GLAVA III

PARAMETRI RASPODELA

Najčešće osnovni statistički skupovi, (S) kao osnovni predmeti statističkog istraživanja su veoma brojni, imaju veliki broj elemenata N. Samo za jedno posmatrano i mereno obeležje X na skupu S dobijaju se velike distribucije podataka, rezultata merenja, uređenih ili kao prosta serija $x_1, x_2, ..., x_i, ..., x_n$ različitih vrednosti ili kao grupisana distribucija apsolutnih frekvencija datih sa (3) i u tabeli 1, ili kao intervalna distribucija apsolutnih frekvencija data u tabeli 3. i poseban primer u tabeli 4. To mnoštvo podataka za veliko N je i za moćne računare veliko opterećenje za memoriju, za korišćenje. To mnoštvo podataka o pojavi na S je tako nepregledno i ne pruža informacije o izučavanoj pojavi, a još manje o zakonitostima promena pojava. Ali iz raspodela (3) i onih datih u tabeli 1. je uočljivo da su vrednosti obeležja najčešće raspoređena tako da se njihove frekvencije koncentrišu oko sredine između najmanje i najveće vrednosti, a ukoliko su udaljenije od sredine frekvencije su im sve manje. S obzirom na takvu tendenciju raspodele možemo govoriti da su izvesne vrednosti više tipične za to obeležje nego druge. Isto se može uočiti stepen koncentracije oko tih vrednosti. Pokazatelji raspodele frekvencija obeležja X koji pokazuju ceo osnovni skup nazivaju se parametri skupa ili parametri raspodele. Postoje tri vrste parametara skupa i to su: srednje vrednosti ili mere centralne tendencije, druge su mere disperzije ili mere varijacije obeležja i treće su mere oblika raspodele, poimane kao grafički prikazi osnovnih podataka u sistemu xOy. Vrlo često se pod osnovnim skupom S poima i skup podataka, rezultata merenja X na S, a datih sa raspodelama frekvencija (2) ili (3) ili serije u tabeli 3. i dr.

1. Srednje vrednosti - mere centralne tendencije

Raspodele obeležja pokazuju strukturu osnovnog skupa u odnosu na posmatrano obeležje X, iskazuje grupisanje jedinica skupa oko pojedinih vrednosti. Srednja vrednost je osnovni pokazatelj centralne tendencije, pokazuje lokaciju osnovnog skupa. To je najznačajniji parametar skupa obeležja X a omogućuje polazno poređenje između dva i više statističkih skupova i svaku analizu podataka.

Srednje vrednosti su aritmetička sredina, geometrijska sredina i harmonijska sredina. Zavise od svih vrednosti obeležja. Svaka od njih veća je od najmanje vrednosti obeležja, a manja od najveće vrednosti obeležja, a ukoliko su sve vrednosti obeležja medusobno jednake i srednje vrednosti jednake su toj vrednosti.

1.1 Aritmetička sredina

Za osnovni skup označava se sa μ. ("mi").

Dobija se za prostu, negrupisanu seriju kad se zbir svih vrednosti obeležja podeli sa njihovim brojem. Za obeležje X sa vrednostima $x_1, x_2, ..., x_N$ aritmetička sredina je

$$\mu = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \text{ili} \quad \mu = \frac{\sum_{1}^{N} x_i}{N} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Prethodni izraz (1) je izraz za prostu aritmetičku sredinu.

Za grupisanu raspodelu frekvencija oblika 3 aritmetička sredina obeležja *X* je

$$\mu = \frac{f_1 \cdot x_1 + f_2 \cdot x_2 + \ldots + f_n x_n}{N} = \frac{1}{N}\sum_{i=1}^{N} f_i \cdot x_i \quad \text{ili} \quad \mu = \frac{\sum_{1}^{N} f_i \cdot x_i}{N} \qquad \ldots\ldots\ldots\ldots\ldots (2)$$

Izraz (2) naziva se ponderisana aritmetička sredina, gde su ponderi $f_1, f_2, \ldots, f_n$ takvi da važi $f_1 + f_2 + \ldots + f_n = N$ i znače učestalost odgovarajućih vrednosti $x_1, x_2, \ldots, x_n$ obeležja *X* na *S*.

Kad se aritmetička sredina obeležja *X* određuje na uzorku $U \subset S$, osnovnog skupa *S*, od *n* elemenata, tada se aritmetička sredina označava sa $\overline{X}$ ili $\overline{x}$ (iks bar) i izračunava po formuli

$$\overline{X} = \frac{x_1 + x_2 + \ldots + x_n}{n} \quad \text{za negrupisanu seriju i}$$

$$\overline{X} = \overline{x} = \frac{f_1 \cdot x_1 + f_2 \cdot x_2 + \ldots + f_k \cdot x_k}{n}, \text{ za } f_1 + f_2 + \ldots + f_k = n$$

za grupisanu seriju iz uzorka.

Primeri:

1. Neka je osnovni skup sastavljen od brojeva: 20, 30, 40, 50, 60, 70, 80 ; N = 7 i aritmetička sredina jednaka je

$$\mu = \frac{20 + 30 + 40 + 50 + 60 + 70 + 80}{7} = \frac{350}{7} = 50.$$

2. Osnovni skup predstavlja 400 čeličnih osovina uređenih u odnosu na prečnik osovine $X_{(mm)}$.

Tabela 1.

X :	14,97	14,98	14,99	15,00	15,01	15,02	15,03
f :	40	60	50	100	80	50	20

Srednja vrednost prečnika osovine dobija se po formuli (2) jednaka je

$$\mu = \frac{40 \cdot 14{,}97 + 60 \cdot 14{,}98 + 50 \cdot 14{,}99 + 100 \cdot 15{,}00 + 80 \cdot 15{,}01 + 50 \cdot 15{,}02 + 20 \cdot 15{,}03}{400} =$$

$$\mu = 14{,}99875 mm.$$

Pokazuje se da je zbir svih odstupanja pojedinih vrednosti statističkog skupa od aritmetičke sredine jednaka 0, naime da važe jednakosti za prostu seriju podataka

$$\sum_{i=1}^{N}(x_i - \mu) = 0$$

i za grupisanu raspodelu frekvencija

$$\sum_{i=1}^{N} f_i(x_i - \mu) = 0 \, .$$

Aritmetička sredina dobro reprezentuje one raspodele kod kojih velike frekvencije imaju vrednosti obeležja koja su najbliža aritmetičkoj sredini. Aritmetička sredina se ne može izračunati za raspodele sa otvorenim grupnim intervalima kod koga su nepoznate sredine tih intervala.

1.2. Geometrijska sredina

Pod uslovom da su sve vrednosti obeležja *X* veće od nule, $x_i > 0$ za $\forall = 1,2,\ldots,N$, geometrijska sredina za prostu seriju podataka, u oznaci *G*, jednaka je

$$G = \sqrt[N]{x_1 \cdot x_2 \ldots x_N} \quad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

To znači da se geometrijska sredina dobija kad se izmnože vrednosti obeležja i izvuče *N*-ti koren iz dobijenog proizvoda.

Za grupisanu seriju podataka (3) ili intervalne distribucije geometrijska sredina je

$$G = \sqrt[N]{x_1^{f_1} \cdot x_2^{f_2} \ldots x_n^{f_n}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (4)$$

uz uslov da je

$$f_1 + f_2 + \ldots + f_n = N$$

Izračunavanje geometrijske sredine vrši se pomoću logaritama, tj. za (3)

$$\log G = \frac{1}{N}(\log x_1 + \log x_2 + \ldots + \log x_n)$$

$$\log G = \frac{1}{N}\sum_{i=1}^{N}\log x_i \ \text{ i } \ G = anti\log\left(\frac{1}{N}\sum_{i=1}^{N}\log x_i\right)$$

a za (4) dobija se da je

$$G = anti\log\left(\frac{1}{N}\sum_{i=1}^{n} f_i \cdot \log x_i\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

Pokazuje se da je za jednu istu raspodelu frekvencija aritmetička sredina veća od geometrijske sredine, tj. da važi odnos

$$G \leq \mu \, .$$

Geometrijska sredina se upotrebljava tada kad se želi da iskaže dirtamika promene obeležja X stopa rasta na osnovu lančanih indeksa, a kod vremenskih serija se koristi za izračunavanje srednjih vrednosti indeksnih brojeva.

Za osnovni skup brojeva: (4, 8, 16, 32, 64), N = 5 geometrijska sredina je

$$G = \sqrt[5]{4\cdot 8\cdot 16\cdot 32\cdot 64} = \sqrt[5]{2^2\cdot 2^3\cdot 2^4\cdot 2^5\cdot 2^6} = \sqrt[5]{2^{20}} = 2^4 = 16.$$

Nije teško sagledati da dati skup brojeva čini geometrijsku progresiju sa količnikom $q = 2$, pa je geometrijska sredina 16.

1.3 Harmonijska sredina

Za prostu seriju podataka kod koje je

$$x_i \neq 0 \text{ za }, \ \forall = 1,2,\ldots,N .$$

Harmonijska sredina, u oznaci H, jeste recipročna vrednost aritmetičke sredine recipročnih vrednosti obeležja i važi formula

$$H = \frac{H}{\frac{1}{x_1}+\frac{1}{x_2}+\ldots+\frac{1}{x_N}} = \frac{H}{\sum_{i=1}^{N}\frac{1}{x_i}} \quad \ldots\ldots (6)$$

Za grupisanu raspodelu fekvencija važi gormula

$$H = \frac{H}{\frac{f_1}{x_1}+\frac{f_2}{x_2}+\ldots+\frac{f_n}{x_n}} = \frac{H}{\sum_{i=1}^{N}\frac{f_i}{x_i}} \quad \ldots\ldots (7)$$

Ne primenjuje se za male vrednosti članova serije. Ima ograničenu primenu u ekonomiji. Izračunavaju se kod određivanja srednjih vrednosti normi za koje se jedan te isti posao može završiti od različitih izvodilaca radova.

1.4 Medijana

Medijana je ona vrednost statističkog obeležja, u oznaci Me, koja polovi statističku masu na dva jednaka dela, što znači da pola jedinica osnovnog skupa ima od Me manju vrednost obeležja x_i, $\left(x_i < Me, \text{za } i = 1,2,\ldots,\frac{N}{2}-1\right)$, a pola jedinica skupa ima vrednost obeležja veću od Me $\left(Me < x_j, \text{za } j = \frac{N}{2}+1, \frac{N}{2}+2,\ldots,N\right)$.

Ako se radi o prostoj seriji od N rezultata merenja, gde je $N = 2n+1$ neparan broj i oni su uređeni po rastućem redosledu

$$\underbrace{x_1, x_2, x_3,\ldots,x_n}_{n}, x_{n+1}, \underbrace{\ldots,x_{2n+2},\ldots,x_{2n+1}}_{n},$$

tada je medijana $Me = x_{n+1}$ jednaka srednjem članu serije.

Ako serija ima paran broj članova $N = 2n$ uređenih po rastućem redosledu

$$\underbrace{x_1, x_2, x_3, \ldots, x_{n-1}}_{n-1\,\text{član}}, x_n, x_{n+1}, \underbrace{x_{2n+2}, x_{2n+3} \ldots, x_{2n}}_{n-1\,\text{član}},$$

onda ta serija ima dva srednja člana x_n i x_{n+1} i za medijanu se uzima svaki broj koji je veći od x_n i manji od x_{n+1} a najčešće medijana je

$$Me = \frac{x_n + x_{n+1}}{2}, \text{ tj.}$$

aritmetička sredina srednjih članova.

To znači za prostu seriju podataka

11,19,24,31,49,52,57

medijana je $Me = 31$, a za prostu seriju parnog broja podataka

15,22,29,38,41,48,51,53

medijana je $Me = \dfrac{38 + 41}{2} = 39{,}5$.

Za grupisanu raspodelu frekvencija obeležja X uređenu po rastućem redosledu u obliku

X	x_1	x_2	$\cdots$	x_n
f	f_1	f_2	$\cdots$	f_n

$$\left(\sum_{i=1}^{n} f_i = N \right)$$

ako je N neparan broj, tada je medijana obeležja X ona vrednost u nizu $x_1, x_2, \ldots x_n$ kojoj odgovara frekvencija koja sabrana sa svim prethodnim frekvencijama daje najmanji broj koji nije manji od $\dfrac{N+1}{2}$.

Zato se napravi kumulanta raspodele dok ne dobijemo indeks k za koji važe nejednakosti

$$f_1 + f_2 + \ldots + f_{k-1} < \frac{N+1}{2} \text{ i}$$

$$f_1 + f_2 + \ldots + f_k \geq \frac{N+1}{2}.$$

Sada je medijana

$$Me = x_k.$$

Ako je N paran broj tada je medijana jednaka

$$M = \frac{x_{\frac{N}{2}} + x_{\frac{N}{2}+1}}{2}.$$

Ako imamo intervalnu distribuciju frekvencija obeležja X prikazanu tabelom 3 (iz 2.5 Grafički prikaz raspodele), tada se toj tabeli doda vrsta za kumulativnu frekvenciju.

Tabela 1.

Obeležje X	$(b_0, b_1]$	$(b_1, b_2]$	$(b_2, b_3]$	...	$(b_i, b_{i+1}]$	...	$(b_{n-1}, b_n]$
Sredine intervala x_1	x_1	x_2	x_3	...	x_i	...	x_n
Apsolutna frekvencija f	f_1	f_2	f_3	...	f_i	...	f_n
Σ	f_1	$f_1 + f_2$	$f_1 + f_2 + f_3$	...	$f_1 + f_2 + ... + f_i$	...	N

Tada se medijana Me izračunava po formuli:

$$Me = L_1 = \frac{\frac{N}{2} - \sum_{1}^{k-1} f_i}{f_{Me}} \cdot i \quad \text{...} (8)$$

gde je L_1 donja granica medijalnog podintervala, f_{Me} je frekvencija medijalnog podintervala, i je dužina podintervala, a $\sum f_i$ je zbir frekvencija do medijalnog podintervala.

Primer:

1. Data je intervalna distribucija obeležja X, gde je X veličina preseka jednog metalnog profila u mm u obliku tabele 2.

Tabela 2.

X u mm	Broj profila		Sredina x_i grupnog intervala	$f_i x_i$	$f_i x_i^2$
	f_i	$\sum f_i$			
[1,51;2,5)	3	3	2	6	12
[2,51;3,5)	6	9	3	18	54
[3,51;4,5)	7	16	4	28	112
[4,51;5,5)	12	28	5	60	300
[5,51 ;6,5)	9	37	6	54	324
[6,51;7,5)	8	45	7	56	392
[7,51;8,5)	2	47	8	16	128
[8,51;9,5)	2	49	9	18	162
[9,51;10,5)	1	50	10	10	100
	50			266	1584

Za seriju podataka u tabeli 2. odrediti medijanu Me po formuli (8).

Ovde je medijalni podinterval [4,51;5,5), kome odgovara $L_1 = 4{,}51$; $f_{Me} = 12$; $\sum f_i = 16$; $i = 0{,}99$; $\frac{N}{2} = 25$.

Tražena medijana je jednaka

$$Me = 4{,}51 + \frac{25-16}{12} \cdot 0{,}99 = 5{,}2425.$$

Medijana se grafički dobija približno kad se kriva kumulativne frekvencije preseče sa pravom $\frac{N}{2}$, koja je paralelna osi obeležja X i tačka preseka sa kumulantom projektuje na osu obležja X .

1.5 Moda ili Modus raspodele

Moda ili modus raspodele je ona vrednost obeležja koja u datoj seriji podataka ima najveću frekvenciju. Zbog toga je moda tipična vrednost serije podataka. Ako je srednja vrednost obeležja tipična vrednost obeležja tada se najčešće poima da je to moda. Oznaka za modu je M_o. U intervalnoj raspodeli frekvencija obeležja X gde je X veličina preseka datog metalnog profila, prikazanoj u tabeli 2. moda je $M_o = 5$, jer njoj odgovara najveća učestalost $f_{M_o} = 12 > f_i$ za $\forall_i \neq 4$.

Kad u jednoj raspodeli frekvencija imamo samo jednu vrednost sa najvećom frekvencijom tada je ta raspodela unimodalna, a ako imaju dve ili više vrednosti obeležja sa najvećim frekvencijama, tada se ta raspodela naziva bimodalnom ili polimodalnom raspodelom. Ako posmatramo prostu seriju podataka $x_1, x_2, \ldots, x_N$ sa $f_i = 1$ $\forall = 1,2,\ldots,N$ kažemo da nema modu ili da nije modalna raspodela.

Za grupisane serije podataka moda M_o se određuje po formuli

$$M_o = L_1 + \frac{f_2 - f_1}{(f_2 - f_1) + (f_2 - f_3)} \cdot i \quad \ldots\ldots\ldots\ldots (9)$$

gde je L_1 donja granica modalnog intervala, f_2 je frekvencija modalnog intervala, f_1 je frekvencija predmodalnog, a f_3 frekvencija poslemodalnog intervala.

Ova formula je naravno formula za modu za intervalne distribucije.

Za raspodelu frekvencija obeležja X u obliku:

$$\left.\begin{array}{llllll} X: & x_1, & x_2, & \ldots, & x_k & \ldots, \; x_n \\ f: & f_1, & f_2, & \ldots, & f_k & \ldots, \; f_n \end{array}\right\}$$

moda je $M_o = x_k$ ako je

$$f_k > f_i \text{, za } i \neq k = 1,2,\ldots,n.$$

Za raspodelu datu u tabeli 2. modalni podinterval je [4,51;5,5), $f_2 = 12$, $f_1 = 7$ i $f_3 = 9$, $L_1 = 4{,}51$; $i = 0{,}99$, pa je po formuli (9) moda jednaka

$$M_o = 4{,}51 + \frac{f_2 - f_1}{(f_2 - f_1) + (f_2 - f_3)} \cdot i = 4{,}51 + \frac{12 - 7}{(12-7)+(12-9)} \cdot 0{,}99$$

$$M_o = 5{,}12875.$$

2. Mere varijabiliteta ili disperzije

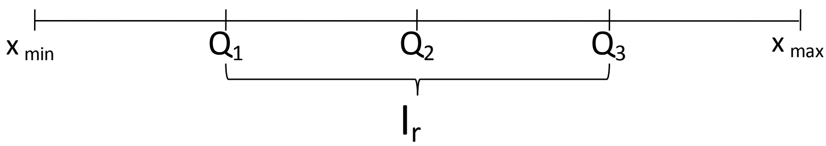
Srednje vrednosti karakterišu raspodelu obeležja u smislu označavanja gde se nalazi pretežni deo vrednosti obeležja, ali ne daju potpunu informaciju o raspodeli obeležja. Mogu sasvim različite raspodele imati iste srednje vrednosti. Zato pored srednjih vrednosti postoje i druge veoma značajne karakteristike obeležja i njegove raspodele. Mere varijacije iskazuju razmak promena obeležja X, meru odstupanja pojedinih vrednosti od srednjih vrednosti. Najgrublja mera varijacije obeležja X jeste razmak varijacije ili interval varijacije

$$W_r = x_{\max} - x_{\min} = W \qquad \text{...} (10)$$

kao razlika izmedu najveće i najmanje vrednosti obeležja. Taj razmak varijacije se često deli na 4 dela u smislu rašćenja obeležja. Prva četvrt Q_1 obuhvata 25% jedinica izvršenih merenja (podataka) i zove se prvi kvartil. Ograničen je odgovarajućom vrednošću obeležja x_k od koje je 25% manje. Drugi kvartil Q_2 ili medijana sadrži 50% podataka i treći kvartil Q_3 sadrži 75% podataka, a određen je vrednošću obeležja x_l od koje je 75% vrednosti manje.

Interkvartilna razlika, kao gruba mera varijacije je razlika između treće Q_3 i prve kvartile Q_1, tj.

$$I_r = Q_3 - Q_1 \qquad \text{...} (11)$$



Slika 5. Razmak varijacije i interkvartilna razlika

2.1 Varijansa (disperzija)

Kako su zbir odstupanja svih pojedinih vrednosti od aritmetičke sredine μ, za svako obeležje X jednako 0, tj. $\sum f_i (x_i - \mu) = 0$, to se uzima kvadrat odstupanja svake vrednosti od $\mu (x_i - \mu)^2$ i taj kvadrat umnoži učestalošću f_i obeležja x_i, pa se dobije $f_i (x_i - \mu)^2$. Svi ti kvadrati odstupanja se saberu za ceo skup S i podeli sa ukupnim brojem N elemenata skupa. Tako dobijeni izraz, u oznaci σ^2 ("sigma kvadrat"), naziva se varijansa obeležja X .

Za negrupisanu, prostu seriju podataka varijansa je

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \quad \text{.......} (12)$$

a za grupisanu seriju varijansa je

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \mu)^2 \quad \text{.......} (13)$$

i

$$f_1 + f_2 + ... + f_n = N .$$

Radni obrazac za varijansu dobija se kad se razvije kvadrat

$$(x_i - \mu)^2 = \left(x_i^2 - 2\mu x + \mu^2\right),$$

μ je konstanta, pa sledi

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \mu)^2 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(x_i^2 - 2\mu\cdot x_i + \mu^2\right) =$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i x_i^2 - 2\mu\cdot\frac{1}{N}\sum_{i=1}^{n} f_i x_i + \mu^2\frac{1}{N}\sum_{i=1}^{n} f_i =$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i x_i^2 - 2\mu\cdot\mu + \mu^2\cdot\frac{1}{N}\cdot N .$$

Konačno

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

ili za prostu seriju

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \mu^2 \quad \text{.......} (14)$$

Prema tome, varijansa je srednja vrednost kvadrata odstupanja pojedinih vrednosti obeležja x_i od svoje srednje vrednosti, ne može biti negativna, $\sigma^2 \geq 0$, $\sigma^2 = 0$ ako su sve vrednosti obeležja x_i međusobno jednake i jednake μ . Ali ti podaci nisu iz statističkog skupa, jer u statističkom skupu obeležje varira od elementa do elementa skupa. Varijansa je izražena i kvadratima mere.

Da bi se varijabilitet mogao porediti za različite skupove i obeležja koristi se pozitivni kvadratni koren iz varijanse, u oznaci σ ("sigma") i naziva se standardnom devijacijom. Standardna devijacija za prostu seriju je

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

a za uređenu, grupisanu raspodelu...(15)

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i x_i^2 - \mu^2} \ .$$

Statistička tabela sa podacima se prilagodi potrebama za izračunavanje. Za poređenje varijabiliteta obeležja dva skupa koristi se koeficijent varijacije, u oznaci V i on je jednak

$$V = \frac{\sigma}{\mu} \quad \text{...} (16)$$

količniku standardne devijacijc obeležja σ i njegove srednje vrednosti μ. Koeficijent varijacije V je jedna relativna mera varijabiiiteta obeiežja i izražena je u jedinicama srednje vrednosti.

Primer:

1. Koristeći tabelu 2. sređene raspodele obeležja X preseka jednog metalnog profila, po formuli (2) aritmetička sredina prečnika je:

   $$\mu = \frac{\sum f_i x_i}{N} = \frac{266}{50} = \frac{532}{100} = 5{,}32mm,$$

   po formuli (14) varijansa σ^2 tog prečnika je

   $$\sigma^2 = \frac{1}{50}1584 - (5{,}32)^2 = 31{,}68 - 28{,}3024$$

   $$\sigma^2 = 3{,}3776,$$

   a standardna devijacija je kvadratni koren iz varijanse i iznosi $\sigma \approx 1{,}8378$.

   Koeficijent varijacije prečnika x datog profila iznosi $V \approx 0{,}3439$.

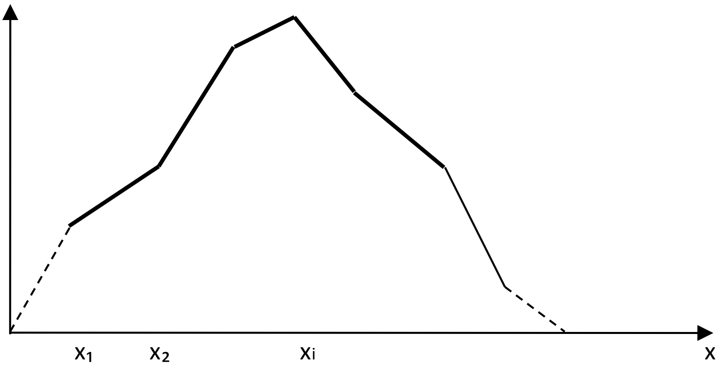
3. Mere oblika raspodele

Raspodela jednog obeležja je uređen niz mogućih vrednosti sa učestalošću

$$\begin{array}{ccccc} X & : & x_1 & x_2 & \dots & x_n \\ & & \updownarrow & \updownarrow & & \updownarrow \\ f & : & f_1 & f_2 & & f_n \end{array},$$

gde je $f_1 + f_2 + \dots + f_n = N$.

To znači da se x_i pojavljuje f_i puta za $i = 1,2,\dots,n$. Kad se odrede tačke $M_1(x_1, f_i)$, $M_2(x_2, f_2),\dots,M_n(x_n, f_n)$ u sistemu xOf i povežu po dve uzastopne tačke $\overline{M_i M_{i+1}}$ dobija se dijagram ili kriva frekvencija

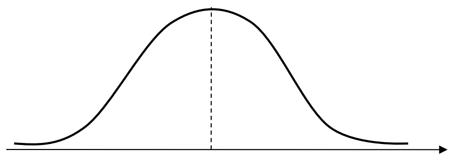


Slika 6. Kriva frekvencija

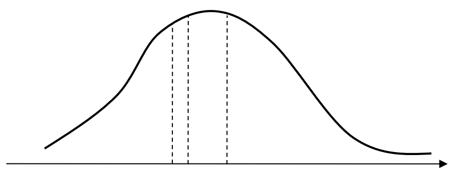
Krive frekvencije imaju različit oblik, koji odražava prirodu statističkog obeležja. Svaka tačka na krivoj frekvencija za svoju apscisu ima odgovarajuću.vrednost obeležja x_i a za svoju ordinatu ima učestalost f_i vrednosti x_i ili relativnu učestalost p_i te vrednosti x_i.

Prema tome kriva frekvencije jeste grafički oblik raspodele obeležja X na S.

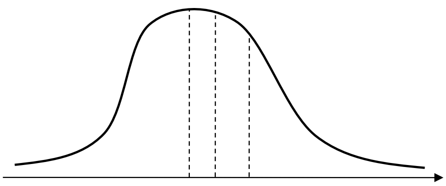
Krive frekvencija mogu biti simetrične u odnosu na jedan od koordinatnih pravaca a može da ima i različitu spljoštenost



a)

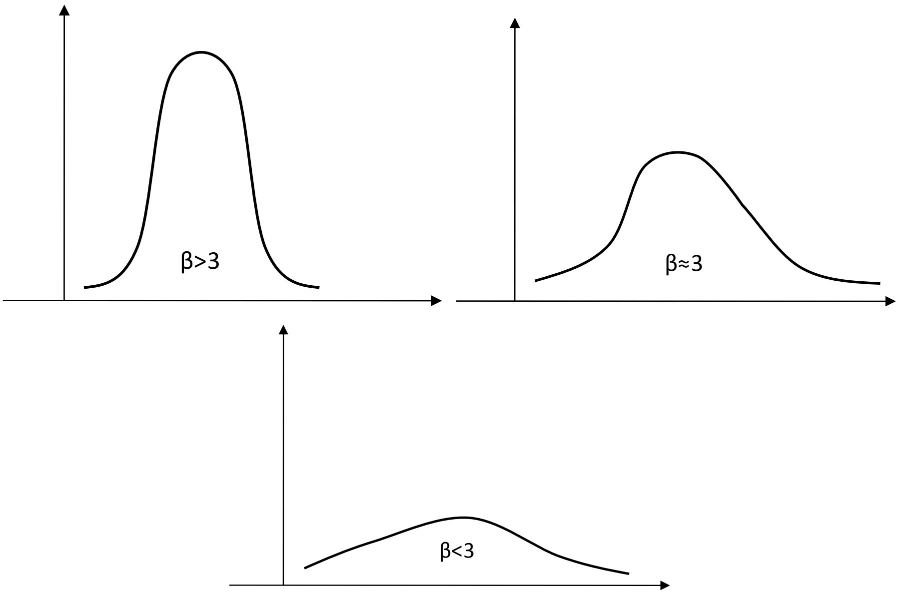


b)



c)

Slika 7. a) simetričan; b) negativno asimetričan i c) pozitivno asimetričan



Slika 8. Oblici spljoštenosti

Za datu seriju izračunava se

$$\mu_3 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(x_i - \mu\right)^3 \quad \text{...} (17)$$

treći centralni momenat raspodele X i

$$\mu_4 = \frac{1}{N}\sum_{i=1}^{n} f_i\left(x_i - \mu\right)^4 \quad \text{...} (18)$$

četvrti centralni momenat raspodele X.
μ_3 je očekivana vrednost trećeg stepena odstupanja od aritmetičke sredine μ, $\left(x_i - \mu\right)^3$, a μ_4 je očekivana vrednost četvrtog stepena odstupanja od aritmetičke sredine μ, $\left(x_i - \mu\right)^4$. Koeficijent asimetrije jednak je

$$\alpha = \frac{\mu_3}{\sigma^3} \quad \text{...} (19)$$

i predstavlja odnos između centralnog momenta trećeg reda μ_3 i standardne devijacije σ podignute na treći, tj. $(\sigma)^3$. μ_3 je „osetljiv“ na negativan znak (-) i zato je uključen u (19), jer $\left(x_i - \mu\right)^3 = \left[(-1)(\mu - x_i)\right]^3 = -(\mu - x_i)^3$. Koeficijent spljoštenosti, u oznaci β, jednak je

$$\beta = \frac{\mu_4}{\sigma^4} \quad \text{...} (20)$$

predstavlja odnos između centralnog momenta četrvtog reda μ_4 i standardne devijacije σ podignute na četvrti, tj σ^4.

U cilju ilustracije postupka i metode izračunavanja parametara statističkog skupa uzimamo jedan primer, gde se osnovni skup sastoji iz 100 tročlanih porodica, $N = 100$, a X = potrošnja kafe u kg zavreme T_o .

1. Data je raspodela potrošnje kafe X u kg u tročlanim domaćinstvima za vreme T_o skupa od 100 domaćinstava u obliku raspodele:

Tabela 3.

X (potrošnja)	do 3,5	$(3{,}5-4{,}5]$	$(4{,}5-5{,}5]$	$(5{,}5-6{,}5]$	$(6{,}5-7{,}5]$	$(7{,}5-8{,}5]$	$(8{,}5-9{,}5]$	9,5 i više
f (domaćinstva)	3	10	14	51	10	5	5	2

Sredine, intervalne sredine su redom

$3, 4,\ 5, 6,\ 7, 8, 9$ i 10,

Izračunati:

1. aritmetičku sredinu, modu i medijanu
2. varijansu σ^2 i standardnu devijaciju
3. koeficijent asimetrije α
4. koeficijent spljoštenosti β

Formirajmo tzv. "radnu" tabelu 4.

Tabela 4.

x	f	$f \cdot x$	Kumulativ "ispod"	$x-\mu$	$(x-\mu)^2$	$(x-\mu)^3$	$(x-\mu)^4$	$f(x-\mu)$	$f(x-\mu)^2$	$f(x-\mu)^3$	$f(x-\mu)^4$
3	3	9	3	-3	9	-27	81	-9	27	-81	243
4	10	40	13	-2	4	-8	16	-20	40	-80	160
5	14	70	27	-1	1	-1	1	-14	14	-14	14
6	51	306	78	0	0	0	0	0	0	0	0
7	10	70	88	1	1	1	1	10	10	10	30
8	5	40	93	2	4	8	16	10	20	40	80
9	5	45	98	3	9	27	81	15	45	135	405
10	2	20	100	4	16	64	256	8	32	128	512
Σ	100	600						0	188	138	1424

1. $\mu = \dfrac{\sum fx}{N} = \dfrac{600}{100} = 6$; $\mu = 6$; $\Sigma f_i(x-\mu) = 0$, $\Sigma f_i(x_i - 6) = 0$

$$M_o = L_1 + \frac{f_2 - f_1}{(f_2 - f_1) + (f_2 - f_3)} i = 5{,}5 + \frac{51-14}{(51-14)+(51-10)} 1 = 5{,}97;$$

jer je $L_1 = 5{,}5$; $f_2 = 51$, $f_1 = 14$, $f_3 = 10$ i $i = 1$

medijana

$$M_e = L_1 + \frac{\frac{N}{2} - \sum f_i}{f_{M_e}} \cdot i = 5{,}5 + \frac{50 - 27}{51} \cdot 1 = 5{,}95$$

$M_e = 5{,}95$, jer je $L_1 = 5{,}5$, medijalna frekvencija je $f_{M_e = 51}$, $\Sigma f_i = 27$, $i = 1$.

2. Varijansa je

$$\sigma^2 = \frac{\sum f(x-\mu)^2}{N} = \frac{188}{100} = 1{,}88$$

Standardna devijacija

$\sigma = \sqrt{1{,}88} \approx 1{,}37$.

3. Koeficijent asimetrije

$\alpha = \frac{\mu_3}{\sigma^3} = \frac{1{,}38}{1{,}37^3} \approx 0{,}54$, jer je $\mu_3 = \frac{138}{100} = 1{,}38$.

Kako je $0 < \alpha$ raspodela je pozitivno asimetrična ili asimetrična „ulevo“. To se zapaža iz odnosa aritmetičke sredine $\mu = 6$ i mode $M_o = 5{,}97$ da je $M_o < \mu$.

4. Koeficijent spljoštenosti

$$\beta = \frac{\mu_4}{\sigma^4} = \frac{14{,}24}{(1{,}37)^4} \approx 4{,}04$$

jer je

$$\mu_4 = \frac{\sum f(x_i - \mu)^4}{N} = \frac{1424}{100} = 14{,}24$$

To znači da je $\beta > 3$, tj. da je manje spljošten od normalnog rasporeda.

Svi parametri osnovnog skupa S od N elemenata, polazeći od mera centralne tendencije μ, G, H, M_o, M_e, mera varijacije σ^2, σ, koeflcijenata varijacije V do parametara oblika α i β su konstante i odredljive su tada i samo tada kad su poznati podaci obeležja za ceo skup S. To se odnosi i na one parametre raspodele koji nisu u ovom kursu pomenuti.

Kako se podaci o celoj raspodeli, odnosno o celom osnovnom skupu mogu dobiti samo iz njegovog popisa, koji se retko obavlja, na primer za stanovništvo na svakih 10 godina, to se može reći da u najvećem broju slučajeva nema podataka za ceo skup. Popisi su veoma skupa istraživanja. I zato umesto celog skupa S obima N bira se na slučaj uzorak $U \subset S$, obima $n < N$, gde se odnos $\frac{n}{N}$ često zove stopa uzorka ili stopa izbora u uzorak.