

STATISTIKA I ANALIZA

SKRIPTA

dr Milica Cvetković



dr Milica Cvetković

Statistika i analiza

Skripta

Niš, 2022.

dr Milica Cvetković
Statistika i analiza-Skripta
ISBN-978-86-81912-17-1

Skripta je urađena u okviru projekta "Primena koncepta zelenih praksi u razvoju programskih sadržaja – GREENP EDU", Ministarstva prosvete, nauke i tehnološkog razvoja republike Srbije -programska aktivnost Razvoj visokog obrazovanja (2021-22).

SADRŽAJ

PREDGOVOR	1
POJAM I PREDMET STATISTIKE	4
ISTORIJAT STATISTIKE	4
DEFINICIJA I PODELA STATISTIKE	4
OSNOVNI STATISTIČKI POJMOVI	6
OSNOVNI KORACI STATISTIČKOG ISTRAŽIVANJA	8
VRSTE STATISTIČKIH VARIJABLI	16
METODE OPISIVANJA KVALITATIVNIH PODATAKA	18
FREKVENCIJE I RELATIVNE FREKVENCIJE	18
FREKVENCIJE I GRAFIČKI PRIKAZ U SOFTVERU <i>STATISTICA</i>	19
KUMULATIVNA FREKVENCIJA	23
METODE OPISIVANJA KVANTITATIVNIH PODATAKA	26
GRUPISANJE KVANTITATIVNIH PODATAKA RAČUNSKI	28
GRUPISANJE KVANTITATIVNIH PODATAKA U SOFTVERU <i>STATISTICA</i>	31
MERE POPULACIJE I UZORKA	35
MERE CENTRALNE TENDENCIJE	36

IZRAČUNATE SREDNJE VREDNOSTI: ARITMETIČKA SREDINA	37
IZRAČUNATE SREDNJE VREDNOSTI: HARMONIJSKA SREDINA	45
IZRAČUNATE SREDNJE VREDNOSTI: GEOMETRIJSKA SREDINA	46
POZICIONE SREDNJE VREDNOSTI: MODUS I MEDIJANA.....	49
MERE DISPERZIJE.....	58
APSOLUTNE MERE DISPERZIJE	59
RELATIVNE MERE DISPERZIJE	64
MERE OBLIKA	67
MERE ASIMETRIJE	68
MERE SPLJOŠTENOSTI	72
ANKETA I OBRADA PODATAKA.....	75
PRIMENA KONCEPTA ZELENIH PRAKSI – REZULTATI ANKETE	77
LITERATURA	82

PREDGOVOR

Veliki razvoj računarske tehnike u drugoj polovini 20. veka uslovio je nagli skok u primeni statističkih metoda u svim granama nauke, kao i u analizama svih masovnih pojava.

Kvantitativni pristup problemima, odabir odgovarajuće statističke metode, statistička obrada podataka, statistička analiza i zaključci kao i ispitivanje javnog mnjenja, od velikog su značaja za ispitivanje mnogih sfera, pa i u oblasti zaštite životne sredine, u monitoringu sistema za upravljanje otpadom, ekonomske analize u tretiranju otpada, u ispitivanju reciklabilnih materijala i sl.

Ova skripta ima za cilj da, sa tehničko-tehnološkog aspekta, predstavi koristan alat, za obradu podataka, analizu, predikciju, kao i mere unapređenja sistema za upravljanje otpadom i zaštitu životne sredine, studentima iz ovih oblasti koji ga mogu koristiti u svim stručnim i stručno-aplikativnim predmetima.

Skripta, pored osnovnih teorijskih prikaza metoda obrada podataka, sadrži i praktičnu primenu vežbi u softveru *Statistica* na konkretnim primerima iz oblasti zaštite životne sredine i upravljanju otpadom.

Na ovaj način će studentima biti približen koncept kreiranja vodiča za Zelene prakse u preduzećima tako što će pratiti i sakupljati podatke iz dela reciklaže, analize životnog ciklusa i merenja i kontrole parametara životne sredine, dobijene podatke obrađivati i analizirati u okviru izabranog softverskog

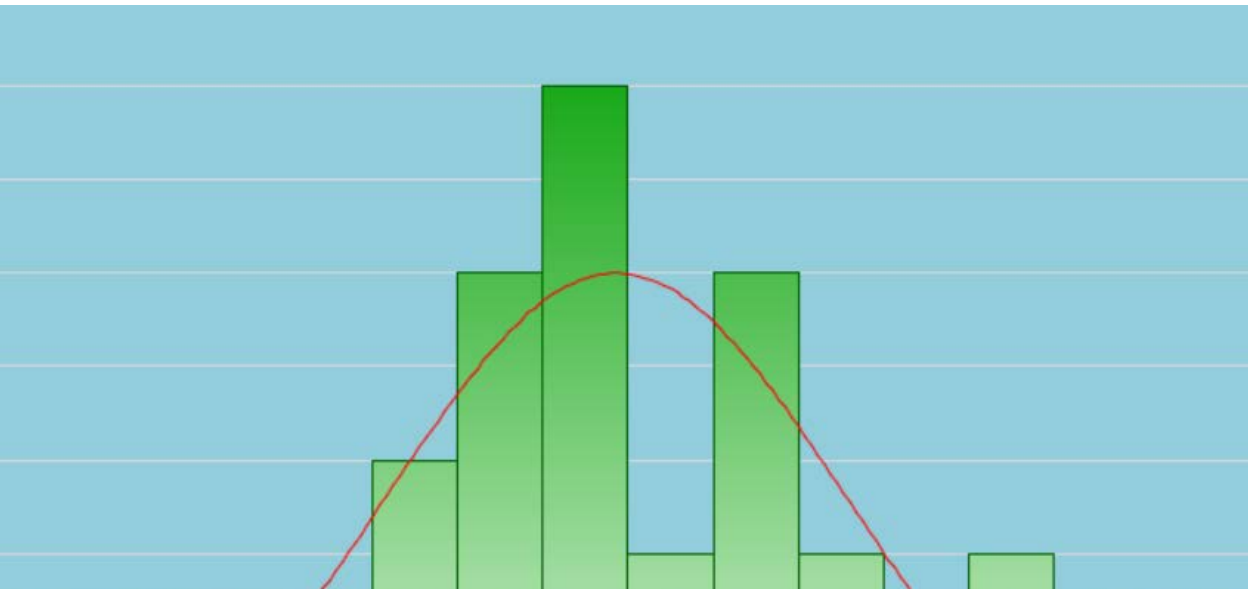
paketa za monitoring životne sredine.

Takođe, u skripti su prikazani rezultati obrađenih podataka ankete u ispitivanju javnog mnjenja o primeni koncepta zelenih praksi, odnosno, o sprovođenja mera u oblasti sprečavanja klimatskih promena i zagađenja, razvoja energije i biodiverziteta, mobilnosti i cirkularne ekonomije, održive poljoprivrede i proizvodnje hrane.

Skripta je namenjena pre svega studentima koji izučavaju inženjerstvo zaštite životne sredine, kao i upravljanje otpadom. Obzirom da je predstavljen koristan alat za obradu i analizu podataka, mogu je koristiti i studenti drugih studijskih programa tehničko-tehnološkog polja.

Skripta je nastala kao rezultat rada autora 2021. i 2022. godine, na aktivnostima u projektu “Primena koncepta zelenih praksi u razvoju programskih sadržaja” (GREENP-EDU), koji se finansira od strane Ministarstva prosvete, nauke i tehnološkog razvoja RS, u okviru programske aktivnosti “Razvoj visokog obrazovanja”.

Autor



POJAM I PREDMET STATISTIKE

Istorijat statistike

Prapočeci statistike kao nauke ogledaju se u organizovanim prebrojavanjima stanovnika još 4000 god.p.n.e u Kini i 3000 god.p.n.e. u Egiptu. Potom su u srednjem veku u Evropi bili organizovani popisi stanovništva, stoke, rođenih, venčanih i umrlih. Kao naučna disciplina, statistika se formira u 17. veku istovremeno u Nemačkoj i Engleskoj. Zahvaljujući razvoju verovatnoće i matematičke analize u 19. veku, dolazi i do razvoja statističkih teorija. U drugoj polovini 20. veka, razvoj računarske tehnike uslovljava nagli skok u primenama statističkih metoda u svim granama nauke, kao i u analizama svih masovnih pojava.

Naziv statistika potiče od latinske reči **Status**, što znači stanje i italijanskog termina **Regione di stato**, što znači država, državni interes, i ukazuje da je država bila područje nastanka i formiranja statistike kao naučne discipline.

Definicija i podela statistike

Statistika je naučna disciplina koja se bavi prikupljanjem, analizom i tumačenjem podataka o masovnim pojavama sa ciljem da:

- Utvrdi zakonitosti masovnih pojava,

- Utvrđene zakonitosti kvantitativno izrazi,
- Utvrđene zakonitosti analizira.

Podela statistike data je sledećim dijagramom:



Podela statistike

Deskriptivna statistika je deo statistike koji se bavi prikupljanjem, obradom i povezivanjem podataka.

Statistička analiza je pribavljanje numeričkih informacija, kvalitativna interpretacija, donošenje zaključaka i formiranje zakonitosti ponašanja posmatranih pojava.

Statistička teorija predstavlja pronalazak novih statističkih metoda, njihovo objašnjavanje, dokazivanje i usavršavanje.

Značaj statistike je nekad bio metod i sredstvo za prikupljanje

najpouzdanijih informacija za vođenje političkog razvoja društva. Danas je statistika sastavni deo svakodnevnog života pojedinca, a naučnom radniku sredstvo do objektivnih činjenica u istraživanju.

Osnovni statistički pojmovi

Masovna pojava je svaka pojava koja se u vremenu i prostoru pojavljuje u velikom broju.

Statistički skup ili **populacija** je skup svih elemenata na kojima se izvesna pojava statistički posmatra. To je skup pojava koje zadovoljavaju neka svojstva (uslove) bitna za cilj analize, odn. sve pojave koje su predmet istraživanja.



Uzorak kao podskup populacije

U slučaju izuzetno masovnih pojava, kada nije moguće izvršiti obradu svih elemenata populacije, određenim metodama teorije verovatnoće vrši se odabir podskupa, uzorka.

Uzorkovanje nije proizvoljan postupak, već mora zadovoljiti određene uslove kako bi uzorak bio reprezentativan.

Reprezentativan uzorak je podskup populacije koji predstavlja karakteristike osnovnog skupa. Nad uzorkom se sprovodi dalja statistička analiza koja rezultira određenim kvantitativnim zaključcima koji se prepisuju čitavoj populaciji.



Veza verovatnoće i statistike

Statistička jedinica je pojedini element populacije ili uzorka ("case").

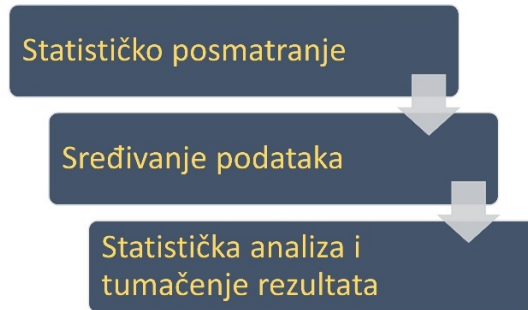
Statističko obeležje ili **varijabla** je svojstvo statističkih jedinica ("variable").

Varijabilitet je promenljivost obeležja od jedinice do

jedinice statističkog skupa (kategorije).

Statistički podatak je zapis o elementima statističkog skupa (broj, reč, simbol, znak,...).

Osnovni koraci statističkog istraživanja



Statističko posmatranje je prikupljanje podataka za celu populaciju ili uzorak.

Sređivanje podataka tabelarno i/ili grafički predstavlja razvrstavanje vrednosti varijabli u grupe (računski ili primenom programskih paketa).

Statistička analiza i tumačenje rezultata je utvrđivanje svojstva populacije, odnosno, dobijanje statističkih parametara.

Najbitnije u analizi podataka je da zaključke treba pregledno i sadržajno obrazložiti.

Temelj svake kvalitetne analize su kvalitetni podaci koji mogu biti:

- Primarni podaci – oni koji se prikupljaju neposredno za namenu pojedine analize;
- Sekundarni podaci – oni koji su već prikupljeni i mogu biti do izvesne mere obrađeni i na raspolaganju korisniku (besplatno ili uz nadoknadu).

U zavisnosti da li se obrada podataka vrši na čitavom statističkom skupu (populaciji) ili na uzorku, razlikujemo sledeće vrste statističke analize:

- Opisnu statističku analizu – područje statistike koje se bavi prikupljanjem i analizom podataka za populaciju u celini;
- Reprezentativnu statističku analizu - područje statistike koje se bavi uopštavanjem, predviđanjem i ocenjivanjem na osnovu nepotpunih podataka (uzorka).



Za svaku statističku akciju potrebno je izabrati najefikasniji metod posmatranja (prikupljanja podataka). U zavisnosti od skupa statističkog posmatranja, metode prikupljanja podataka mogu biti potpuno (na čitavoj populaciji) ili delimično posmatranje (na uzorku).

Statistički popis obuhvata sve jedinice statističkog skupa u određenom "kritičnom" momentu, npr. popis stanovništva na 10 godina, 31. marta. Ovaj metod daje potpuni uvid u stanje i strukturu skupa po raznim obeležjima. Veoma je skup pa se zato organizuje u dužim vremenskim intervalima.

Izveštajni metod prati kontinuirano događaje čiji je varijabilitet tokom vremena jače izražen. Sprovode ga lica ili institucije radi svojih poslovnih potreba. Statističkim izveštajem u sukcesivnim vremenskim momentima (npr. stanje novca u blagajni u mesecu) ili intervalima vremena (npr. prirodno kretanje stanovništva), vrši se posmatranje promena statističkog skupa.

Kvalitet prikupljenih podataka zavisi od:

- Specifičnosti istraživanja,
- Instrumenata istraživanja,
- Uslova istraživanja,
- Stava i ponašanja davalaca podataka.

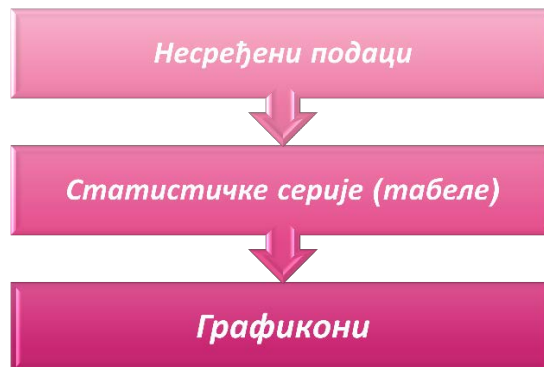
Greške su neminovni pratilac statističkih istraživanja. Mogu biti:

- Slučajne greške – greške koje nemaju poseban uticaj na kvalitet podataka;
- Sistemske greške – greške koje su u samom sistemu statističkog istraživanja i uvek utiču na podatke.

Podaci koji su prikupljeni jednom od metoda, predstavljaju sirov materijal koji treba srediti i obraditi. Podaci zapisani po redosledu prikupljanja, pre nego što se uredi po veličini ili grupišu na neki način, zovu se neuređeni, negrupisani ili sirovi podaci.

Sređivanje podataka predstavlja tehničko-metodološki postupak u kojem se, prema šemi grupisanja u cilju istraživanja, prikupljeni statistički materijal svrstava u statističke serije koje predstavljaju statistički način istraživanja. Statističke serije su rezultat sređivanja, grupisanja statističkih podataka prema određenim kriterijumima unutar utvrđene grupe.

Statističke serije se pregledno prikazuju tabelama i/ili grafikonima.



Statistička tabela je osnovna forma prikazivanja statističkih podataka. Može biti:

- Prosta – sadrži samo jednu varijablu;
- Složena – sadrži više varijabli;

- Kombinovana – dobija se ukrštanjem dve ili više varijabli.

Grafikon je osnovna forma grafičkog prikazivanja statističkih podataka. Grafičkim prikazivanjem statističkih podataka uočavaju se osnovne karakteristike posmatrane pojave.

Svaki grafikon, osim sadržaja (pojave koju prikazuje), mora da ima i sastavne elemente:

- Naslov (oznaka predmeta);
- Teritoriju (oznaka mesta);
- Vreme (oznaka vremena);
- Legendu (objašnjavanje simbola);
- Jedinice mere (u kojima su izraženi rezultati).

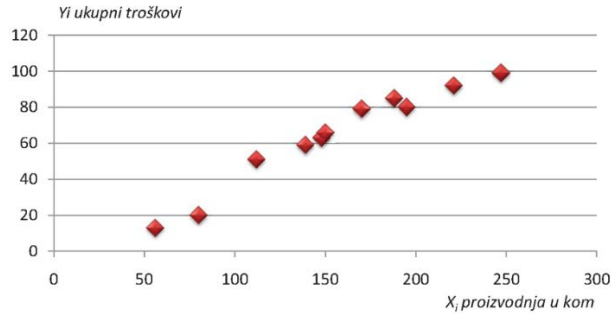
Prema elementima koje sadrže, razlikujemo:

- Dijagrame (predstavljene geometrijskim pojmovima);
- Kartograme (grafikoni na geografskim kartama);
- Piktograme (slikovito prikazane pojave).

U zavisnosti kojim geometrijskim pojmom (2D ili 3D) je predstavljen, dijagram može biti: tačkasti (stigmogram), linijski (poligon), površinski (histogram) i prostorni (stereogram).

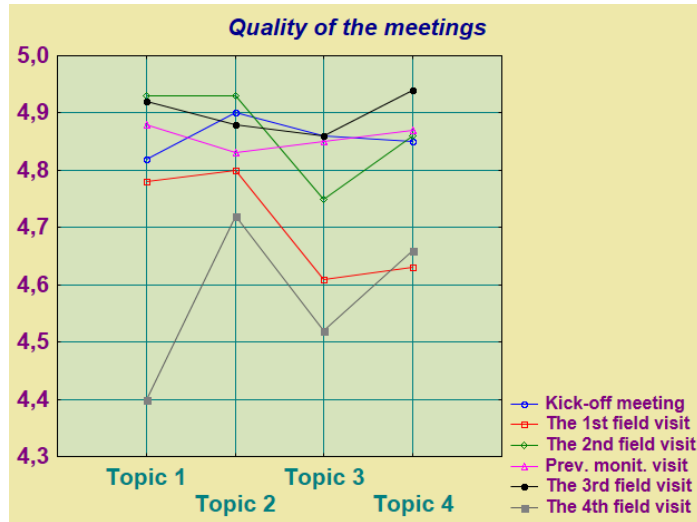
Površinskim dijagramima prikazuju se obim, veličina ili struktura jedne ili više masovnih pojava. Najčešće se koristi u obliku pravougaonika ili krugova.

Proizvodnja i ukupni troškovi

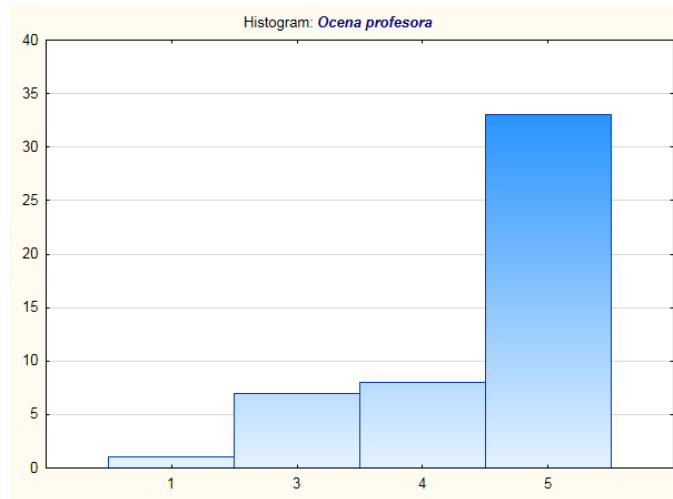


Izvor: Podaci poduzeća "A"

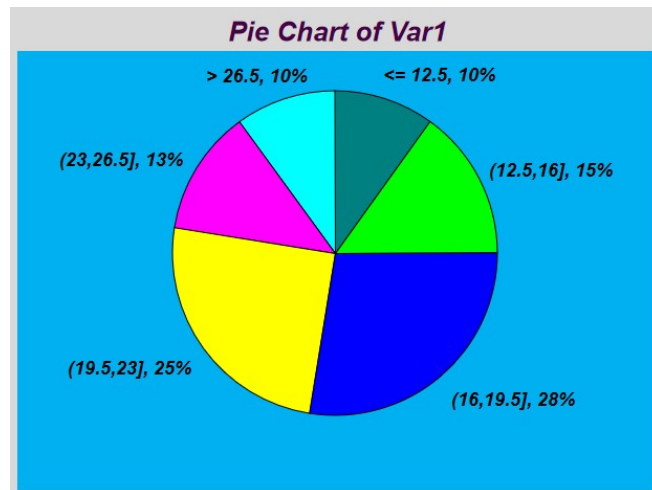
Dijagram – Tačkasti dijagram



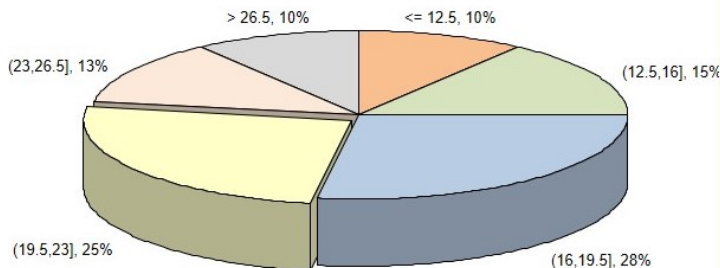
Dijagram – Linijski dijagram



Dijagram – Površinski dijagram, histogram



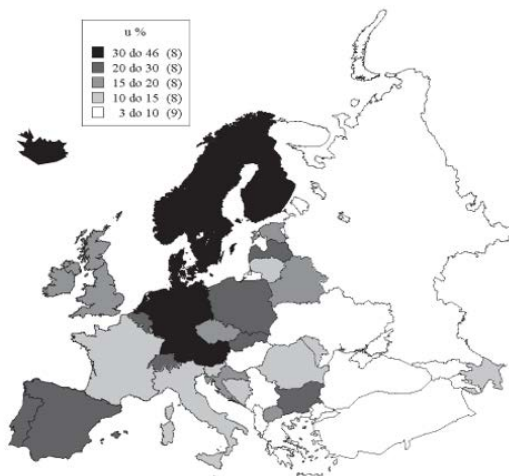
Dijagram – Površinski dijagram, strukturni krug



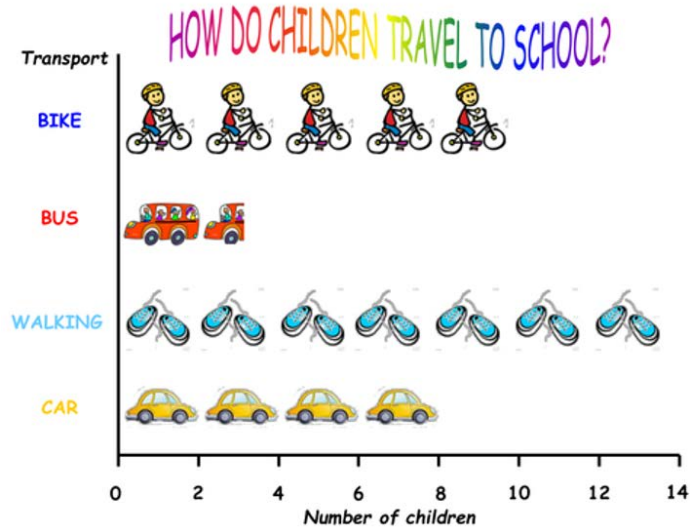
Dijagram – Prostorni dijagram, strukturni 3D krug

STATISTIČKA KARTA

EUROPSKE ZEMLJE PREMA % ŽENA U PARLAMENTU



Kartogram



Piktogram

Vrste statističkih varijabli

Na osnovu vrste izražavanja vrednosti varijabli razlikujemo:

- **Kvantitativne varijable** (numeričke) - kod kojih se vrednosti izražavaju brojevima;
- **Kvalitativne varijable** (atributivne, opisne) – kod kojih se vrednosti opisuju rečima ili simbolima.

Kvantitativne varijable mogu biti:

1. Diskretne (prekidne) – kod kojih obeležje ima konačan broj vrednosti (dozvoljava samo celobrojni zapis);
2. Kontinuirane (neprekidne) – kod kojih obeležje može

imati bilo koju vrednost konačnog ili beskonačnog intervala (dozvoljavaju i decimalni zapis).

Kvalitativne varijable možemo dalje podeliti na :

1. Nominalno-atributivne – varijable čije su vrednosti date opisno, rečima ili simbolima;
2. Nominalno-geografske – varijable vezane za neki geografski pojam;
3. Redosledne (uređajne) – varijable čije su vrednosti date opisno ali među njima postoji neka gradacija (zato se mogu simbolično i brojevima izraziti i obrađivati kao numeričke varijable).

Sledećom tabelom sumiran je prikaz vrsta i podvrsta varijabli, kao i odgovarajući primeri:

Vrsta varijable	Podvrsta	Primer
Kvantitativne	Diskretne	<i>Broj studenata, broj zaposlenih, broj stanovnika,...</i>
	Kontinuirane	<i>Visina, težina, daljina, visina plate, stopa inflacije,...</i>
Kvalitativne	Nominalno-atributivne	<i>Pol, zanimanje, boja kose, studentski prog.,...</i>
	Nominalno-geografske	<i>Mesto boravka, zemlja izvoza, mesto rođenja,...</i>
	Redosledne (uređajne)	<i>Školska sprema, kvalitet usluge,...</i>

METODE OPISIVANJA KVALITATIVNIH PODATAKA

Kvalitativne varijable imaju vrednosti koje se razvrstavaju u kategorije. Pri proučavanju takvih varijabli pažnju usmeravamo na zastupljenost pojedine kategorije u uzorku na kojem sprovodimo istraživanje ili na čitavoj populaciji. Osnovna mera kojom opisujemo zastupljenost jedne kategorije u statističkom skupu jeste frekvencija kategorije.

Frekvencije i relativne frekvencije

Frekvencija f_i , neke određene kategorije x_i , varijable X , je broj pojavljivanja te vrednosti u posmatranom skupu podataka.

Za lakše poređenje i tumačenje rezultata raznih istraživanja, u opisu zastupljenosti jedne kategorije u uzorku, koristi se i relativna frekvencija.

Relativna frekvencija f_{ri} , kategorije x_i , je broj izmerenih vrednosti varijable koje pripadaju kategoriji x_i , podeljen ukupnim brojem podataka, odnosno, ako je N broj podataka

u skupu:

$$f_{ri} = \frac{f_i}{N} ,$$

Procentualna frekvencija biće jednaka:

$$f_{ri} \times 100 = \frac{f_i}{N} \times 100 .$$

Frekvencije i grafički prikaz u softveru *Statistica*

Programski paket *Statistica* je proizvod kompanije “StatSoft” i predstavlja celovit programski paket za statističko-grafičku obradu podataka sa izuzetno velikim izborom kako osnovnih tako i naprednih statističko-matematičko-grafičkih analiza.

Uprkos gotovo isključivo profesionalnoj nameni, *Statistica* je vrlo “prijateljski raspoložena” prema korisniku (ima lagodno korisničko okruženje). Primenjene su gotovo sve napredne tehnike koje se koriste pod *Windows*-om.

Statistica ima velike mogućnosti manipulacije sa podacima. Uneti podaci mogu se pretraživati, transformisati, menjati. Pored toga veliki broj standardnih formata direktno je podržan preko interne konverzije. Tako je moguće koristiti podatke iz Excela i drugih sličnih formata. Veličina datoteka s podacima za obradu gotovo je neograničena (može imati 32000 varijabli).

Grafičke mogućnosti *Statistica* nadilaze sve ostale programske pakete tog tipa. Tako je na raspolaganju veliki broj raznih tipova 2D, 3D grafičkih prikaza. *Statistica* može

napraviti ili snimiti grafikon u *.wmf*, *.bmp*, *.ps*, *.hpgl* formatu.

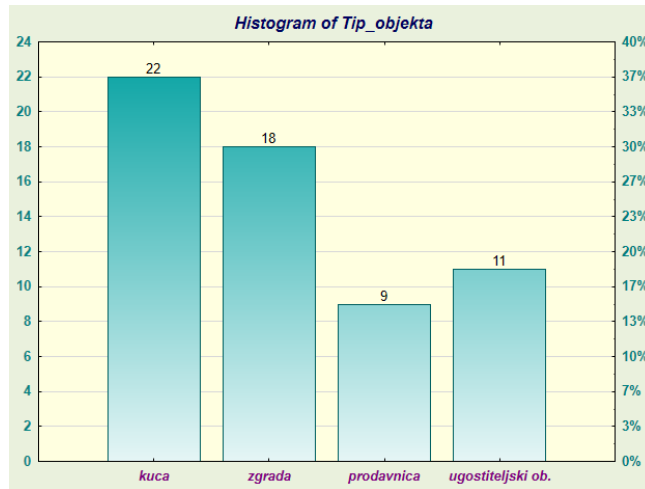
Statistica, pored procedura opšte analize, grafičkog prikaza i upravljanja bazama podataka, uključuje i implementaciju specijalizovanih metoda za analizu podataka, kao što su predikcija, primene u inženjerstvu, istraživanja u društvenim, poslovnim, biomedicinskim i drugim naukama, a pre svega predstavlja jako koristan alat za edukaciju studenata strukovnih studija iz oblasti tehničko-tehnološkog polja.

Narednim primerom biće prikazano izračunavanje frekvencija i relativnih frekvencija u programskom paketu *Statistica*, kao i grafički prikaz frekvencija preko histograma i strukturnog kruga za pojedine varijable iz date baze podataka.

Primer 1. Data je baza podataka *Podaci, upravljanje otpadom.sta*. Odrediti frekvenciju i procentualnu (relativnu) frekvenciju varijabli *Tip_objekta* i *Br_kontejnera* u programskom paketu *Statistica*. Rezultate predstaviti tabelarno i grafički.

Category	Frequency table: Tip_objekta (Podaci, upravljanje otpadom)	
	Count	Percent
kuca	22	36.66667
zgrada	18	30.00000
prodavnica	9	15.00000
ugostiteljski ob.	11	18.33333
Missing	0	0.00000

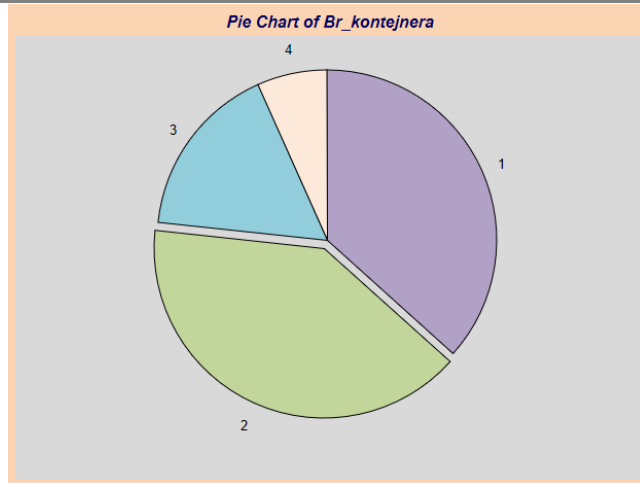
Frekvencija i relativna frekvencija: Tip_objekta



Grafički prikaz frekvencije i relativne frekvencije: Tip_objekta

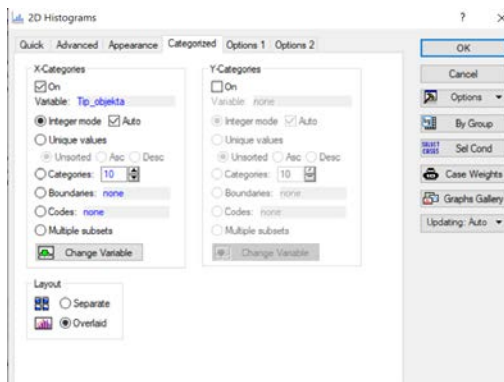
Frequency table: Br_kontejnera (Podaci, upravljanje otpadom)		
Category	Count	Percent
1	22	36.66667
2	24	40.00000
3	10	16.66667
4	4	6.66667
Missing	0	0.00000

Frekvencija i relativna frekvencija: Br_kontejnera

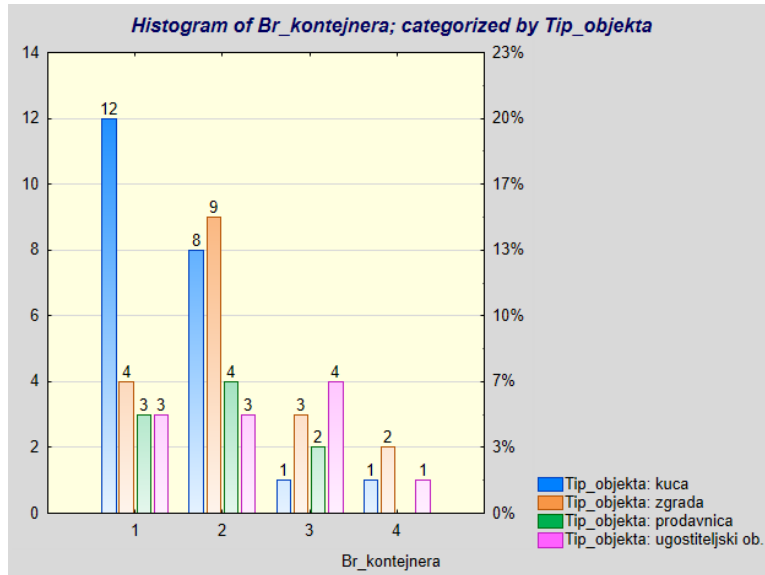


Grafički prikaz frekvencije i relativne frekvencije: Br_kontejnera

Grafički u programskom paketu *Statistica* možemo posmatrati frekvenciju jedne varijable po kategorijama druge varijable, koristeći opciju podešavanja 2D grafičkog prikaza za "Overlaid" prikaz histograma:



Dobija se uporedni prikaz dve varijable:



Grafički prikaz frekv. Br_kontejnera po kategorijama Tip_objekta

Kumulativna frekvencija

Pored frekvencije i relativne frekvencije, u ispitivanju i analizi kako kvalitativnih, tako i kvantitativnih varijabli, veliki značaj ima kumulativna frekvencija.

Rastuća kumulativna frekvencija (“Cumulative frequencies”) predstavlja pridruživanje (sabiranje) frekvencija prethodnih podataka (intervala) ili samih podataka redom. Koristi se u

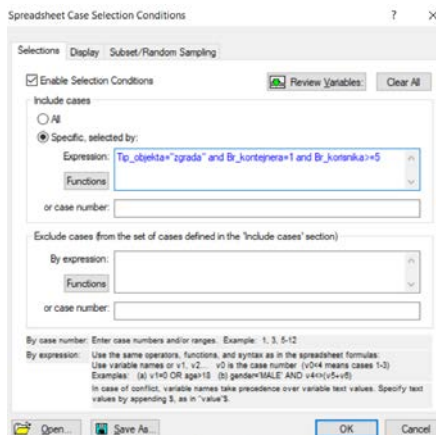
ispitivanju frekvencija ili vrednosti zadate gornjom granicom, odnosno, na pitanje: Koliko najviše?

Opadajuća kumulativna frekvencija (“100% minus cumulative percentages”) predstavlja oduzimanje frekvencija prethodnih podataka (intervala) ili samih podataka redom od zbirne frekvencije (ukupnog broja podataka). Koristi se u ispitivanju frekvencija ili vrednosti zadate donjom granicom, odnosno, na pitanje: Koliko najmanje?

U programskom paketu *Statistica* moguće je ispitivati frekvencije varijabli u okviru pojedinih kategorija nekih drugih varijabli.

Primer 2. Data je baza podataka *Podaci, upravljanje otpadom.sta*. Koliko kuća, koje poseduju 1 kontejner, a imaju 5 ili više korisnika, generiše najviše 32kg otpada nedeljno, a koliko njih generiše najmanje 25kg (broj i %)?

Koristeći specifikaciju podataka (“Selection Conditions”), u softveru:



i ispitivanjem frekvencija, rastuće i opadajuće kumulante, dolazi se do rešenja zadatka.

Frequency table: <i>Kolicina otpada</i> (Podaci, upravljanje otpadom)						
Include condition: Tip objekta="kuca" and Br kontejnera=1 and Br korisnika>						
Category	Count	Cumulative Count	Percent	Cumulative Percent	100% - Percent	
19	1	1	10.00000	10.0000	100.0000	
22	1	2	10.00000	20.0000	90.0000	
23	1	3	10.00000	30.0000	80.0000	
25	1	4	10.00000	40.0000	70.0000	
27	1	5	10.00000	50.0000	60.0000	
31	2	7	20.00000	70.0000	50.0000	
32	1	8	10.00000	80.0000	30.0000	
34	1	9	10.00000	90.0000	20.0000	
49	1	10	10.00000	100.0000	10.0000	
Missing	0	10	0.00000	100.0000	0.0000	

Odgovor na pitanje: Koliko kuća, koje poseduju 1 kontejner, a imaju 5 ili više korisnika, generiše najviše 32kg otpada nedeljno, dobijamo koristeći rastuću kumulantu, dakle, 8, odnosno, 80% (što je označeno plavom bojom).

Odgovor na pitanje: Koliko kuća, koje poseduju 1 kontejner, a imaju 5 ili više korisnika, generiše najmanje 25kg otpada nedeljno, dobijamo koristeći opadajuću kumulantu, a vrednost ćemo dobiti oduzimanjem od zbirne frekvencije frekvencija prethodnih vrednosti od 25 ili sabiranjem frekvencija za vrednost 25 i veće, dakle, 7, odnosno, 70% (što je označeno ljubičastom bojom).

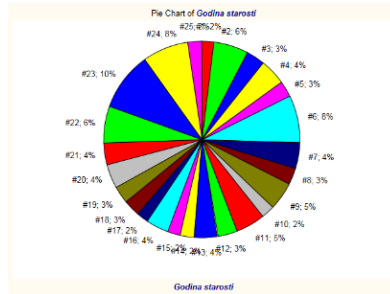
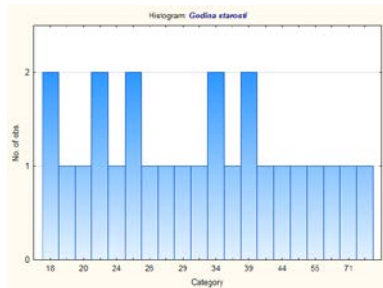
METODE OPISIVANJA KVANTITATIVNIH PODATAKA

Kvantitativni podaci mogu biti prikupljeni posmatranjem (merenjem) numeričkih ili redoslednih (uređajnih) podataka.

Redosledne varijable najčešće imaju samo nekoliko različitih vrednosti. Kod numeričkih varijabli, posebno kod kontinuiranih, može se desiti da imaju jako veliki broj različitih vrednosti.

Ako su numeričke varijable diskretne sa malo mogućih vrednosti, ili ako su varijable redosledne (uređajne), za opis podataka možemo koristiti iste metode kao pri opisivanju kvalitativnih podataka: frekvencije i relativne frekvencije.

Ako numeričke varijable imaju veliki broj različitih vrednosti, tabelarni i grafički prikazi frekvencija mogu biti nedovoljno informativni, a obrada i analiza podataka takvim postupkom ne služi svrsi, što ilustruje sledeći grafički i tabelarni prikaz primera numeričke varijable sa velikim brojem različitih vrednosti u varijabli.



Frequency table: <i>Godina starosti</i>				
Category	Count	Cumulative Count	Percent	Cumulative Percent
18	2	2	8,000000	8,0000
19	1	3	4,000000	12,0000
20	1	4	4,000000	16,0000
21	2	6	8,000000	24,0000
24	1	7	4,000000	28,0000
25	2	9	8,000000	36,0000
26	1	10	4,000000	40,0000
28	1	11	4,000000	44,0000
29	1	12	4,000000	48,0000
33	1	13	4,000000	52,0000
34	2	15	8,000000	60,0000
36	1	16	4,000000	64,0000
39	2	18	8,000000	72,0000
42	1	19	4,000000	76,0000
44	1	20	4,000000	80,0000
52	1	21	4,000000	84,0000
55	1	22	4,000000	88,0000
69	1	23	4,000000	92,0000
71	1	24	4,000000	96,0000
88	1	25	4,000000	100,0000
Missing	0	25	0,000000	100,0000

Grafički i tabelarni prikaz frekvencije, numerička var. sa velikim brojem različitih vrednosti

U slučaju diskretnih varijabli sa velikim brojem različitih vrednosti ili u slučaju kontinuiranih varijabli, potrebno je izvršiti kategorizaciju, odnosno, grupisanje vrednosti.

Grupisanje kvantitativnih podataka računski

Grupisanje se vrši u disjunktним intervalima po kriterijumu za koji smatramo da će nam dati željene rezultate.

Kategorizaciju numeričkih varijabli moguće je izvesti na više načina. Jedan od načina i najčešći je podela skupa podataka na jednake disjunktne skupove (intervale).

Za grupisanje kvantitativnih podataka koristi se *Sturges*-ova formula koja nam daje broj disjunktних intervala k , na koji treba podeliti dati statistički skup numeričkih podataka i širinu intervala i :

$$k = 1 + 3,3 \log N ,$$

$$i = \frac{x_{max} - x_{min}}{k} ,$$

gde je:

- k – broj intervala;
- N – broj podataka;
- i – širina intervala;
- X_{max} – najveća vrednost varijable;
- x_{min} – najmanja vrednost varijable.

Intervalna numerička serija se sastoji iz donje i gornje granice

intervala. Ovakve serije se, radi matematičke obrade, pretvaraju u neintervalne numeričke serije metodom razredne sredine: Zbir donje i gornje granice intervala podeljen sa 2.

Raspodela frekvencija grupisanih numeričkih podataka je tabelarni prikaz dva niza podataka:

1. Vrednosti variable prikazane grupnim intervalima;
2. Broj podataka koji pripadaju datom intervalu (frekvencija).

Ovako prikazani podaci nazivaju se grupisani podaci.

		Frequenc
From	To	Count
22	$\leq x < 32$	4
32	$\leq x < 42$	5
42	$\leq x < 52$	9
52	$\leq x < 62$	13
62	$\leq x < 72$	7
72	$\leq x < 82$	7
82	$\leq x < 92$	5

Grupisani podaci

Granice intervala mogu biti:

- **Prave** - ako je donja granica svakog intervala jednaka gornjoj granici prethodnog intervala;
- **Nominalne** – ako nisu prave.

U obradi podataka se koriste prave granice, zato se nominalne prevode u prave tako što se uzima granična vrednost intervala - Poluzbir donje nominalne granice intervala i gornje nominalne

granice njemu prethodnog intervala.

Nominalne granice	Frekvencija intervala	Granična vrednost intervala (prevođenje u prave granice)	Širina intervala	Razredna sredina intervala
18 - 30	12	17.5 – 30.5	13	24
31 - 43	19	30.5 – 43.5	13	37
44 - 56	14	43.5 – 56.5	13	50
57 - 69	5	56.5 – 69.5	13	63

Prevođenje nominalnih u prave granice

Primer 3. Data je serija podataka. Grupisati podatke, odrediti razrednu sredinu, frekvenciju, relativnu (procentualnu) frekvenciju, rastuću i opadajuću kumulantu.

30	13	21	9	19	17	15	23	20	23
24	19	11	15	22	16	17	29	18	19
23	21	19	24	18	17	25	18	26	27
13	26	11	30	20	16	10	23	20	19

Broj podataka je $N = 40$, pa je primenom *Sturges*-ove formule:

$$k = 1 + 3,3 \log 40 = 6,286 \approx 6.$$

Najveći i najmanji podatak u seriji su: $x_{max} = 30$, $x_{min} = 9$, pa je:

$$i = \frac{30-9}{6} = 3,5 \approx 4.$$

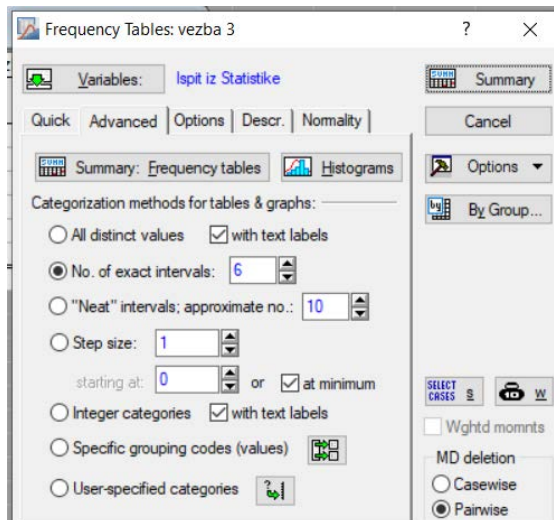
Nakon podele date serije na 6 disjunktnih intervala širine 4 i prebrojavanjem podataka, dobijaju se traženi podaci:

Prave granice (x_i)	Frekvencije (f_i)	Razredna sredina (x_i')	Relativna frekvencija % (f_{ri})	Rastuća kumulanta ($\sum f_i \nearrow$)	Opadajuća kumulanta ($\sum f_i \searrow$)
9 - 13	4	11	10	4	40
13 - 17	6	15	15	10	36
17 - 21	14	19	35	24	30
21 - 25	9	23	22.5	33	16
25 - 29	4	27	10	37	7
29 - 33	3	31	7.5	40	3
	$\Sigma=40$		$\Sigma=100\%$		

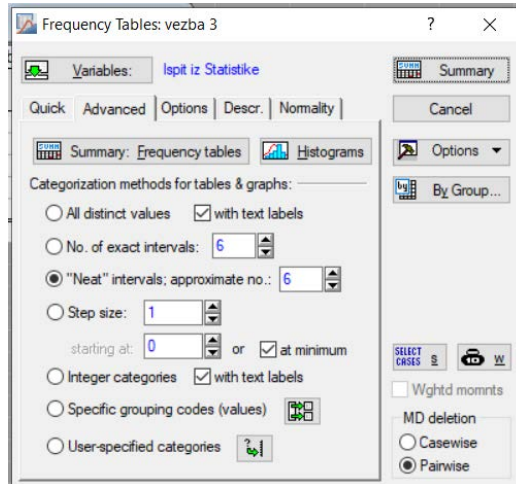
Grupisanje kvantitativnih podataka u softveru *Statistica*

U programskom paketu *Statistica* moguće je vršiti grupisanje numeričke serije podataka na više načina:

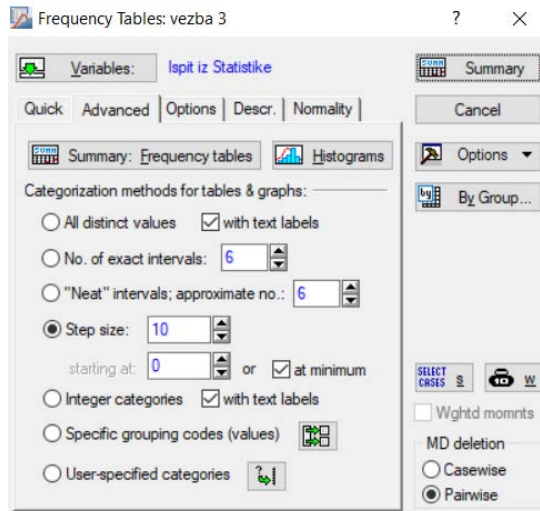
- Zadavanjem broja intervala:



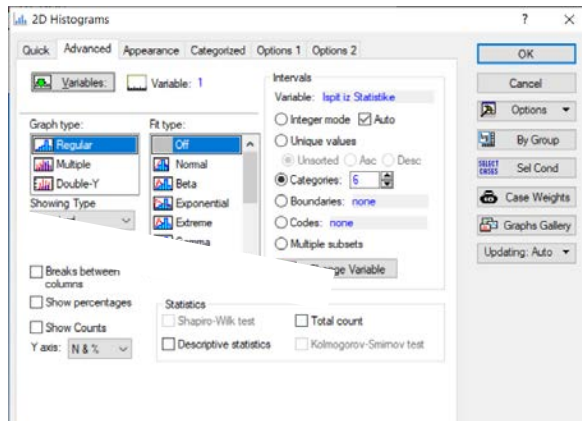
- Prepuštanjem softveru da odredi “pravilne” intervale zadavajući okvirno broj intervala:



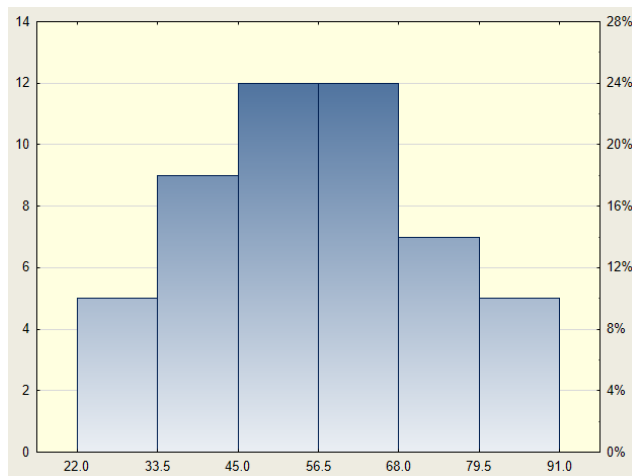
- Zadavanjem širine intervala (što ćemo koristiti u radu):



Moguće je u programskom paketu *Statistica* i grafički predstaviti frekvencije za grupisane podatke zadavanjem broja intervala:

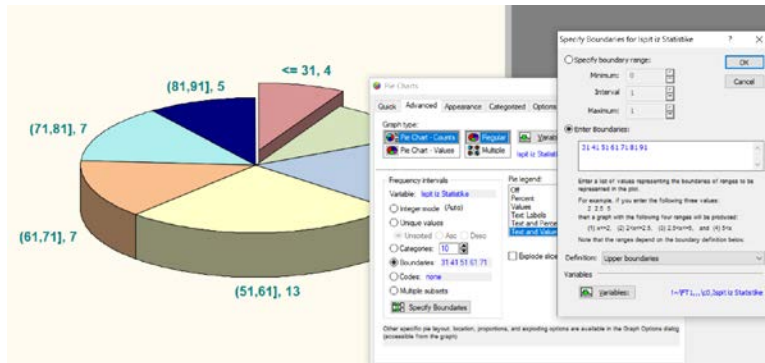


Dobija se histogram frekvencija intervala za zadat broj intervala koje softver generiše:



Histogram grupisanih podataka, za zadat broj intervala

Takođe je moguće uneti samostalno granice intervala koje se dobijaju *Sturges*-ovom formulom, ali će pritom za generisanje grafika softver uključiti gornju a ne donju granicu intervala, za razliku od računске obrade podataka gde je obrnuto. Treba imati to u vidu pri zadavanju gornje granice. Iste mogućnosti i principi važe za grafički prikaz i histograma i strukturnog kruga.



Strukturni krug grupisanih podataka, za zadate granice intervala

MERE POPULACIJE I UZORKA

Istraživanje statističkog skupa polazi od pojedinačnih vrednosti obeležja, a zaključci o celom skupu ne mogu se izvoditi izolovanim posmatranjem tih podataka. Zato se serija podataka zamenjuje malim brojem novih veličina. Te veličine treba što bolje da informišu o posmatranom skupu i pruže najvažnije informacije o rasporedu vrednosti posmatrane varijable statističkog skupa.

Koristeći relativne brojeve i raspodelu frekvencija može se steći izvestan globalni utisak o posmatranoj pojavi. Ali za dalju, svrsishodniju analizu potrebne su preciznije metode za obradu mase statističkih podataka i donošenje zaključaka. Zato se definišu izvesni pokazatelji ili parametri. Te pokazatelje koncentracije, oblika i disperzije podataka u nekom statističkom skupu možemo podeliti u tri grupe:

1. **Mere centralne tendencije** (mere koncentracije ili srednje vrednosti);
2. **Mere disperzije** (mere varijacije podataka);
3. **Mere oblika** (mere asimetrije i spljoštenosti raspodele podataka).

MERE CENTRALNE TENDENCIJE

Srednje vrednosti čine veoma značajnu grupu statističkih indikatora u istraživanju i analizi masovnih pojava. Najčešće se primenjuju u praksi i na njima se zasniva veliki deo statističke analize. Srednja vrednost je kvantitativni reprezent svih individualnih vrednosti nekog numeričkog obeležja posmatrane pojave.

Na osnovu toga da li se mere ove grupe parametara dobijaju izračunavanjem ili na osnovu njihove pozicije u uredjenoj numeričkoj seriji podataka, možemo ih podeliti u dve grupe: izračunate i pozicione.



Podela i vrste mera centralne tendencije

Izračunate srednje vrednosti: Aritmetička sredina

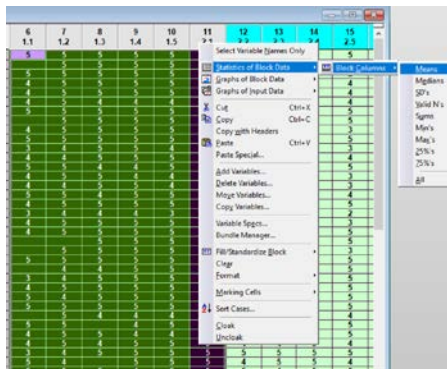
Izračunate srednje vrednosti dobijamo izračunavanjem bilo za negrupisane serije podataka (proste vrednosti), bilo za grupisane serije podataka (ponderisane vrednosti).

Prosta aritmetička sredina (μ) je mera centralne tendencije koja se koristi kod negrupisanih podataka. Izračunava se po formuli:

$$\mu = \frac{x_1 + x_2 + \dots + x_N}{N},$$

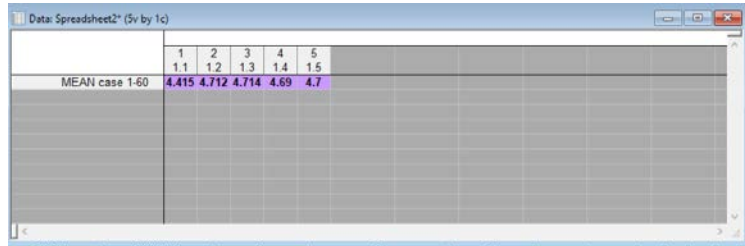
gde je N broj podataka u statističkoj seriji.

U softveru *Statistica* moguće je lako računati prostu aritmetičku sredinu bilo po kolonama (varijabli) bilo po vrstama (za vrednosti jednog ispitanika "case") ili sve zajedno.

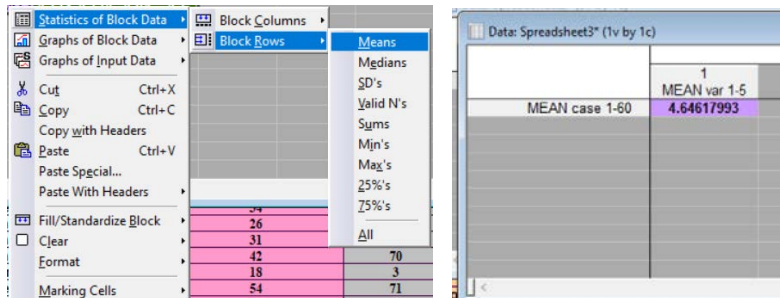


Izračunavanje aritmetičke sredine po kolonama

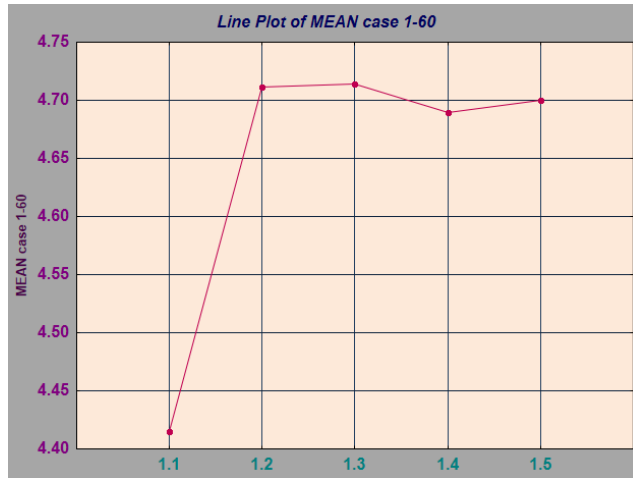
Time se dobija novi “spreadsheet” sa podacima koje je moguće dalje obrađivati.



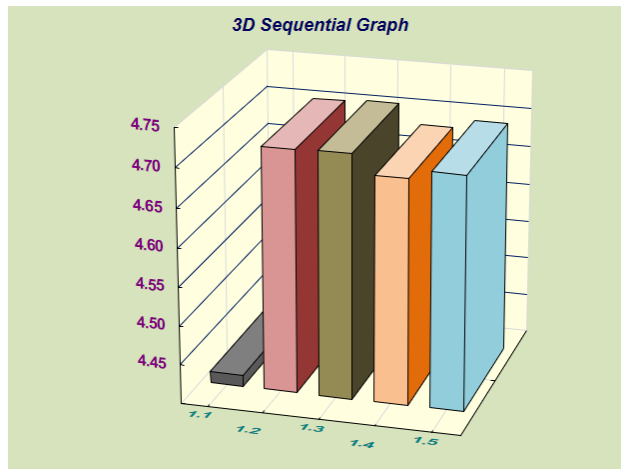
Ponavljanjem postupka ali sada za vrstu, dobija se ukupna aritmetička sredina svih podataka u matrici.



U softveru *Statistica* moguće je i grafički prikazati dobijene vrednosti aritmetičkih sredina više varijabli. Linijskim dijagramom ili 3D grafikom (“Raw Data Plots”) moguće je vršiti uporednu analizu.



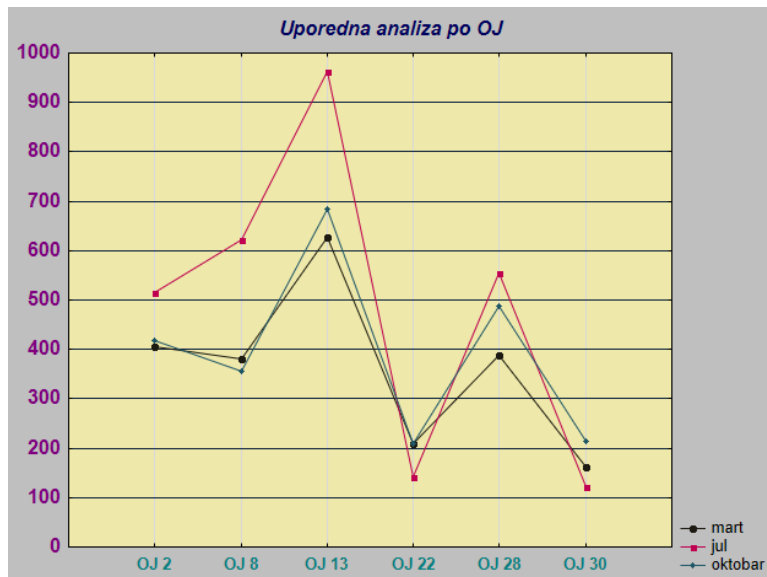
Linijski dijagram aritmetičkih sredina varijabli od 1.1-1.5



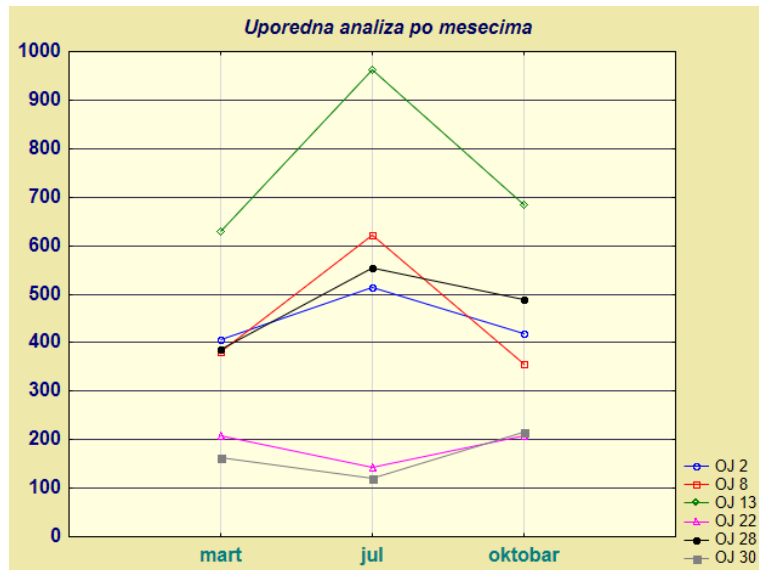
3D grafik aritmetičkih sredina varijabli od 1.1-1.5

Primer 4. Data je baza podataka količine tretiranog infektivnog medicinskog otpada UKC Niš po mesecima i po organizacionim jedinicama (OJ). Napraviti uporednu analizu prosečne tretirane količine infektivnog medicinskog otpada u programskom paketu *Statistica* (linijskim dijagramom) Organizacionih jedinica: 2, 8, 13, 22, 28 и 30 u periodu od 7. do 25. u mesecima mart, jul i oktobar.

Nakon formiranja tabela za odgovarajuće periode i tražene organizacione jedinice, dobija se uporedna analiza koja se može posmatrati na dva načina: upredna analiza po OJ u datom periodu i uporedna analiza po mesecima za sveku OJ.



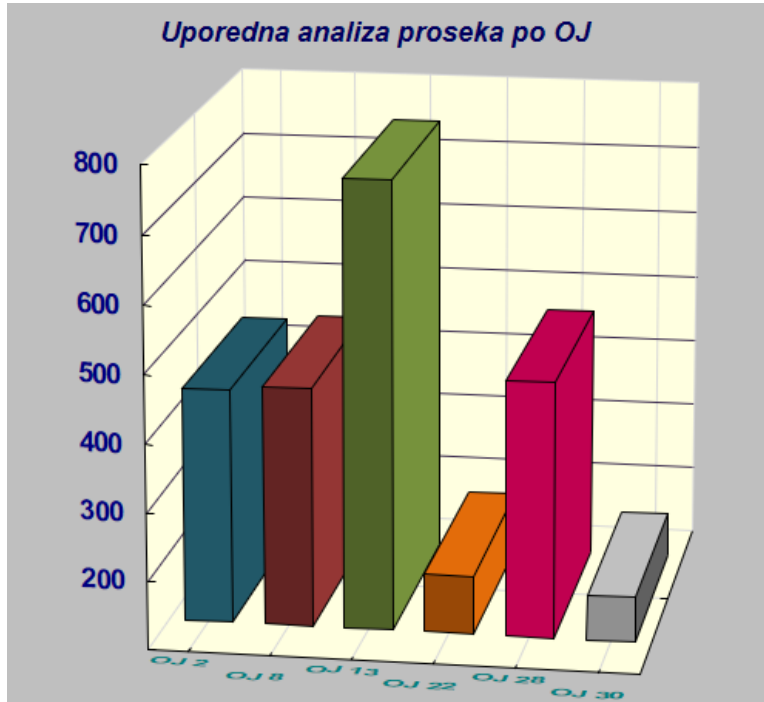
Uporedna analiza količine otpada (u kg) po OJ



Uporedna analiza količine otpada (u kg) po mesecima

Primer 5. Data je baza podataka količine tretiranog infektivnog medicinskog otpada UKC Niš po mesecima i po organizacionim jedinicama (OJ). Izračunati ukupnu prosečnu količinu tretiranog medicinskog otpada za svaki od tri meseca: mart, jul i oktobar u odnosu na posmatrani period od 7. - 25. u pomenutim mesecima, Organizacionih jedinica: 2, 8, 13, 22, 28 i 30 i grafički je prikazati za svaku jedinicu u programskom paketu *Statistica*, (3D grafik).

Određivanjem prosečnih vrednosti (aritmetičkih sredina) za sva tri meseca zajedno po OJ, dobija se grafički prikaz uporedne analize po OJ u pomenutom periodu.



Uporedna analiza prosečne količine otpada (u kg) po OJ

U slučaju grupisanih podataka nije moguće vršiti obradu podataka u softveru, već računski. Samim tim nije moguće ni grafički prikaz obrađenih podataka.

Ponderisana aritmetička sredina (μ) je mera centralne tendencije koja se koristi kod grupisanih podataka. Ako je data serija grupisana u k disjunktih intervala, gde su x_1', x_2', \dots, x_k' razredne sredine tih intervala, a f_1, f_2, \dots, f_k , odgovarajuće frekvencije intervala, ponderisana aritmetička sredina se zračunava po formuli:

$$\mu = \frac{f_1 x_1' + f_2 x_2' + \dots + f_k x_k'}{N},$$

gde je $N = \sum_{i=1}^k f_i$.

Primer 6. Za datu grupisanu seriju podataka izračunati ponderisanu aritmetičku sredinu.

x_i	f_i
22 - 32	4
32 - 42	5
42 - 52	9
52 - 62	13
62 - 72	7
72 - 82	7
82 - 92	5
	$\Sigma=50$

Kako bi izračunali ponderisanu aritmetičku sredinu, datu tabelu proširćemo potrebnim kolonama: kolonom razrednih sredina intervala i kolonom proizvoda vrednosti razredne sredine i odgovarajuće frekvencije intervala.

x_i	f_i	x_i'	$f_i \cdot x_i'$
22 - 32	4	27	108
32 - 42	5	37	185
42 - 52	9	47	423
52 - 62	13	57	741
62 - 72	7	67	469
72 - 82	7	77	539
82 - 92	5	87	435
	$\Sigma=50$		$\Sigma=2900$

Poslednja kolona u tabeli, tačnije zbir vrednosti te kolone, daje nam potreban podatak za izračunavanje ponderisane aritmetičke sredine:

$$\mu = \frac{2900}{50} = 58 .$$

Aritmetička sredina ima sledeće osobine:

- Veća je od najmanje, a manja od najveće vrednosti posmatrane varijable: $x_{min} < \mu < x_{max}$;
- Ako je $x_1 = x_2 = \dots = x_N = a \Rightarrow \mu = a$;
- Zbir odstupanja vrednosti varijable od aritmetičke sredine jednak je nuli.

Izračunate srednje vrednosti: Harmonijska sredina

Harmonijska sredina upotrebaljava se u onim slučajevima kada numerička vrednost obeležja i obim pojave stoje u obrnutoj srazmeri i kada su vrednosti obeležja za koje treba izračunati sredinu izražene u vidu recipročnih odnosa.

Taj odnos reciprociteta sastoji se u tome što se vrednost tih obeležja smanjuje kada se pojava povećava i obrnuto.

Harmonijska sredina je recipročna aritmetička sredina recipročnih vrednosti podataka.

Prosta harmonijska sredina (H) je mera centralne tendencije koja se koristi kod negrupisanih podataka. Izračunava se po formuli:

$$H = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_N}},$$

Ponderisana harmonijska sredina (H) je mera centralne tendencije koja se koristi kod grupisanih podataka.

Ako je data serija grupisana u k disjunktnih intervala, gde su x_1', x_2', \dots, x_k' razredne sredine tih intrevala, a f_1, f_2, \dots, f_k , odgovarajuće frekvencije intervala, ponderisana harmonijska sredina se zračunava po formuli:

$$H = \frac{N}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \dots + \frac{f_k}{x_k}} ,$$

gde je $N = \sum_{i=1}^k f_i$.

Harmonijska sredina se najčešće koristi kod prosečne proizvodnje u jedinici proizvoda. Ne može se koristiti ako se u varijabli pojavljuju negativne vrednosti ili nula.

Harmonijska sredina je manja i od aritmetičke i od geometrijske sredine istog statističkog skupa.

Izračunate srednje vrednosti: Geometrijska sredina

Geometrijska sredina se koristi za izračunavanje serije podataka koja ima ubrzan rast. Predstavlja vrednost koja daje prosek relativnih pokazatelja, bilo da su indeksi ili proporcionalne promene između vrednosti podataka posmatrane pojave.

Prosta geometrijska sredina (G) je mera centralne tendencije koja se koristi kod negrupisanih podataka. Izračunava se po formuli:

$$G = \sqrt[N]{x_1 \cdot x_2 \cdot \dots \cdot x_N} ,$$

odnosno, logaritmovanjem obe strane jednakosti dolazi se do formule lakšeg izračunavanja:

$$\log G = \frac{\log x_1 + \log x_2 + \dots + \log x_N}{N}$$

Ponderisana geometrijska sredina (G) je mera centralne tendencije koja se koristi kod grupisanih podataka.

Ako je data serija grupisana u k disjunktih intervala, gde su x_1', x_2', \dots, x_k' razredne sredine tih intrevala, a f_1, f_2, \dots, f_k , odgovarajuće frekvencije intervala, ponderisana harmonijska sredina se zračunava po formuli:

$$G = \sqrt[N]{x_1^{f_1} \cdot x_2^{f_2} \cdot \dots \cdot x_k^{f_k}}$$

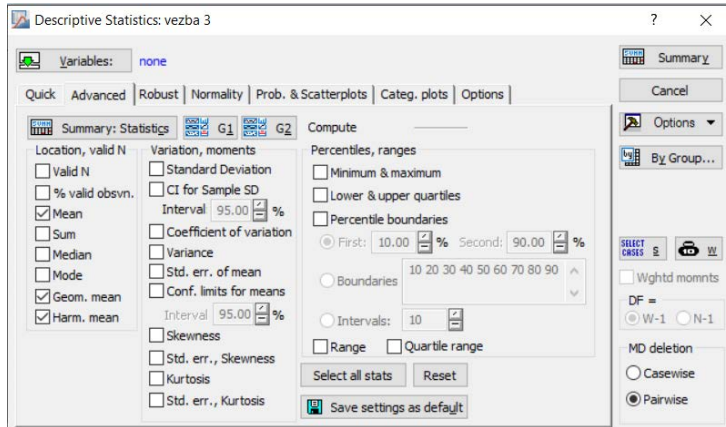
gde je $N = \sum_{i=1}^k f_i$.

Odnosno, logaritmovanjem se dobija:

$$\log G = \frac{f_1 \cdot \log x_1 + f_2 \cdot \log x_2 + \dots + f_k \cdot \log x_k}{N}$$

Za aritmetičku, harmonijsku i geometrijsku sredinu istog statističkog skupa važi relacija: $H < G < \mu$.

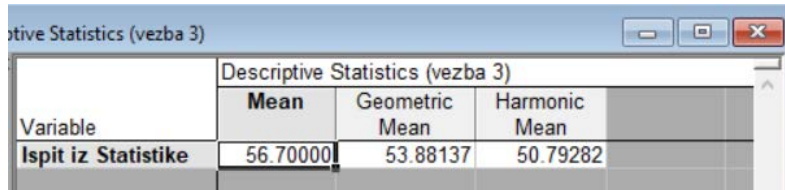
U programskom paketu *Statistica* se izračunavanje harmonijske i geometrijske sredine lako izračunava u delu "Descriptive statistics" čekiranjem traženih veličina:



Primer 7. Na ispitu iz Statistike, 50-oro studenata je polagalo i ostvarilo bodove date u tabeli. Za datu statističku seriju podataka *Ispit iz statistike* izračunati aritmetičku, harmonijsku i geometrijsku sredinu u softveru *Statistica*.

44	22	65	47	43	33	51	79	59	78
66	37	36	31	57	62	37	84	53	76
57	45	43	52	64	74	85	88	46	58
25	41	49	53	66	82	91	54	58	68
77	56	55	52	73	26	78	54	43	62

Unošenjem datih podataka u “spreadsheet” softvera, čekiranjem traženih vrednosti u delu “Descriptive statistics”, dobija se traženo rešenje:



Variable	Mean	Geometric Mean	Harmonic Mean
Ispit iz Statistike	56.70000	53.88137	50.79282

Pozicione srednje vrednosti: Modus i Medijana

Pozicione srednje vrednosti određuju se pozicijom koju zauzimaju u datoj seriji podataka. Nalaze se na onom mestu koje zauzima bilo dominantan (najznačajniji), bilo centralni (središnji) položaj u seriji. Pre nego što se pristupi određivanju pozicionih srednjih vrednosti, potrebno je datu seriju poredati po veličini modaliteta. U pozicione srednje vrednosti spadaju modus i medijana.

Modus (*Mo*) je obeležje u seriji koje ima najveću frekvenciju pojavljivanja. Najčešće se javlja pa je najtipičnija (dominantna) vrednost u seriji. Zato se često naziva i dominantna ili normala. Obzirom da je vezan za frekvenciju kategorije varijable, modus možemo računati i kod atributivnih varijabli. Modus ne mora biti jedinstven.

Seriya može biti:

- Bez modusa,
(Pr.: 12.3, 24.4, 46.8, 32.5, 18.2, 43.3, 53.2);
- Unimodalna (sa jednim modusom),
(Pr.: Boja kose studenata u grupi je: smeđa, plava, smeđa,

smeđa, riđa, smeđa, plava, smeđa; M_o =smeđa);

- Bimodalna (sa dva modusa),
(Pr.: Godine starosti studenata u grupi: 18, 19, 20, 20, 19, 19, 20, 18, 19, 20; M_o su 19 i 20);
- Multimodalna (sa više modusa).

U praksi se modus može određivati i kod negrupisanih i kod grupisanih podataka.

Kod negrupisanih podataka se uređivanjem serije i određivanjem frekvencija podataka određuje modus.

Primer 8. На испиту је 33 студента добило следеће оцене: 6, 7, 6, 5, 10, 7, 9, 6, 5, 7, 9, 8, 7, 6, 6, 10, 7, 6, 7, 8, 7, 5, 6, 5, 10, 5, 7, 9, 8, 8, 6, 7, 7. Коју оцену су студенти најчешће добијали?

Formiranjem tabele sa datim vrednostima u rastućem poretku i određivanjem frekvencija, dobija se da najveću frekvenciju ima podatak 7.

x_i	f_i
5	5
6	8
7	10
8	4
9	3
10	3
	$\Sigma=33$

Dakle, studenti su najčešće dobijali ocenu 7.

U slučaju grupisanih podataka, modus se izračunava po formuli:

$$M_o = a_{M_o} + i \frac{f_{M_o} - f_{M_o-1}}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})},$$

gde je:

- a_{M_o} = donja granica modalnog intervala,
- i = širina intervala,
- f_{M_o} = frekvencija modalnog intervala,
- f_{M_o-1} = frekvencija predmodalnog intervala,
- f_{M_o+1} = frekvencija poslemodalnog intervala.

Primer 9. Za datu grupisanu seriju podataka odrediti modus.

x_i	f_i
22 - 32	4
32 - 42	5
42 - 52	9
52 - 62	13
62 - 72	7
72 - 82	7
82 - 92	5
	$\Sigma=50$

Kako je najveća frekvencija u intervalu od 52-62, modus ćemo izračunati po zadatoj formuli za grupisane podatke:

$$M_o = 52 + 10 \frac{13-9}{(13-9)+(13-7)} = 56.$$

Modus predstavlja grubu meru proseka i poželjno ga je koristiti da se isključe ekstremne vrednosti, ali nije poželjno ukoliko ekstremna vrednost ima najveću frekvenciju.

Nedostaci modusa su što uzima u obzir samo vrednost sa najvećom frekvencijom a ostale zanemaruje. Nepouzdan je za izrazito asimetrične raspodele u kojima je modalni interval otvoren.

Antimodus je vrednost sa najmanjom frekvencijom.

Grafički se modus prikazuje histogramom. U slučaju negrupisanih podataka, modus je kategorija varijable sa najvećom ordinatom. Kod grupisanih serija, modus se grafički može samo približno naći, odnosno, samo interval u kom se modus nalazi.

Medijana (M_e) je poziciona srednja vrednost i jednaka je vrednosti središnjeg člana numeričke serije podataka koji su rangirani u rastućem poretku. Medijana deli grupisanu seriju numeričkih podataka na dva jednaka dela, tako da prvu polovinu čine vrednosti manje od nje, a drugu veće vrednosti.

Ako je broj numeričkih podataka u seriji neparan, medijana će biti jednaka vrednosti središnjeg člana. Ako je broj numeričkih podataka u seriji paran, vrednost medijane je aritmetička sredina dva središnja podatka u seriji.

U slučaju negrupisanih podataka, gde se svaki podatak javlja

samo po jednom (nema frekvencija), medijanu određujemo u zavisnosti od parnosti broja podataka:

$$x_1 \leq x_2 \leq \dots \leq x_n ,$$

- $M_e = x_{\frac{n+1}{2}}$, ako je n neparan broj;
- $M_e = \frac{1}{2} (x_{\frac{n}{2}} + x_{\frac{n}{2}+1})$, ako je n paran broj.

Primer 10. Izračunati medijane sledećih serija:

- a) 27, 36, 25, 31, 15;
- b) 28, 38, 33, 37, 14, 26.

Nakon uređivanja serija u rastućem poretku i određivanjem parnosti broja podataka, dobija se:

- a) 15, 25, 27, 31, 36 ; $n=5$ - neparan broj,

$$M_e = x_3 = 27 .$$

- b) 14, 26, 28, 33, 37, 38 ; $n=6$ - paran broj,

$$M_e = \frac{1}{2} (x_3 + x_4) = \frac{28+33}{2} = 30,5 .$$

U slučaju negrupisanih podataka sa ponavljanjem, koristi se rastuća kumulanta i potrebno je najpre odrediti poziciju medijane u zavisnosti od parnosti broja podataka:

- $(\text{poz. } M_e) = \frac{N}{2}$, ako je N - neparan broj;
- $(\text{poz. } M_e) = \frac{N+1}{2}$, ako je N - paran broj.

Određivanjem vrednosti u rastućoj kumulanti, počev od najmanje redom, u kom je pozicija medijane sadržana (prvi broj jednak poziciji ili veći od nje), podatak kome odgovara određena vrednost u kumulanti, biće jednak medijani.

Primer 11. Izračunati medijanu broja članova domaćinstva. Podaci su dati tabelom:

Број чланова домаћинства (x_i)	Број домаћинства (f_i)	Растућа кумулат. фрекв. ($\sum f_i$)
1	10	10
2	16	26
3	24	50
4	34	84
6	29	113
8	13	126
Σ	$\Sigma f_i = 126$	/

Kako je broj podataka 126 – paran broj, pozicija medijane jednaka je:

$$(\text{poz. } M_e) = \frac{126+1}{2} = 63,5 .$$

Određivanjem vrednosti u rastućoj kumulanti u kom je pozicija medijane (broj 63,5) sadržan (prvi jednak njemu ili veći od njega), dobija se broj 84, koji odgovara podatku 4. Na taj način, preko rastuće kumulante određujemo medijanu.

Број чланова домаћинства (x_i)	Број домаћинства (f_i)	Растућа кумулат. фрекв. ($\sum f_i$)
1	10	10
2	16	26
3	24	50
4	34	84
6	29	113
8	13	126
Σ	$\Sigma f_i=126$	/

Dakle, $M_e = 4$.

U slučaju grupisanih podataka, takođe se koristi rastuća kumulanta i pozicija medijane u zavisnosti od parnosti broja podataka, na isti način kao kod negrupisanih podataka sa ponavljanjem. Na osnovu pozicije medijane određuje se medijalni intervala preko rastuće kumulante. Medijana se u tom slučaju izračunava formulom:

$$M_e = a_{M_e} + \frac{i}{f_{M_e}} \left((\text{poz. } M_e) - W_e \right),$$

gde je:

- a_{M_e} = donja granica medijalnog intervala,
- i = širina intervala,
- f_{M_e} = frekvencija medijalnog intervala,
- $(\text{poz. } M_e)$ = pozicija medijane,
- W_e = kumulanta predmedijalnog intervala.

Primer 12. Za datu grupisanu seriju podataka iz *Primer 9*. odrediti medijanu.

Kako je ukupan broj podataka $N=50$ – paran broj, pozicija medijane biće jednaka:

$$(\text{poz. } M_e) = \frac{51}{2} = 25,5 .$$

U kumulanti broj u kome je pozicija medijane sadržana je 31,

x_i	f_i	$\Sigma f_i \nearrow$
22 - 32	4	4
32 - 42	5	9
42 - 52	9	18
52 - 62	13	31
62 - 72	7	38
72 - 82	7	45
82 - 92	5	50
	$\Sigma=50$	

što nam daje odgovor u kom intervalu je medijana pa se primenom formule za njeno izračunavanje, dobija:

$$M_e = 52 + \frac{10}{13} (25,5 - 18) = 57,77 .$$

Osobine medijane su sledeće:

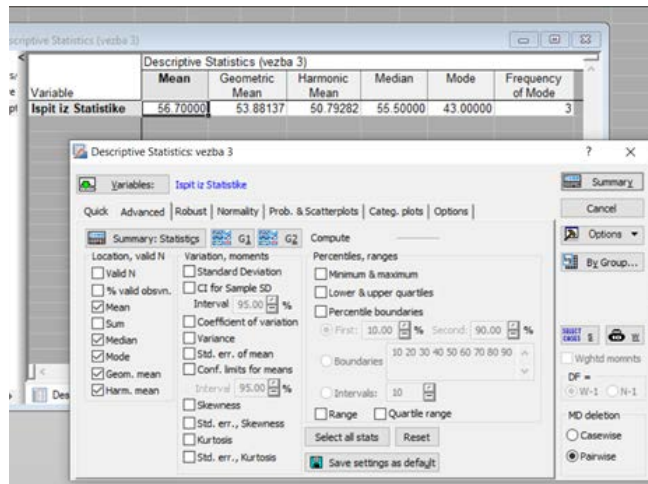
- Jednostavno se računa;

- Koristi se samo kod numeričkih serija;
- Nije pod uticajem ekstremnih vrednosti;
- „Grublja“ je procena od aritmetičke sredine, ali nekad objektivnija, npr.:

Primer 13. Zašto je nekad bolje, objektivnije, koristiti medijanu a ne aritmetičku sredinu: *“U našem preduzeću prosečna plata je 400 evra, $\mu = 400$. Preduzeće ima 6 radnika sa platama: 100, 100, 150, 150, 400, 1500. $M_e = 150$.”*

U slučaju simetrične raspodele podataka, vrednosti aritmetičke sredine, modusa i medijane se poklapaju.

Mere centralne tendencije u softveru *Statistica* moguće je računati samo za negrupisane (date prave) podatke čekiranjem traženih vrednosti u delu “Descriptive statistics”.



MERE DISPERZIJE

Mere centralne tendencije (aritmetička, geometrijska i harmonijska sredina, modus i medijana), ne objašnjavaju u potpunosti raspodelu određene serije podataka. Npr. Dve serije sa istom aritmetičkom sredinom mogu imati potpuno različite raspodele frekvencije, tj. njihova raspršenost može biti potpuno različita, npr.

a) 1,1,2,2,3,4,4,5,5.

b) 3,3,3,3,3,3,3,3,3.

Zbog toga se pored srednjih vrednosti uvode mere varijacije ili mere disperzije. To su pokazatelji varijacije (disperzije varijabiliteta) od neke srednje vrednosti. Odstupanje svake pojedinačne vrednosti od srednje vrednosti zove se devijacija, a odstupanje svih vrednosti od srednje vrednosti je disperzija.

Mere disperzije ili mere varijabiliteta ili mere raspršenosti podataka predstavljaju drugu grupu parametara kojima se opisuju rezultati nekog obeležja na datom statističkom skupu. Mere disperzije su pokazatelji relativnih i apsolutnih odstupanja vrednosti obeležja od neke mere centralne tendencije, najčešće od aritmetičke sredine.

Mere disperzije ukazuju na dve činjenice:

- Koliko su srednje vrednosti obeležja dobri predstavnici svih vrednosti na posmatranom skupu podataka;

- Koliko se sve vrednosti datog obeležja na posmatranom skupu međusobno razlikuju. Odnosno: Što je varijansa manja, to je srednja vrednost bolji predstavnik i obrnuto.

Mere disperzije su numeričke vrednosti kojima se izražava stepen varijabilnosti statističkih podataka. Unutar statističkog skupa.

Mere disperzije možemo podeliti u dve grupe:

1. Apsolutne mere disperzije;
2. Relativne mere disperzije.

Apsolutne mere disperzije

Apsolutne mere disperzije predstavljaju varijaciju u apsolutnim iznosima u istim mernim jedinicama kao dato obeležje (*m, kg, min, kom,...*). U ovu grupu mera spadaju:

- *Raspon (interval) varijacije (Rv)*;
- *Interkvartilno odstupanje (iq)*;
- *Srednje apsolutno odstupanje (srednja devijacija) (d)*;
- *Varijansa (σ^2)*;
- *Standardna devijacija (σ)*.

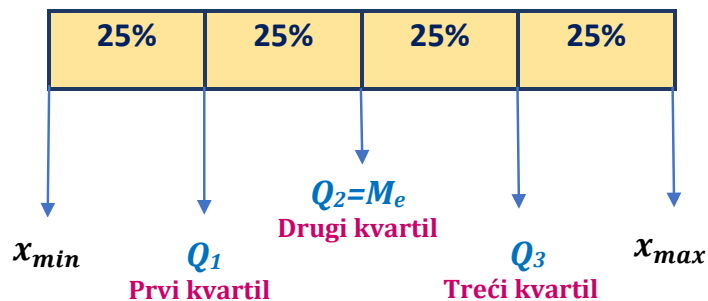
Apsolutne mere možemo podeliti u dve grupe: pozicione i izračunate, u zavisnosti od načina njihovog izračunavanja.



Raspon (interval) varijacije je poziciona apsolutna mera disperzije koja je jednaka razlici najveće i najmanje vrednosti u seriji podataka:

$$R_V = x_{max} - x_{min} .$$

Kvartili su tri deskriptivne mere Q_1, Q_2, Q_3 , koje dele rangiranu seriju podataka na četiri jednaka dela.



Kvartile, prvi (donji) i treći (gornji) računamo slično kao i drugi kvartil – medijanu. Potrebno je odrediti njihovu poziciju koja u ovom slučaju ne zavisi od parnosti ukupnog broja podataka. Takođe je potrebna rastuća kumulanta.

U slučaju grupisane serije podataka, nakon određivanja pozicija:

$$(\text{poz. } Q_1) = \frac{N}{4} ; \quad (\text{poz. } Q_3) = \frac{3N}{4} ,$$

donji i gornji kvartil računamo po sledećim formulama:

$$Q_1 = a_{Q_1} + \frac{i}{f_{Q_1}} ((\text{poz. } Q_1) - W_1) ,$$

$$Q_3 = a_{Q_3} + \frac{i}{f_{Q_3}} ((\text{poz. } Q_3) - W_3) ,$$

gde je:

- $a_{Q_{1,3}}$ = donja granica kvartilnog intervala,
- i = širina intervala,
- $f_{Q_{1,3}}$ = frekvencija kvartilnog intervala,
- $(\text{poz. } Q_{1,3})$ = pozicija kvartila,
- $W_{1,3}$ = kumulanta predkvartilnog intervala.

Interkvartilna razlika (odstupanje) je apsolutna poziciona mera disperzije jednaka razlici gornjeg i donjeg kvartila:

$$i_Q = Q_3 - Q_1 .$$

Interkvartilna devijacija je:

$$Q = \frac{Q_3 - Q_1}{2} .$$

Srednja devijacija je apsolutna izračunata mera disperzije jednaka aritmetičkoj sredini apsolutnih vrednosti odstupanja podataka (u slučaju negrupisanih), odnosno, razredne sredine intervala (u slučaju grupisanih) od aritmetičke sredine.

Za negrupisane podatke, srednja devijacija se računa po formuli:

$$d = \frac{1}{N} \sum_{i=1}^N |x_i - \mu| ,$$

odnosno, za grupisane podatke:

$$d = \frac{1}{N} \sum_{i=1}^k |x_i - \mu| f_i .$$

Varijansa (σ^2) ili fluktoacija je apsolutna izračunata mera disperzije jednaka aritmetičkoj sredini kvadrata odstupanja podataka (u slučaju negrupisanih), odnosno, razredne sredine intervala (u slučaju grupisanih) od aritmetičke sredine.

Za negrupisane podatke, varijansa se računa po formuli:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 ,$$

odnosno, za grupisane podatke:

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

Standardna devijacija (σ) je apsolutna izračunata mera disperzije jednaka pozitivnoj vrednosti kvadratnog korena varijanse. Njena vrednost je u intervalu $[0, +\infty)$.

$$\sigma = +\sqrt{\sigma^2}$$

Standardna devijacija predstavlja najznačajniju meru disperzije koja pokazuje, procenjuje koliko vredi aritmetička sredina (koliko je odstupanje od nje).

Relativne mere disperzije

Relativne mere disperzije predstavljaju varijaciju u procentima u odnosu na neku veličinu. U ovu grupu mera spadaju:

- *Koeficijent varijacije* (K_v);
- *Relativna varijansa* (V^2);
- *Relativno linearno odstupanje* (R_{lo});
- *Koeficijent interkvartilne devijacije* (V_Q);
- *Normalizovano (standardizovano) odstupanje (Z-vrednosti)* (Z).

Koeficijent varijacije (K_v) predstavlja procentualni odnos standardne devijacije i aritmetičke sredine.

$$K_v = \frac{\sigma}{\mu} \cdot 100 .$$

Koeficijent varijacije prikazuje prosečno odstupanje od aritmetičke sredine. Koristi se kod poređenja varijabilnosti različitih skupova. Vrednost K_v od 30% je granična vrednost ispod koje se statistički skup smatra *homogenim* a iznad je *heterogen*.

Relativna varijansa (V^2) predstavlja odnos varijanse i kvadrata aritmetičke sredine.

$$V^2 = \frac{\sigma^2}{\mu^2} .$$

Relativno linearno odstupanje (R_{lo}) predstavlja procentualni odnos srednje devijacije i aritmetičke sredine.

$$R_{lo} = \frac{d}{\mu} \cdot 100 .$$

Koeficijent interkvartilnog odstupanja (V_Q) predstavlja odnos interkvartilne razlike i zbira kvartila.

$$V_Q = \frac{Q_3 - Q_1}{Q_3 + Q_1} .$$

Koeficijent interkvartilnog odstupanja pripada intervalu (0,1). Što je vrednost bliža nuli, stepen homogenosti serije je veći i obrnuto.

Koeficijent interkvartilnog odstupanja je zapravo količnik apsolutne mere disperzije i mere centralne tendencije:

$$V_Q = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{\frac{1}{2} (Q_3 - Q_1)}{\frac{1}{2} (Q_3 + Q_1)} = \frac{Q}{M_e} ,$$

odnosno, količnik interkvartilne devijacije i medijane.

Normalizovano odstupanje (Z) predstavlja odstupanje pojedinačnog podatka od aritmetičke sredine u odnosu na standardnu devijaciju:

$$Z = \frac{x_i - \mu}{\sigma}$$

U softveru *Statistica* moguće je računati samo za negrupisane (date prave) podatke većinu navedenih mera disperzije čekiranjem traženih vrednosti u delu “Descriptive statistics”.

Variable	Minimum	Maximum	Lower Quartile	Upper Quartile	Range	Quartile Range	Variance	Std. Dev.	Coef. Var.	Standard Error
Ispit iz Statistike	22.00000	91.00000	44.00000	68.00000	69.00000	24.00000	299.1939	17.29722	30.50656	2.446197

The screenshot shows the 'Descriptive Statistics: vezba 3' dialog box in Statistica. The 'Compute' section is active, showing options for 'Minimum & maximum', 'Lower & upper quartiles', and 'Percentile boundaries'. The 'Range' and 'Quartile range' options are checked. The 'Interval' is set to 95.00%.

MERE OBLIKA

Dve statističke serije mogu imati iste aritmetičke sredine, varijanse, srednje devijacije i koeficijente varijacije, ali različite rasporede frekvencija (raspodele).

Mere disperzije opisuju varijacije u vrednostima, ali ne pokazuju smer varijacija u odnosu na aritmetičku sredinu, kao ni stepen koncentracije vrednosti oko aritmetičke sredine, tj. oblik rasporeda frekvencija.

Raspored je **simetričan** kada frekvencije vrednosti obeležja ravnomerno opadaju ili rastu u odnosu na aritmetičku sredinu levo i desno. Numerička serija je **asimetrična** ako frekvencije vrednosti obeležja pokazuju tendenciju grupisanja više levo ili više desno od aritmetičke sredine.

U zavisnosti od odnosa frekvencija, raspored podataka je više ili manje spljošten.

Postoje dva prvca u analizi oblika:

1. Ispitivanje simetrije podataka;
2. Ispitivanje visine rasporeda frekvencija.

Na osnovu toga mere oblika delimo na:

1. **Mere asimetrije;**
2. **Mere spljoštenosti.**

Za određivanje mera u statistici koriste se odstupanja pojedinačnih vrednosti obeležja od aritmetičke sredine stepenovana određenim stepenom, tzv. centralni momenti rasporeda (M). Prvi centralni moment koristi se u izračunavanju srednje devijacije, drugi u izračunavanju varijanse i standardne devijacije.

Za merenje asimetrije koristi se treći centralni moment (M_3), a za merenje spljoštenosti koristi se četvrti centralni moment (M_4).

Mere asimetrije

Mere asimetrije predstavljaju mere rasporeda članova statističkog skupa prema određenoj vrednosti, (najčešće prema aritmetičkoj sredini), odnosno, prema osi simetrije.

Raspored je:

- **Simetričan** – ako svakom odstupanju vrednosti od aritmetičke sredine negativnog predznaka odgovara jednak broj odstupanja pozitivnog predznaka;
- **Pozitivno asimetričan** – ako preovladava odstupanje pozitivnog predznaka;
- **Negativno asimetričan** – ako preovladava odstupanje negativnog predznaka.

Za merenje asimetrije polazna veličina je treći centralni moment M_3 , koji se za negrupisane podatke računa po formuli:

$$M_3 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^3,$$

odnosno, za grupisane podatke:

$$M_3 = \frac{1}{N} \sum_{i=1}^k (x_i - \mu)^3 f_i.$$

Veličina asimetrije izražava se **koeficijentom asimetrije (α_3)** koji se naziva prvi *Pearson*-ov koeficijent (u softveru "Skewness") i izračunava formulom:

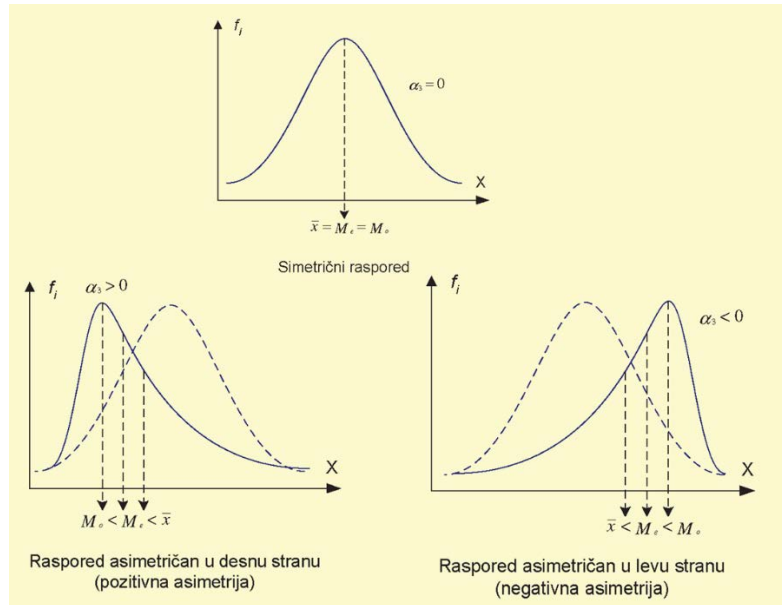
$$\alpha_3 = \frac{M_3}{\sigma^3}.$$

Na osnovu dobijenih vrednosti koeficijenta asimetrije, važi:

- $\alpha_3 = 0 \Rightarrow$ **serija je simetrična;**
- $\alpha_3 > 0 \Rightarrow$ **serija je pozitivno asimetrična;**
- $\alpha_3 < 0 \Rightarrow$ **serija je negativno asimetrična.**

Takođe se može analizirati i jačina asimetrije:

$$\begin{aligned} |\alpha_3| < 0.1 & \text{ nema asimetrije} \\ 0.1 \leq |\alpha_3| < 0.25 & \text{ mala asimetrija} \\ 0.25 \leq |\alpha_3| < 0.5 & \text{ srednja asimetrija} \\ |\alpha_3| \geq 0.5 & \text{ jaka asimetrija} \end{aligned}$$



Grafički prikaz simetričnih, odnosno, asimetričnih podataka

Asimetriju možemo ispitivati (ali ne i njenu jačinu) preko:

- *Pearson*-ove mere asimetrije koja se temelji na odnosu aritmetičke sredine, modusa i medijane;
- *Bowley*-jeve mere asimetrije koja se temelji na odnosu kvartila i medijane.

***Pearson*-ova mera asimetrije:**

(Karl Pearson, 1857-1936, British statistician)

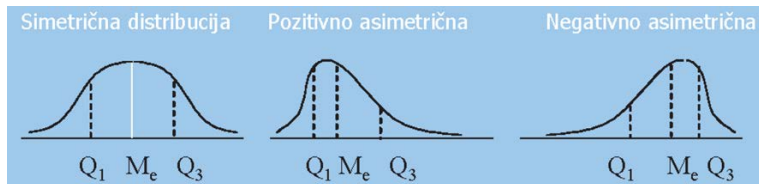
- $\mu = M_e = M_o \Rightarrow$ serija je simetrična;
- $\mu > M_e > M_o \Rightarrow$ serija je pozitivno asimetrična;
- $\mu < M_e < M_o \Rightarrow$ serija je negativno asimetrična.



Bowley- jeva mera asimetrije:

(Sir Arthur Lyon Bowley, 1869-1957, British statistician)

- $Q_3 - M_e = M_e - Q_1 \Rightarrow$ serija je simetrična;
- $Q_3 - M_e > M_e - Q_1 \Rightarrow$
serija je pozitivno asimetrična;
- $Q_3 - M_e < M_e - Q_1 \Rightarrow$
serija je negativno asimetrična.



Mere spljoštenosti

Mere spljoštenosti (zaobljenosti) upotpunjuju sliku o izgledu distribucije (raspodele). Predstavljaju mere zaobljenosti raspodele u okolini modalnog vrha. Zaobljenost se upoređuje i meri u odnosu na zaobljenost modalnog vrha – normalne ili *Gausove raspodele* (najpoznatije i najvažnije teorijske distribucije).

Raspodela može biti:

- **Iste spljoštenosti** kao normalna raspodela;
- **Izdužena** u odnosu na normalnu raspodelu;
- **Spljoštena** u odnosu na normalnu raspodelu.

Za merenje spljoštenosti polazna veličina je četvrti centralni moment M_4 , koji se za negrupisane podatke računa po formuli:

$$M_4 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^4,$$

odnosno, za grupisane podatke:

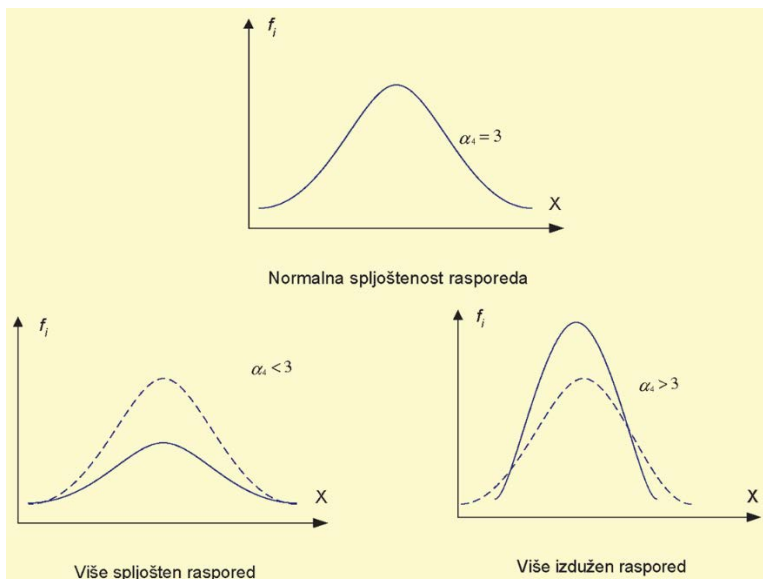
$$M_4 = \frac{1}{N} \sum_{i=1}^k (x_i - \mu)^4 f_i.$$

Veličina spljoštenosti izražava se **koeficijentom spljoštenosti** (α_4) koji se naziva drugi *Pearson*-ov koeficijent i izračunava formulom:

$$\alpha_4 = \frac{M_4}{\sigma^4} .$$

Na osnovu dobijenih vrednosti koeficijenta spljoštenosti, važi:

- $\alpha_4 = 3 \Rightarrow$ **normalna spljoštenost;**
- $\alpha_4 > 3 \Rightarrow$ **izdužena raspodela;**
- $\alpha_4 < 3 \Rightarrow$ **spljoštena raspodela.**



Grafički prikaz spljoštenosti raspodele

U softveru *Statistica* se kao mera spljoštenosti posmatra veličina “Kurtosis” i on analizira u odnosu na nulu jer važi da je:

$$k = \alpha_4 - 3.$$

U tom slučaju važe relacije:

- $k = 0 \Rightarrow$ normalna spljoštenost;
- $k > 0 \Rightarrow$ izdužena raspodela;
- $k < 0 \Rightarrow$ spljoštena raspodela.

Primer 14. Za datu statističku seriju podataka *Ispit iz statistike* (iz *Primer 7.*) ispitati asimetriju, odrediti njenu jačinu i oblik raspodele.

		Descriptive Statistics (vezba 3)	
Variable	Skewness	Kurtosis	
Ispit iz Statistike	0.060157	-0.630758	

U softveru *Statistica* nakon izračunavanja potrebnih veličina, zaključak je da je serija podataka pozitivno asimetrična ali kako je apsolutna vrednost manja od 0,1, asimetrija se zanemaruje, smatra se da nema asimetrije. Negativna vrednost “Kurtosis” pokazuje da seriju karakteriše spljoštena raspodela.

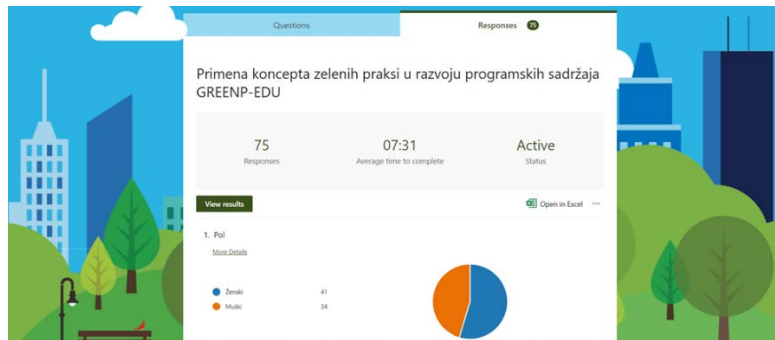
ANKETA I OBRADA PODATAKA

U okviru projekta “Primena koncepta zelenih praksi u razvoju programskih sadržaja” (GREENP-EDU), koji se finansira od strane Ministarstva prosvete, nauke i tehnološkog razvoja RS, u okviru programske aktivnosti “Razvoj visokog obrazovanja”, sprovedena je online anketa preko *Microsoft* platforme na temu projekta i ispitano javno mnjenje studenata osnovnih i master strukovnih studija iz oblasti zaštite životne sredine Akademije tehničko-vaspitačkih strukovnih studija, kao i pojedinih profesora.

Pored osnovnog cilja ankete dobijanja povratne informacije o znanju i mišljenju studenata strukovnih studija o primeni koncepta zelenih praksi, odnosno, o sprovođenja mera u oblasti sprečavanja klimatskih promena i zagađenja, razvoja energije i biodiverziteta, mobilnosti i cirkularne ekonomije i o unapređenju životne sredine u globalu, cilj ankete je bio i korišćenje alata za statističku obradu dobijenih (generisanih) podataka za dalju, svrsishodniju analizu, od strane studenata druge godine osnovnih studija studijskog programa *Zaštita životne sredine* u okviru predmeta **Statistika i analiza**.

Primena koncepta zelenih praksi – rezultati ankete

Anketirano je ukupno 75 ispitanika, u okviru 30 pitanja, od kojih je 5 bilo vezano za opšte informacije o ispitaniku. Ostala pitanja vezana za temu ispitivanja bila su sa ponuđenim odgovorima ili skalom ocenjivanja, a poslednje, predlog lokalnoj samoupravi za unapređenje kvaliteta životne sredine, bilo je sa slobodnim odgovorom.



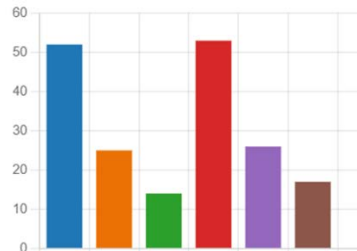
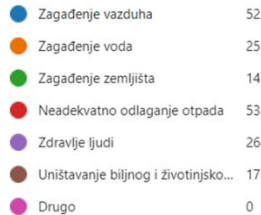
Interfejs online ankete

Najviše anketiranih je imalo između 20 i 29 godina, 55% ženskog i 45% muškog pola, sa završenom srednjom školom, bez radnog statusa, odnosno, po zanimanju student, većina iz Niša i okoline ali i iz delova južne, jugoistočne i zapadne Srbije.

Među najznačajnijim ekološkim problemima mesta u kome živi, anketirana populacija navela je neadekvatno odlaganje otpada i zagađenje vazduha:

6. Koji su najznačajniji ekološki problemi mesta u kome živite? Možete izabrati do tri odgovora.

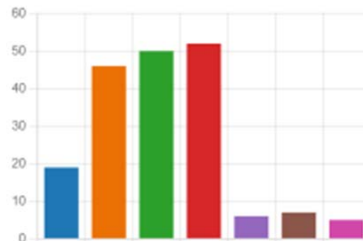
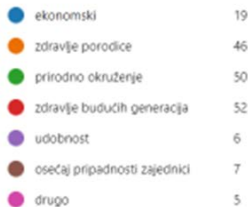
[More Details](#)



Negativni uticaj problema životne sredine najviše pogađa zdravlje budućih generacija a životna sredina mesta iz koga dolaze označena je kao zagađena.

7. Na koji način problemi životne sredine negativno utiču na Vaš život i život Vaše porodice? Možete da odaberete do tri odgovora.

[More Details](#)



Više od polovine anketiranih smatra da stanuje u energetski efikasnom domu i njih 56% je zainteresovano da sprovede još mera po tom pitanju.

Na skali od 1 (nimalo) do 5 (potpuno) u kojoj meri ste spremni

da promenite svoj životni stil/komfor kako biste ublažili klimatske promene i zagađenje, uglavnom je prosečna ocena bila 4 za sledeće ponuđene opcije:

- 1) Kada sam u mogućnosti biram proizvode koji su skuplji, ali ekološki prihvatljivi (ne zagađuju, štede struju i vodu i dr.) – **prosečna ocena 3,69**;
- 2) Trudim se da smanjim upotrebu jednokratnih proizvoda kad god je to moguće (npr. plastične kese, cevčice, pribor za jelo i dr. – **prosečna ocena 3,93**;
- 3) Kupujem lokalne i sezonske proizvode kada god je to moguće– **prosečna ocena 3,81**;
- 4) Trudim se da smanjim otpad– **prosečna ocena 4,25**;
- 5) Redovno razdvajam otpad za reciklažu– **prosečna ocena 3,37**;
- 6) Pri kupovini kućnih aparata, npr. mašine za veš, frižidera ili televizora, biram one koji su energetske efikasniji– **prosečna ocena 3,93**;
- 7) Retko koristim svoj automobil i idem peške, biciklom ili javnim prevozom kada je to moguće – **prosečna ocena 3,15**.

Za 5 godina polovina anketiranih smatra da će zagađenje biti veće, a čak 92% da je potrebno učešće građana u donošenju odluka o životnoj sredini.

Kako bi pomogli lokalnoj zajednici da se izbori sa ekološkim problemima, anketirani najčešće predlažu razgovore s ljudima i pomoć u organizaciji akcija.

27. Za 5 godina zagađenje će biti

[More Details](#)

[Insights](#)

● Manje	4
● Veće	38
● Nepromenjeno	8
● Nisam siguran/na	25



28. Da li je potrebno učesće građana u donošenju odluka o životnoj sredini?

[More Details](#)

[Insights](#)

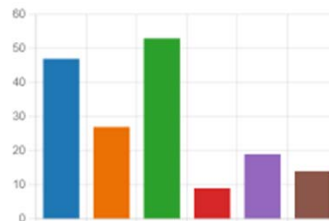
● Ne	2
● Da	69
● Ne znam	4



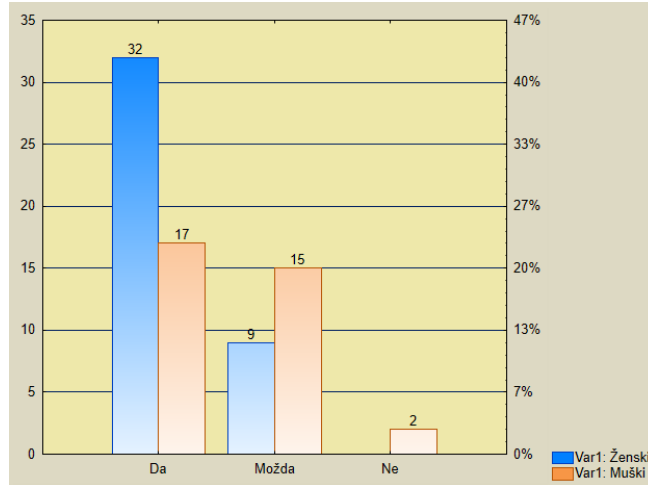
29. Kako biste mogli pomoći lokalnoj zajednici da se izbori sa ekološkim problemima? Možete izabrati do 3 odgovora.

[More Details](#)

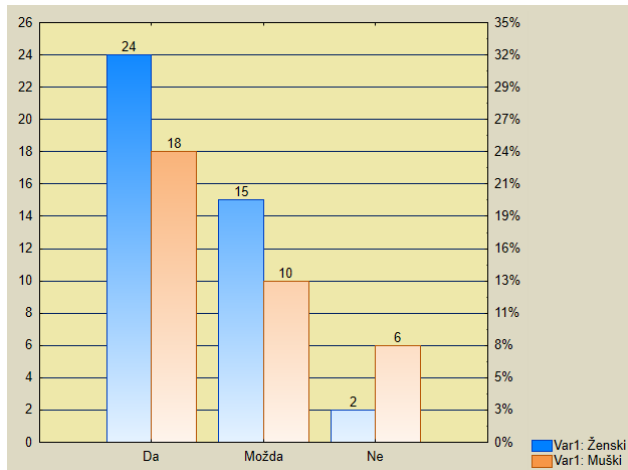
● Pomoć u organizaciji akcija	47
● Priprema informacija	27
● Razgovori sa ljudima	53
● Telefoniranje	9
● Pisanje članka (u lokalnim novin...	19
● Drugo	14



Detaljnijom analizom studenata u softveru *Statistica* došlo se do zaključka da je ženska populacija spremnija po pitanju aktiviranja u pomoći zajednici da se izbori sa ekološkim problemima, ali i zainteresovanija da sprovede mere energetske efikasnosti u domaćinstvu.

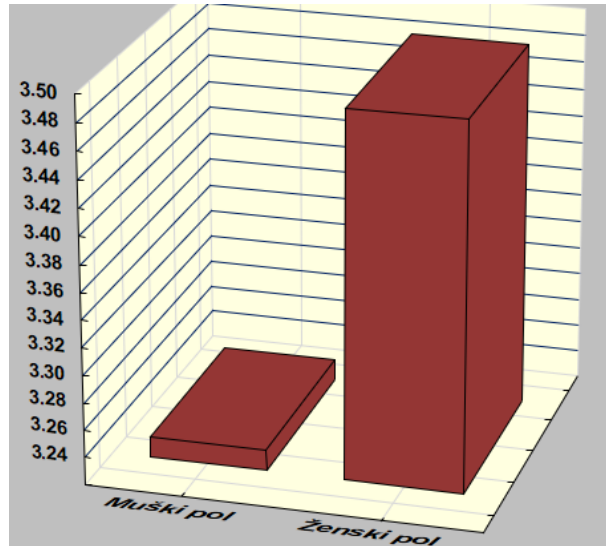


Histogram o spremnosti aktiviranja, po polu

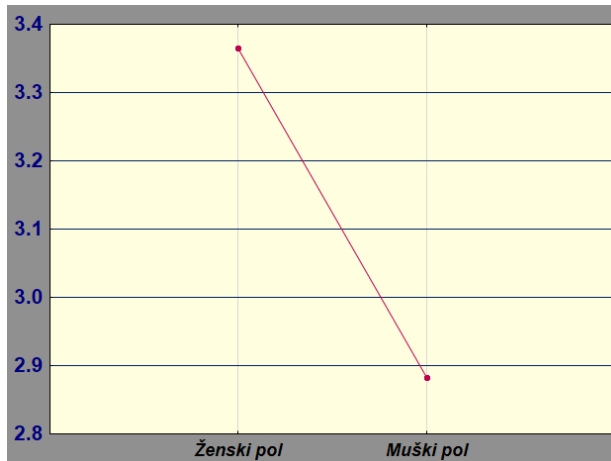


Histogram o zainteresovanosti za sprovođenje mera, po polu

Takođe, ženski pol prednjači u razdvajanju otpada za reciklažu:



Ali i pri prevozu, ženski pol ređe koristi automobil i ide peške, biciklom ili javnim prevozom kada je to moguće:



LITERATURA

- [1] Popović B., Nastić A., Đorđević M.: **Zbirka zadataka iz matematičke statistike**, Univerzitet u Nišu, Prirodno-matematički fakultet, Niš, 2014.
- [2] Aleksić V., Stanković M.: **Elementi teorije verovatnoće i matematičke statistike**, Univerzitet u Nišu, Fakultet zaštite na radu, Niš, 2012.
- [3] Benšić M., Šuvak N.: **Primijenjena statistika**, Sveučilište J. J. Strossmayera u Osijeku, Odjel za matematiku, Osijek, 2013.
- [4] Šekarić M.: **Statističke metode**, Univerzitet Singidunum, Beograd 2010.

