen.

$$\mu_3 = \frac{1.849.449,32}{67.957} = 27,215$$

$$\sigma^2 = \frac{2.190.185,44}{67.957} = 32,23$$

$$\sigma = \sqrt{32,23} = 5,68$$

$$\sigma^3 = 5,68^3 = 182,97$$

Sada možemo izračunati vrednost koeficijanta asimetrije.

$$\alpha_3 = \frac{27,215}{182,97} = 0,15$$

On u ovom slučaju poprima blagu pozitivnu vrednost (0,15), te upućuje na pozitivno asimetričnu seriju.

Statistička serija je u ovom slučaju unimodalna, te je prihvatljivo izračunavanje Pirsonove mere asimetrije. Već imamo određenu aritmetičku sredinu i standardnu devijaciju, jedino nam je potrebno da odredimo medijanu. Kako se radi o intervalnoj neparnoj seriji, upotrebićemo izraz [2.13].

$$M_e = 25 + \frac{29 - 25}{41.695\ 18.874} \cdot \left(\frac{67.957}{2} - 18.874\right) = 27,65$$

Pirsonova mera je, kao što smo već objasnili, zasnovana na drugim osnovama, ali i njena vrednost upućuje na isti zaključak. Pomeranje u odnosu na koeficijent asimetrije je neznatan, a nastao je usled toga što se radi o nepotpunoj meri asimetrije.

$$S_k = \frac{3\,(27{,}97 - 27{,}65)}{5{,}68} = 0{,}17$$

Da bismo analizu priveli kraju ostalo je da odredimo prvi i treći kvartil i izračunamo Bovlijevu meru. Njena vrednost najviše odstupa od koeficijenta asimetrije (0,33), jer je bazirana samo na osnovu pozicionih mera centralne tendencije.

$$Q_1 = 20 + \frac{\dfrac{67.957}{4} - 4.109}{14.756} \cdot 5 = 24{,}36$$

$$Q_3 = 30 + \frac{\dfrac{3 \cdot 67.957}{4} - 41.695}{17.728} \cdot 5 = 32{,}62$$

$$B_o = \frac{24{,}36 + 32{,}62 - 2 \cdot 27{,}65}{32{,}62 - 27{,}65} = 0{,}33$$

4.2 MERE ZAOBLJENOSTI

Zaobljenost modalnog vrha distribucije meri se koeficijentom zaobljenosti. Obeležava se sa α_4 i predstavlja odnos četvrtog centralnog momenta i standardne devijacije podignute na četvrti stepen. Alternativno, u imeniocu razlomka može stajati i drugi centralni momenat (varijansa) podignut na drugi stepen:

$$\alpha_4 = \frac{\mu_4}{\sigma_4} = \frac{\mu_4}{\mu_2^2} \qquad [4.8]$$

Kako je četvrti centralni momenat prosečno odstupanje vrednosti numeričke promenljive od njene aritmetičke sredine podignute na četvrti stepen, kod negrupisanih podataka dobijamo ga na sledeći način:

$$\mu_4 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^4}{N} \qquad [4.9]$$

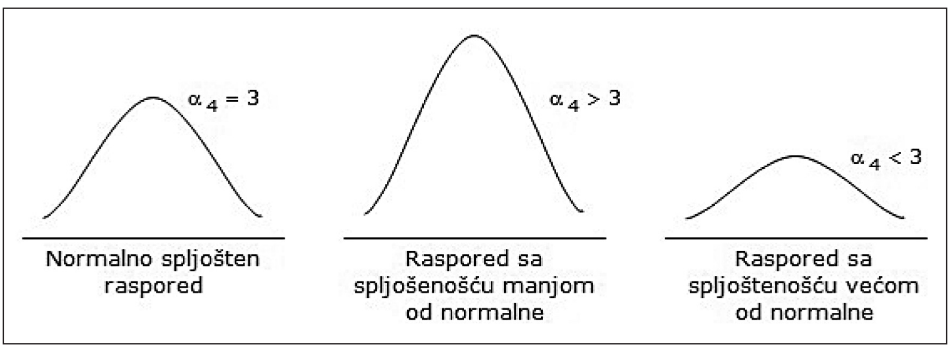
Za distribuciju frekvencija koristićemo sledeći izraz:

$$\mu_4 = \frac{\sum_{i=1}^{N} f_i (x_i - \bar{x})^4}{N} \qquad [4.9a]$$

U slučaju normalne distribucije, koeficijent zaobljenosti iznosi 3, ukoliko je veći od tri tumačimo da je distribucija više izdužena u odnosu na normalnu, dok koeficijent zaobljenosti manji od tri ukazuje na distribuciju koja je više spljoštena u odnosu na normalnu. Normalna distribucija je najvažnija teorijska distribucija u statistici. Ona je osnova od koje se polazi prilikom testiranja statističkih hipoteza. Međutim, ovom prilikom nećemo ulaziti u dublje razmatranje njene osobenosti i svrhe. Za deskriptivnu statistiku – kojoj je ovaj udžbenik i namenjen – dovoljno je da pokažemo pokazatelje koji je određuju.

Mogući oblici spljoštenosti prikazani su u narednom grafikonu (4.2).

Grafikon 4.2 *Oblici spljoštenosti*



Primer 3

Izračunavanje koeficijenta zaobljenosti prezentovaćemo na podacima iz tabele 4.1.

Tabela 4.1b *Radna tabela za izračunavanje koeficijenta zaobljenosti*

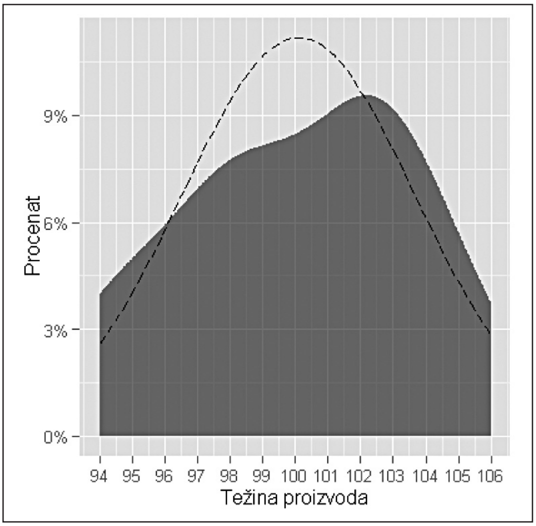
x_i	f_i	$f_i\,(x_i - \bar{x})^4$
94	2	2.769,17
95	1	676,52
97	2	184,70
98	2	38,90
99	2	2,93
100	1	0,00
101	1	0,66
102	3	39,10
103	3	212,18
104	1	231,34
105	1	576,48
106	1	1.211,74
Ukupno	20	5.943.71

$$\mu_4 = \frac{5.943,71}{20} = 297,19$$

$$\alpha_4 = \frac{297,19}{162,05} = 1,83$$

Kako je 1,83 < 3 zaključujemo da je serija više spljoštena u odnosu na normalnu (grafikon 4.3)

Grafikon 4.3 *Težine (u gr) slučajno odabranih proizvoda – spljoštenost*



ZADACI[24]

1. Zadata je sledeća distribucija frekvencija:

Intervali	15–19	20–24	25–29	30–34	35–39	40–44
Frekvencije	96	271	503	483	175	39

a) Kolika je vrednost koeficijenta zaobljenosti distribucije?
b) Izračunajte vrednosti koeficijenta asimetrije, Pirsonove i Bovlijeve mere asimetrije.
c) Prikažite distribuciju grafički i označite položaj aritmetičke sredine, modusa i medijane.

[24] Zadaci se mogu preuzeti i sa sajta Filozofskog fakulteta u Novom Sadu: http://ff.uns.ac.rs/studenti/oglasne%20table/studenti_oglasne_table_sociologija.html

2. Prvi kvartil jedne distribucije frekvencija iznosi 4243,24, drugi kvartil 5923,43, a treći kvartil 7760,68.[25]

   a) Odredite vrednost odgovarajuće mere asimetrije.
   b) Koristeći se datim veličinama, izračunajte vrednost odgovarajućeg pokazatelja disperzije.

3. U jednoj distribuciji izračunata je aritmetička sredina 2422, standardna devijacija 2383, prvi kvartil 683, medijana 1766 i treći kvartil 3370.

   a) Izračunajte vrednost koeficijenta asimetrije.
   b) Odredite vrednost Bovlijeve i Pirsonove mere.
   c) Prikažite navedenu distribuciju grafički i protumačite ga.

4. Tokom 20 radnih dana zabeležen je sledeći broj kupaca u jednoj trgovini.

45	74	89	89	23
23	56	84	31	63
65	53	98	45	31
72	65	78	24	32

   a) Merama asimetrije i zaobljenosti opišite dati niz.
   b) Nacrtajte tačkasti dijagram i na njemu odredite aritmetičku sredinu, modus i medijanu.

5. U sledećem primeru prikazano je šest unimodalnih distribucija. U jednoj od njih razlika između medijane i donjeg kvartila je manja od razlike gornjeg kvartila i medijane. Odredite u kojoj od prikazanih distribu-

[25] Podaci za drugi, četvrti i peti primer preuzeti su iz knjige: Šošić, 2006:121.

cija vrednost aritmetičke sredine, medijane i modusa zadovoljava gore opisan odnos medijane i kvartila.

aritmetička sredina	medijana	modus
24	22	25
24	25	22
22	24	25
22	25	24
25	24	22
25	22	24

LITERATURA

Angel R. Allen, Christine D. Abbot and Dennis C. Runde (2005). *A Survey of Mathematics with Applications*. Pearson Education.

Demografska statistika 2002 i 2003 (2006). Beograd: Republički zavod za statistiku.

Demografska statistika 2008 (2009). Beograd: Republički zavod za statistiku.

Demografska statistika 2009 (2010). Beograd: Republički zavod za statistiku.

Demografska statistika u Republici Srbiji 2011 (2011). Beograd: Republički zavod za statistiku.

Hegediš Antal i Katarina Čobanović (1991). *Demografska i agrarna statistika Vojvodine 1767–1867*. Novi Sad: Institut za istoriju, Institut za ekonomiku poljoprivrede i sociologiju sela.

Lieberson Stanley (1969). Measuring Population Diversity. *American Sociological Review* 34(6): 850-862.

Mann S. Prem (2009). *Uvod u statistiku*. Beograd: Ekonomski fakultet.

Opštine i regioni u Republici Srbiji 2011 (2011). Beograd: Republički zavod za statistiku.

Petz Boris (2007). *Osnovne statističke metode za nematematičare*. Zagreb: Naklada Slap.

Petz Boris, Vladimir Kolesarić i Dragutin Ivanec (2012). *Petzova statistika. Osnovne statističke metode za nematematičare*. Zagreb: Naklada Slap.

Popis stanovništva, domaćinstva i stanova u 2002. Stanovnoštvo. Školska sprema i pismenost. Podaci po opštinama (2003). Beograd: Republički zavod za statistiku.

Popis stanovništva, domaćinstva i stanova u 2002. Stanovništvo. Veroispovest, maternji jezik i nacionalna ili etnička pripadnost prema starosti i polu. (2003). Beograd: Republički zavod za statistiku.

Popis stanovništva, domaćinstva i stanova u 2002. Stanovništvo. Žensko stanovništvo staro 15 i više godina. Podaci po opštinama (2004). Beograd: Republički zavod za statistiku.

Popis stanovništva, domaćinstva i stanova 2011. u Republici Srbiji. Fertilitet ženskog stanovništva. Podaci po opštinama i gradovima (2013). Beograd: Republički zavod za statistiku.

Popis stanovništva, domaćinstva i stanova 2011. u Republici Srbiji. Starost i pol. Podaci po naseljima. (2012). Beograd: Republički zavod za statistiku.

Popis stanovništva, domaćinstva i stanova u 2011. Stanovništvo. Veroispovest, maternji jezik i nacionalna ili etnička pripadnost prema starosti i polu. Podaci po opštinama i gradovima (2013). Beograd: Republički zavod za statistiku.

Statistički godišnjak Republike Srbije 2011 (2011). Beograd: Republički zavod za statistiku.

Stigler, M. Stephen (1986). *The history of statistics*. Cambridge: The Belknap press of Harvard university press.

Stojković, Milutin (2001). *Statistika*. Subotica: Ekonomski fakultet.

Tepavčević, Andreja i Zorana Lužanin (2006). *Matematičke metode u taksonomiji*. Novi Sad: Prirodno matematički fakultet, Departman za matematiku i informatiku.

Šošić, Ivan i Vladimir Serdar (2002). *Uvod u statistiku*. Zagreb: Školska knjiga.

Šošić, Ivan (2006). *Primijenjena statistika*. Zagreb: Školska knjiga.

Wilcox, R. Allen (1973). Indices of Qualitative Variation and Political Measurement. *The Western Political Quarterly* 26(2): 325–343.

Willcox, F. Walter (1934). Note on the Chronology of Statistical Societies. *Journal of the American Statistical Association* 29(188): 418–420.

Zarkovich, Slobodan (1990). The Beginning of Statistics in Yugoslavia. *International Statistical Review* 58(1): 19–28.

Žene i muškarci u Srbiji (2008). Beograd: Republički zavod za statistiku.

Žižić, Mileva, Miodrag Lovrić i Dubravka Pavličić (2007). *Metodi statističke analize*. Beograd: Ekonomski fakultet.

Korišćeni internet izvori:
www.nsz.gov.rs
www.stat.gov.rs