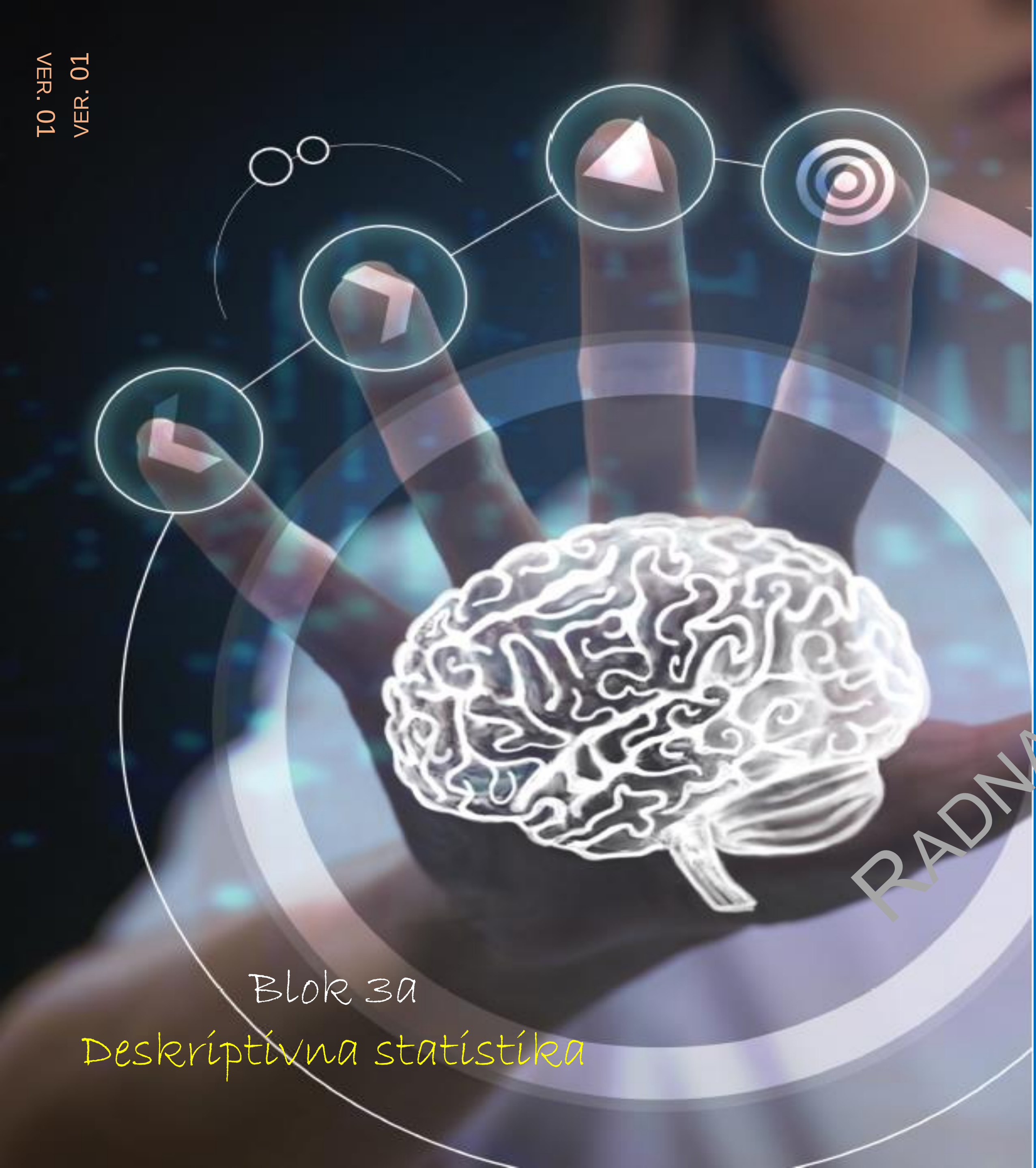


UVOD U NAUKU O PODACIMA

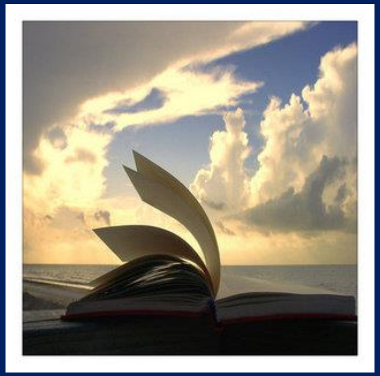
DR BOBAN STOJANOVIĆ / DR MILOVAN MILIVOJEVIĆ

RADNA VERZIJA

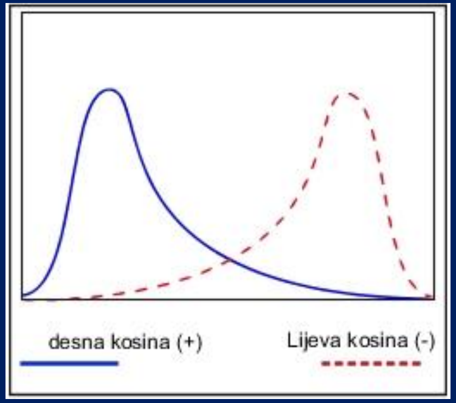
PRIRODNO-MATEMATIČKI FAKULTET
UNIVERZITET U KRAGUJEVCU

Literatura

1. M.Žižić, M.Lovrić, D. Pavličić, *Metodi statističke analize*, Ekonomski fakultet Beograd, Beograd, 2006., [pp. 11-7]
2. P.S.Mann, *Uvod u statistiku*, John-Wiley&Sons Inc. – Ekonomski fakultet Beograd, Beograd, 2009., [pp. 1-46]
3. V. Simonović, *Uvod u teoriju verovatnoće i matematičku statistiku*, Tekon, Beograd, 1995., [pp. 34-40; 54-66]
4. ....



RADNA VERZIJA



VER. 01 VER. 01

Deskriptivna Statistika

Slučajne veličine – Pojam i podela

Pojam
Podela

A	B	C	D...
• Diskretne	• Zavisne	• Atributivne	
• Kontinualne	• Nezavisne	• Numeričke	

RADNA VERZIJA

Tabela 1. Primer kontinualnih slučajnih veličina
[Rm-zatezna čvrstoća, Rp-granica razvlačenja, A-izduženje]

TRJ	Fe[%]	Mn[%]	Mg[%]	Si[%]	Rm	Rp	A	Šarža
9639	0.349	0.376	0.566	0.169	177.41	150.87	10.05	11077
9634	0.349	0.376	0.566	0.169	178.63	152.21	10.02	11077
9635	0.349	0.376	0.566	0.169	173.53	148.76	10.04	11077
9638	0.349	0.376	0.566	0.169	169.34	141.48	10.25	11077
9618	0.349	0.376	0.566	0.169	186.61	160.72	8.41	11077
9619	0.349	0.376	0.566	0.169	164.77	130.78	12.11	11077
9620	0.349	0.376	0.566	0.169	176.48	151.51	10.45	11077
9621	0.349	0.376	0.566	0.169	171.64	144.72	10.51	11077

Oblasti za analizu

- Prikupljanje podataka
- Sređivanje podataka
- Tabelarni prikaz podataka
- Grafički prikaz podataka
- Frekvencija pojave slučajne veličine
- Mere centralne tendencije
- Mere disperzije
- Mere oblika rasporeda...

Domen DESKRIPTIVNE STATISTIKE:

- Prikupljanje, sređivanje i prikazivanje podataka
- Metode određivanja parametara skupova

VER. 01

Deskriptivna Statistika / Merne skale

Postoje četiri merne skale: nominalna, ordinalna, intervalna i skala odnosa.

Nominalna skala svodi merenje na klasifikaciju. U ovoj skali brojevi se koriste kod pojava koje se mogu klasifikovati samo na određeni broj i tip modaliteta. Tako se klasifikuju pol, bračno stanje itd. Broj je upotrebljen samo kao simbol i ne iskazuje kvalitet, već služi da se različiti modaliteti odvoje, obeleže.

Ordinalna skala svodi merenje modaliteta (različite vrednosti slučajne veličine) na njihovo rangiranje po značaju s obzirom na usvojene kriterijume i to brojevima koji označavaju rang, ali ne pokazuju veličinu njihovog razlikovanja.

Intervalnu skalu karakteriše određena jedinica mere, kao na primer za kalendarsko vreme, temperaturu i dr.

Skala odnosa postiže najviši nivo merenja, jer se njome obezbeđuje značenje bilo kog odnosa merenih objekata kao što su: visina u centimetrima, starost u godinama, prihodi u dinarima... Ova skala nam dopušta da iskažemo proporcionalan odnos modaliteta koje merimo. Pakovanje šećera, na primer, koje ima tri puta više mernih jedinica od drugog pakovanja.

RADNA VERZIJA

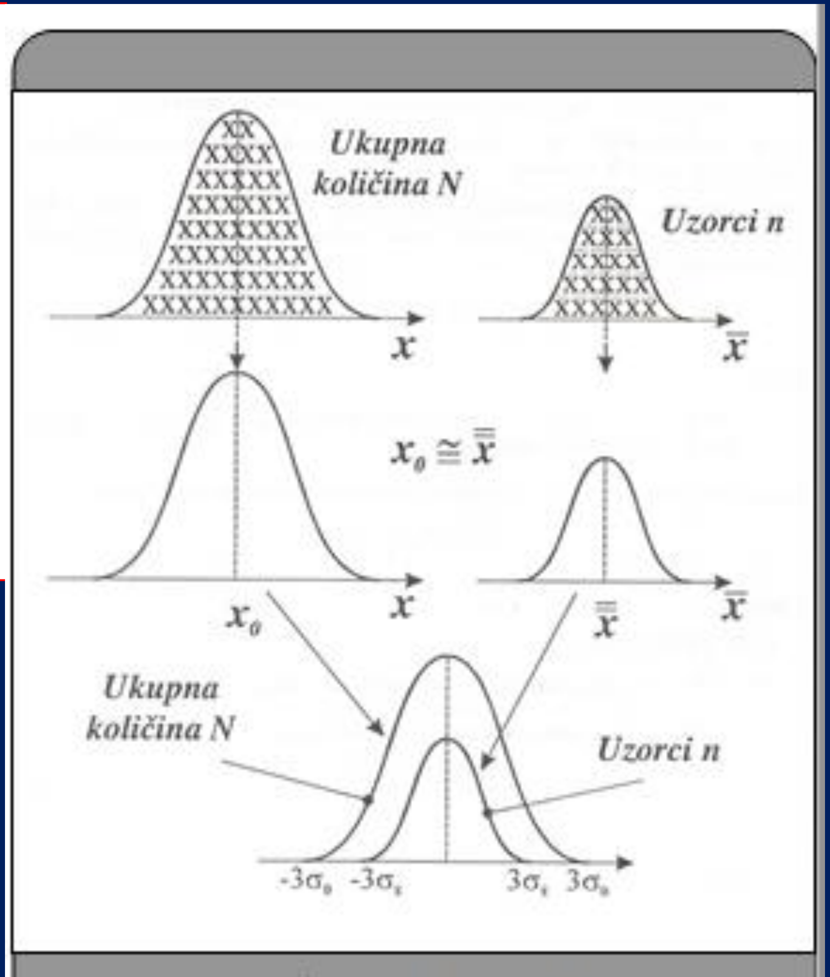
Deskriptivna Statistika / Statistički skup [Osnovni skup-**Populacija**]. Statistički uzorak [**Uzorak**]

Izbor, obim, cilj, metode.

Postoji zavisnost između parametara koji opisuju UZORAK i parametara koji opisuju OSNOVNI SKUP (populaciju). Na osnovu „dovoljnog" broja uzoraka mogu se doneti zaključci o ponašanju OS.

Za svaki uzorak se izračunava:

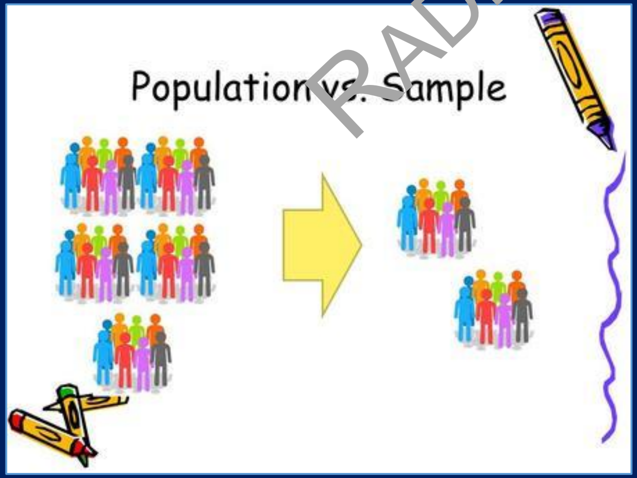
- $\bar{x}$ - **aritmetička sredina**
- σ - **standardna devijacija** (s je ocena za σ)



Sl. 1.Odnos OS i Uzorak

$$\mu = \bar{\bar{x}}$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$



RADNA VERZIJA

Deskriptivna Statistika / Broj intervala

Kod neprekidnih obeležja (ili prekidnih sa velikim brojem modaliteta) postavlja se pitanje broja grupa, odnosno veličine grupnih intervala. Rešenje ovog pitanja treba da zadovolji i zatev za informacijama i zahtev preglednosti.

Veličina grupnih intervala određuje se ponekad na taj način što se prvo odredi broj grupa pomoću tzv. *Sturges*-ovog pravila

$$k = 1 + 3.3 \cdot \log n$$

gde je n-ukupan broj podataka, a potom i veličina intervala

$$i = \frac{x_{max} - x_{min}}{1 + 3.3 \cdot n}$$, gde je

i – veličina intervala, x_{max} – podatak sa najvećom vrednošću, x_{min} – podatak sa najmanjom vrednošću.

Tako, na primer, za 1000 podataka koji variraju od 500 do 1215, broj grupa bi iznosio:

$k = 1 + 3.3 \cdot \log 1000 \approx 11$ a veličina intervala

$$i = \frac{1215 - 500}{11} = 65$$

Pri formiranju grupnih intervala preporučljivo je početi sa vrednošću manjom od najmanje u seriji i završiti sa većom od najveće vrednsoti u seriji. Takođe, treba voditi računa o pravilnom razgraničenju grupnih intervala.

Potpuno određena pravila za izbor broja podintervala ne postoje, ali iskustva istraživača preporučuju da se broj odredi ili pomoću Sturges-ovog pravila ili nekim od sledećih obrazaca:

$$r = \sqrt{n} \quad ili \quad r = 2 \cdot \sqrt[3]{n} \quad ili \quad r = 5 \cdot \log n$$

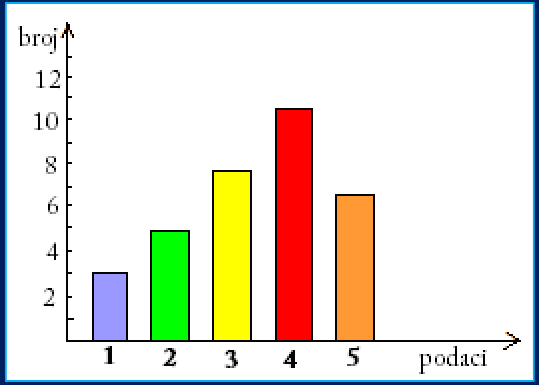
RADNA VERZIJA

Rasporedi frekvencija predstavljaju naročito važan predmet statističke analize. Oni otkrivaju ne samo koncentraciju individualnih vrednosti nego i prirodu varijabiliteta. Takođe, obezbeđuju uspešnu komparaciju dveju ili više serija, grafički ili raznim deskriptivnim merama. Pored navedenog, rasporedi frekvencija, se mogu matematički obrađivati kao empirijske funkcije.

Ako frekvenicju (f_i) izvesnog atributa ili izvesne vrednosti obeležja stavimo u odnos prema ukupnom broju jedinica toga skupa (Σf_i), onda dobijamo **relativnu frekvenciju**.

Relativne frekvencije povećavaju mogućnost analize statističkog skupa. Na ovaj način izražena struktura skupa omogućava poređenje sa strukturom podskupova ili poređenje sa strukturom sličnih skupova.

f_i

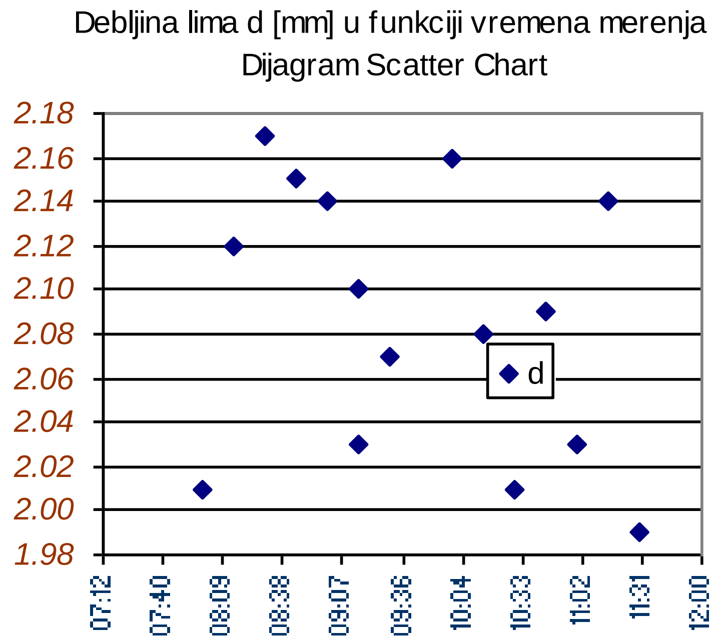


f_r

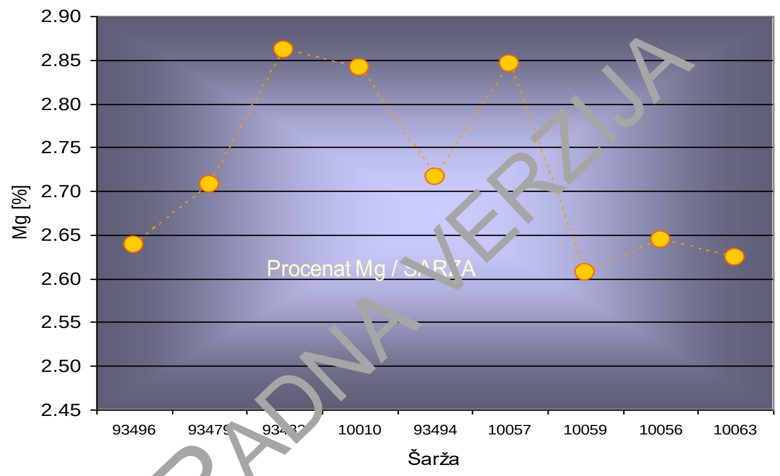
RADNA VERZIJA

Deskriptivna Statistika / VIZUELIZACIJA / Dijagrami

Scatter Chart

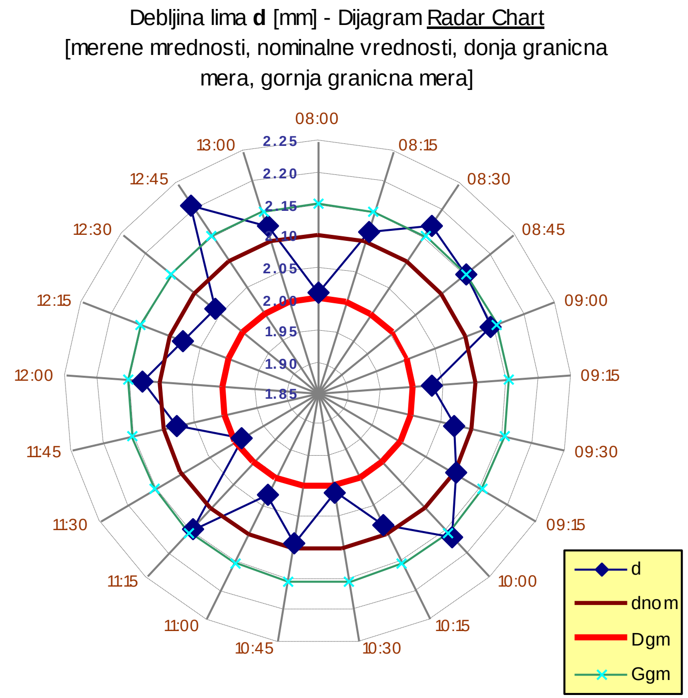


Line Chart



RADNA VERZIJA

Radar Chart



Dijagrami u obliku histograma / Dijagrami sa stubićima 2D i 3D

Bar Chart

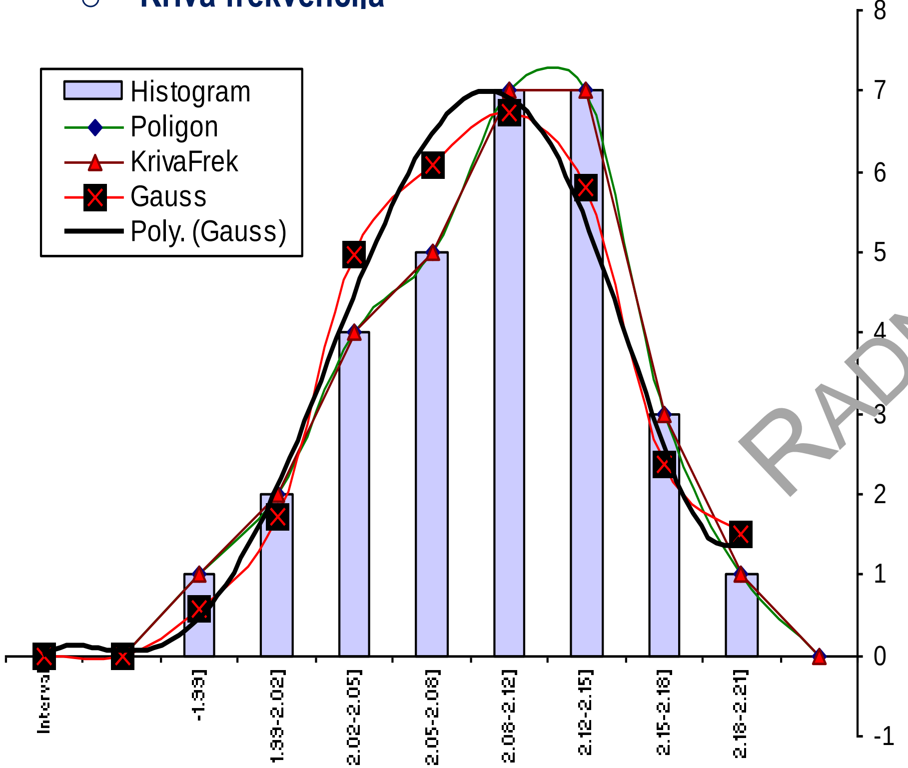
- **Histogram frekvencija**
- **Poligon Frekvencija**
- **Kriva frekvencija**

VER. 01
VER. 01

Deskriptivna Statistika / VIZUELIZACIJA / Dijagrami

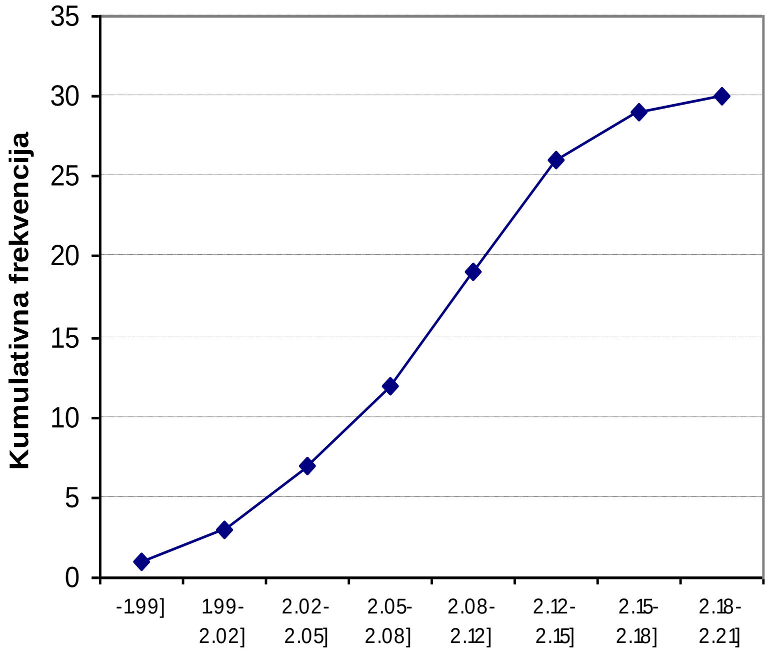
Dijagrami u obliku histograma / Dijagrami sa stubićima 2D i 3D <u>*Bar Chart*</u>

- **Histogram frekvencija**
- **Poligon Frekvencija**
- **Kriva frekvencija**



RADNA VERZIJA

Kumulanta



Deskriptivna Statistika / Mere centralne tendencije

Pokazatelje rasporeda frekvencija koji pokazuju ceo Osnovni Skup nazivamo **parametrima skupa** i svrstavamo ih u tri grupe. Jednu sačinjavaju srednje vrednosti kao mere centralne tendencije rasporeda, drugu mere disperzije (raspršenosti) i treću mere oblika rasporeda.

Deskriptivne mere koje se odnose na *sve jedinice skupa* nazivaju se **parametrima skupa**, a deskriptivne mere koje se odnose na *uzorak* su **statistike uzorka**.

Mere centralne tendencije

Rasporedi frekvencija pružaju široku mogućnost za otkrivanje karakteristika skupova, analizu njihove strukture i unutrašnjih odnosa. Oni okrivaju tendenciju grupisanja pojedinih vrednosti obeležja oko vrednosti karaktersitične za dati skup. Srednje vrednosti su pokazatelji centralne tendencije i lokacije skupa. To su najznačajniji pokazatelji numeričkih serija koji po datim merilima reprezentuju čitav skup i omogućuju poređenje između raznih skupova. Srednje vrednosti se koriste u svim oblastima statističke analize.

U zavisnosti od načina određivanja centralne vrednosti se dela na:

- **Izračunate** (aritmetička sredina, geometrijska, harmonijska sredina), koje se izračunavaju na osnovu svih vrednosti obeležja
- **Pozicione** (modus / M_o / i medijama / M_e / koje se određuju položajem u seriji.

Aritmetička sredina

Najbitnija za statističku analizu.

Prosek/Aritmeička sredina **Osnovnog Skupa** (OS):

- Negrupisani podaci

$$\mu = \frac{\sum x}{N},$$

- Grupisani podaci / Ponderisana aritmetička sredina OS

$$\mu = \frac{1}{N}\sum_{i=1}^{k} f_i \cdot x_i, \quad N = \sum_{i=1}^{k} f_i$$

gde je: N – broj podataka u OS
f_i - frekvencija pojavljivanja obeležja u skupu
k – broj grupa/intervala

Deskriptivna Statistika / Mere centralne tendencije

Prosek/Aritmeička sredina **Uzorka** (U):

- Negrupisani podaci

  $$\bar{x} = \frac{\sum x}{n}$$, gde je n – broj podataka u Uzorku || $\bar{x}$ - Čita se x bar ||

- Grupisani podaci / Ponderisana aritmetička sredina uzorka

  $$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i \cdot x_i \ , \ n = \sum_{i=1}^{k} f_i$$

  gde je: n – broj podataka u Uzorku
  f_i - frekvencija pojavljivanja obeležja u uzorku
  k – broj grupa/intervala

Ako je raspored frekvencija dat u vidu *k* grupnih intervala, Aritmetička sredina OS, kao i Uzorka, izračunava se kao ponderisana aritmetička sredina središnjih vrednosti grupnih intervala.

Aritmetička sredina je pokazatelj lokacije osnovnog skupa ili uzorka.

Osobine aritmetičke sredine /*navode se bitnija svojstva*/:

- Zbir odstupanja aritmetičke sredine od pojedinih vrednosti obeležja jednak je nuli

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

  $$\sum_{i=1}^{k} f_i \cdot (x_i - \mu) = 0$$ za grupisane podatke

- Ako su dva obeležja vezana linearnom funkcijom tada su i njihove aritmetičke sredine vezane istom aritmetičkom funkcijom

  $$y = b_0 + b_1 \cdot x \qquad \mu_y = b_0 + b_1 \cdot \mu_x$$

Deskriptivna Statistika / Mere centralne tendencije

Geometrijska sredina

Geometrijska sredina je srednja vrednost koja izravnjava relativne ili proporcionalne promene između vrednosti podataka, za razliku od aritmetičke sredine koja izravnjava apsolutne razlike između vrednosti podataka posmatrane serije. Ona se, prema tome ne dobija iz zbira nego iz proizvoda vrednosti podataka, s tim što se iz ovog uzima pozitivna vrednost korena čiji je izložitelj jednak njihovom broju.

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

RADNA VERZIJA

Harmonijska sredina

Harmonijska sredina je recipročna vrednost aritmetičke sredine recipročnih vrednosti obeležja.
Slično kao kod aritmetičke i geometrijske sredine postoji prosta i ponderisana harmonijska sredina.

Prosta harmonijska sredina

$$H = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}}$$

Ponderisana harmonijska sredina

$$H = \frac{N}{\sum_{i=1}^{k} \frac{f_i}{x_i}}$$

Deskriptivna Statistika / Mere centralne tendencije

Modus / Moda / Mod / M_o

Modus je vrednost obeležja koja u posmatranoj seriji ima najveću frekvenciju – najčešće se javlja i zato je najtipičnija vrednost u seriji.

Kada je u jednoj seriji samo jedna vrednost obeležja sa najvećom frekvencijom, onda se ona naziva unimodalnom a ako postoje dve ili više takvih vrednosti onda je bimodalna ili multimodalna.

Može se desiti da modusa nema / primer: 34,12,13,15,25/

Za seriju grupisanih podataka, odnosno neprekidnih podataka modus nije tako lako uočljiv. Treba ga tražiti u intervalu sa najvećom frekvencijom koji se naziva modalnim intervalom.

Obrazac za određivanje modusa

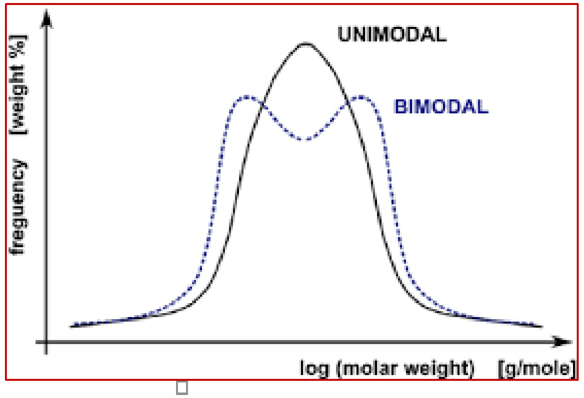
$$Modus(Mo) = L_1 + \frac{f_2 - f_1}{(f_2 - f_1) + (f_2 - f_3)} \cdot i$$

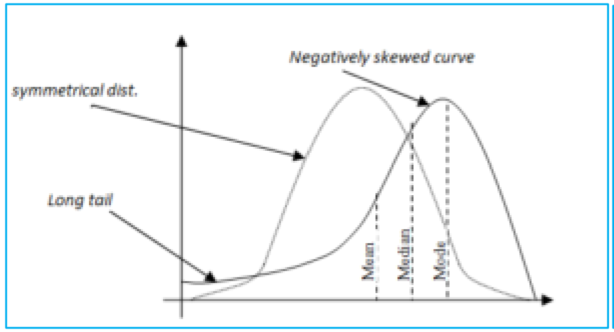
gde je

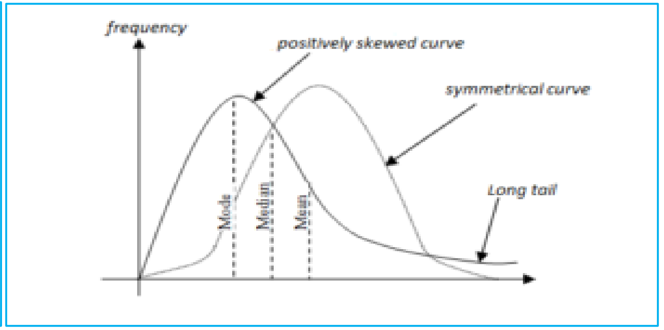
L_1 – donja granica modalnog intervala

i - dužina grupnog intervala

f_1, f_2, f_3 – redom frekvencije predmodalnog, modalnog i poslemodalnog inervala







RADNA VERZIJA

Deskriptivna Statistika / Mere centralne tendencije

Medijana / M_e

Medijana je ona vrednost obeležja koja se nalazi u sredini serije uređene po veličini obeležja, odnosno to je vrednost obeležja koja deli sumu svih frekvencija na dva jednaka dela, tako da jedna polovina obuhvaćenih slučajeva ima manju, a druga polovina veću vrednost od medijane.

Ako su vrednosti obeležja poređane po veličini i od njih obrazovana serija negrupisanih podataka:

$$x_1, x_2, \ldots, x_N$$

- Ako je N – neparan broj, onda je medijana jednaka srednjem članu /primer: 23,25,26,29,36 / Me=26
- Ako je N - paran broj , onda je medijana jednaka aritmetičkoj sredini dva središnja člana /primer: 18,19,20,22,24,26/ Me=21

Za serije grupisanih podataka medijana se dobija interpolacijom između donje i gornje granice intervala grupe u kojoj se medijana nalazi

$$Medijana(Me) = L_1 + \frac{\frac{N}{2} - \sum f_1}{f_{Me}} \cdot i$$

gde je

L_1 – donja granica medijalnog intervala
N – broj članova serije
Σf_1- zbir frekvencija predmedijalnog intervala
i - dužina medijalnog intervala
f_{Me} – frekvencija medijalnog inervala

Medijana se ponekad naziva i drugim kvartilom s obzirom na mogućnost podele jedne serije na četiri jednaka dela.

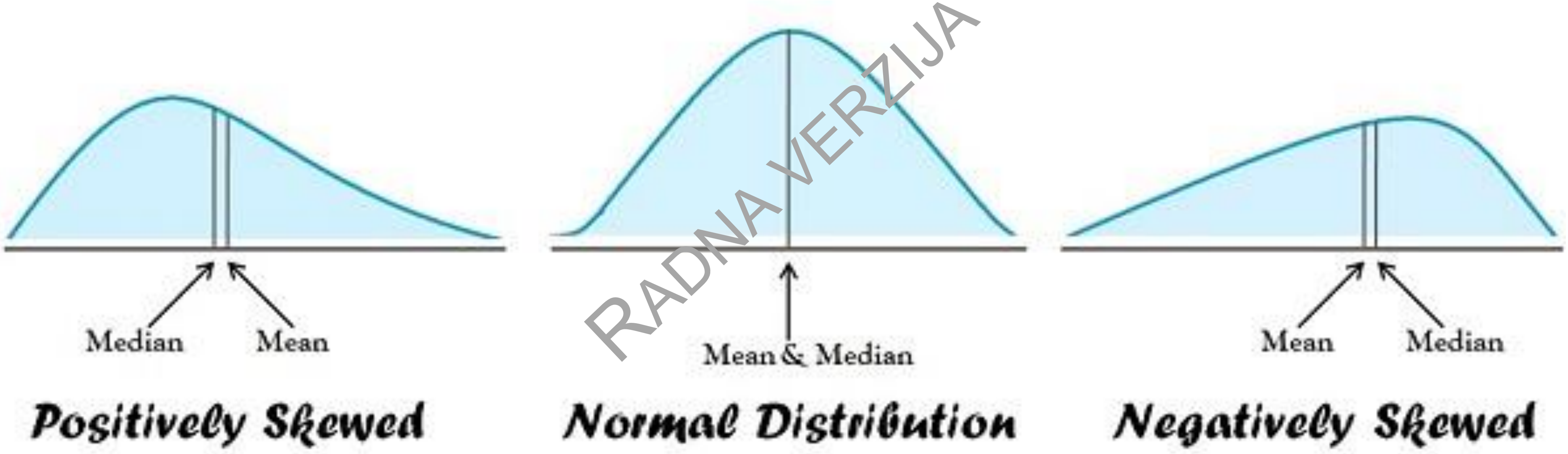
Ako se serija podataka rangiranih po veličini podeli u četiri jednaka dela, vrednosti obeležja koje ih dele nazivaju se kvartilima:

- Prvi kvartil - Q_1, vrednost obeležja od koje 25% elemenata skupa uređenog po veličini ima manju ili jednaku vrednost
- Drugi kvartil – Q_2, vrednost obeležja od koje 50% elemenata skupa uređenog po veličini ima manju ili jednaku vrednost
- Treći kvartil – Q_3, vrednost obeležja od koje 75% elemenata skupa uređenog po veličini ima manju ili jednaku vrednost.

RADNA VERZIJA

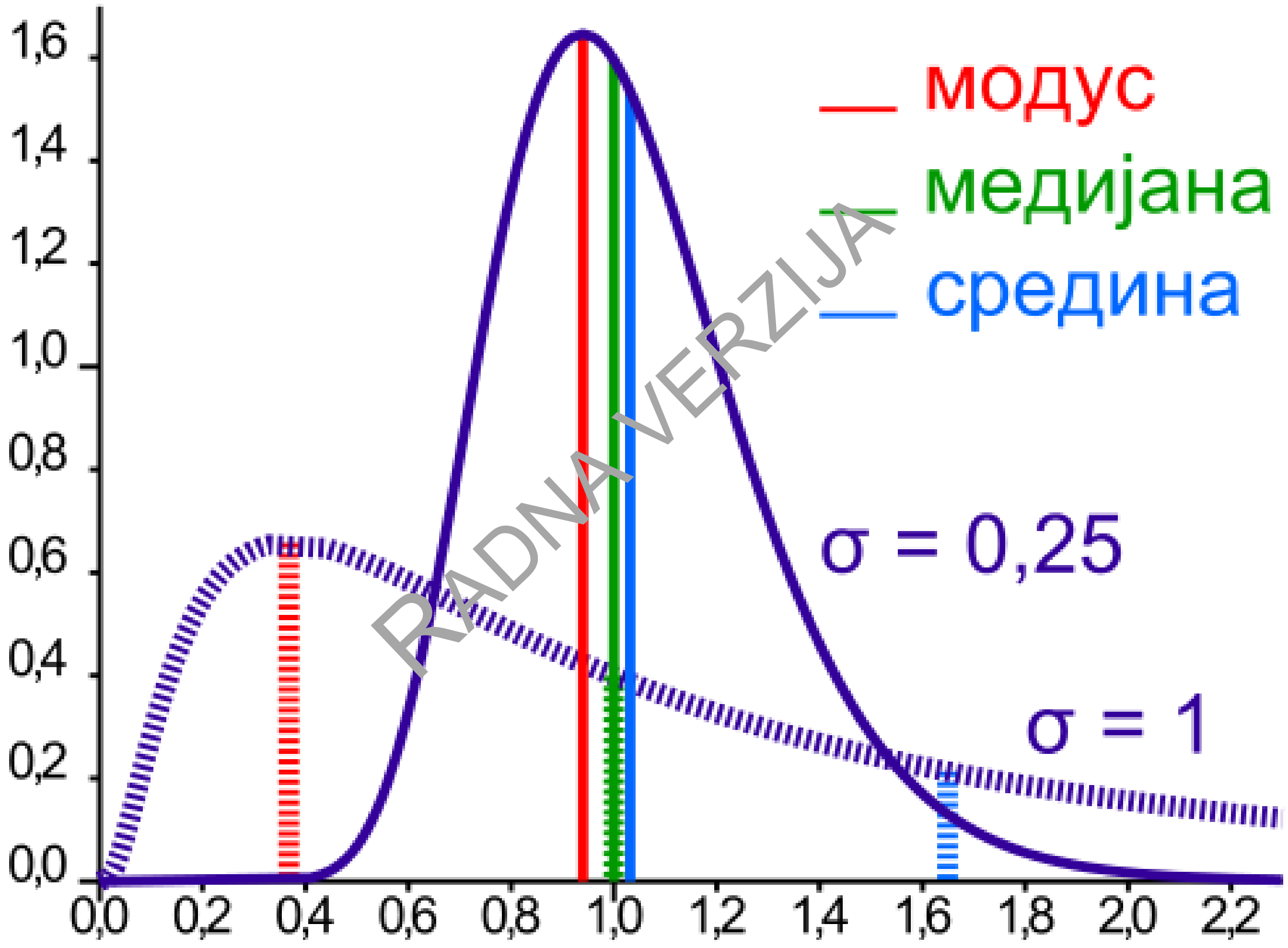
VER. 01

Deskriptivna Statistika / Mere centralne tendencije



RADNA VERZIJA

Deskriptivna Statistika / Mere centralne tendencije



RADNA VERZIJA

Deskriptivna Statistika / Mere disperzije

Apsolutne mere disperzije

Apsolutne mere disperzije iskazuju varijabilitet u apsolutnim iznosima onih mernih jedinica u kojima su dati modaliteti posmatranog obeležja. Ove mere, kao i mere lokacije mogu biti pozicione i izračunate.

Pozicione mere apsolutne disperzije

Razmak / Interval varijacije

Razmak je razlika između najviše i najniže vrednosti obeležja u seriji

$$R = x_{\max} - x_{\min}$$

Razlika / Interval varijacije

Interkvartilna razlika

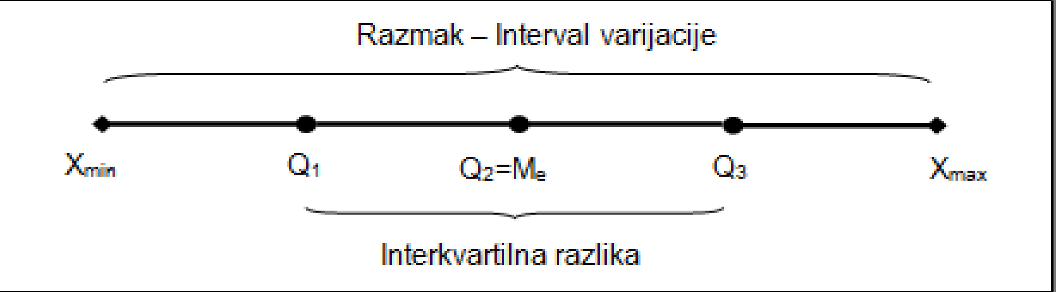
Da bi se eliminisao uticaj ekstremnih vrednosti na iznos razlika, odnosno intervala varijacije, izračunava se kao dopunska mera interkvartilan razlika, tj. razlika između trećeg i prvog kvartila

$$i_Q = Q_3 - Q_1$$

Interkvartilna razlika

Ona isključuje 25% podataka sa najnižim vrednostima i 25% podataka sa najvišim vrednostima.

RADNA VERZIJA



Sl. 3. Interkvartilana razlika

Ako je Razmak, odnosno Interval varijacije veliki a Interkvartilna razlika mala znači da na krajevima intervala postoje ekstremne vrednosti, ali da ostali članovi serije ne pokazuju veliki varijabilitet. Kada je Interkvarilna razlika velika, slika o varijabilitetu serije nije dovoljno jasna.

Deskriptivna Statistika / Mere disperzije

Izračunate mere apsolutne disperzije

Disperzija/Varijansa i Srednje kvadratno odstupanje/Standardna devijacija **Osnovnog Skupa**

Kao što je ranije navedeno, suma pojedinih odstupanja vrednosti obeležja od aritmetičke sredine jednaka je nuli. Tako da ova vrednost nije mogla biti mera disperzije. Zato je kao mera disperzije uzet prosek kvadrata odstupanja. Ova mera naziva se disperzija ili varijansa.

Disperzija/Varijansa **Osnovnog Skupa** (OS):

- Negrupisani podaci

$$D = \sigma_0^{\,2} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 ,$$

gde je: *N* – broj podataka u OS
μ – aritmetička sredina OS

- Grupisani podaci

$$D = \sigma_0^{\,2} = \frac{1}{N}\sum_{i=1}^{k} f_i \cdot (x_i - \mu)^2 , \quad N = \sum_{i=1}^{k} f_i$$

gde je: *N* – broj podataka u OS
f_i - frekvencija pojavljivanja obeležja u skupu
k – broj grupa/intervala

ili za lakši manuelni obračun

$$\sigma_0^{\,2} = \frac{1}{N}\sum_{i=1}^{k} f_i \cdot x_i^{\,2} - \mu^2$$

Standardna devijacija **Osnovnog Skupa** (OS):

- Negrupisani podaci

$$\sigma_0 = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} ,$$

gde je: *N* – broj podataka u OS
μ – aritmetička sredina OS

- Grupisani podaci

$$\sigma_0 = \sqrt{\frac{1}{N}\sum_{i=1}^{k} f_i \cdot (x_i - \mu)^2} , \quad N = \sum_{i=1}^{k} f_i$$

gde je: *N* – broj podataka u OS
f_i - frekvencija pojavljivanja obeležja u skupu
k – broj grupa/intervala

ili za lakši manuelni obračun

$$\sigma_0 = \sqrt{\frac{1}{N}\sum_{i=1}^{k} f_i \cdot x_i^{\,2} - \mu^2}$$

RADNA VERZIJA

VER. 01

Deskriptivna Statistika / Mere disperzije

Disperzija/Varijansa

Standardna devijacija **Uzorka** (U):

- Negrupisani podaci

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1},$$

gde je: n – broj podataka u Uzorku

$\bar{x}$ – aritmetička sredina uzorka

- Grupisani podaci

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

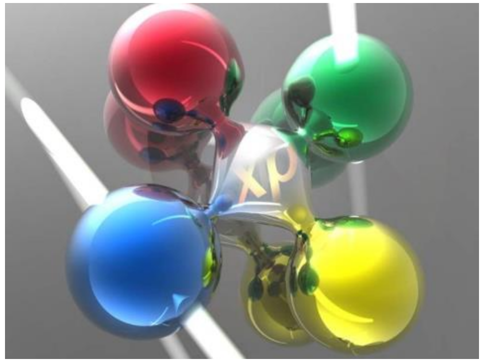
gde je: n – broj podataka u uzorku

f_i - frekvencija pojavljivanja obeležja u uzorku

k – broj grupa/intervala

ili za lakši manuelni obračun

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i \cdot x_i^2 - n \cdot \bar{x}^2}{n-1}$$



RADNA VERZIJA

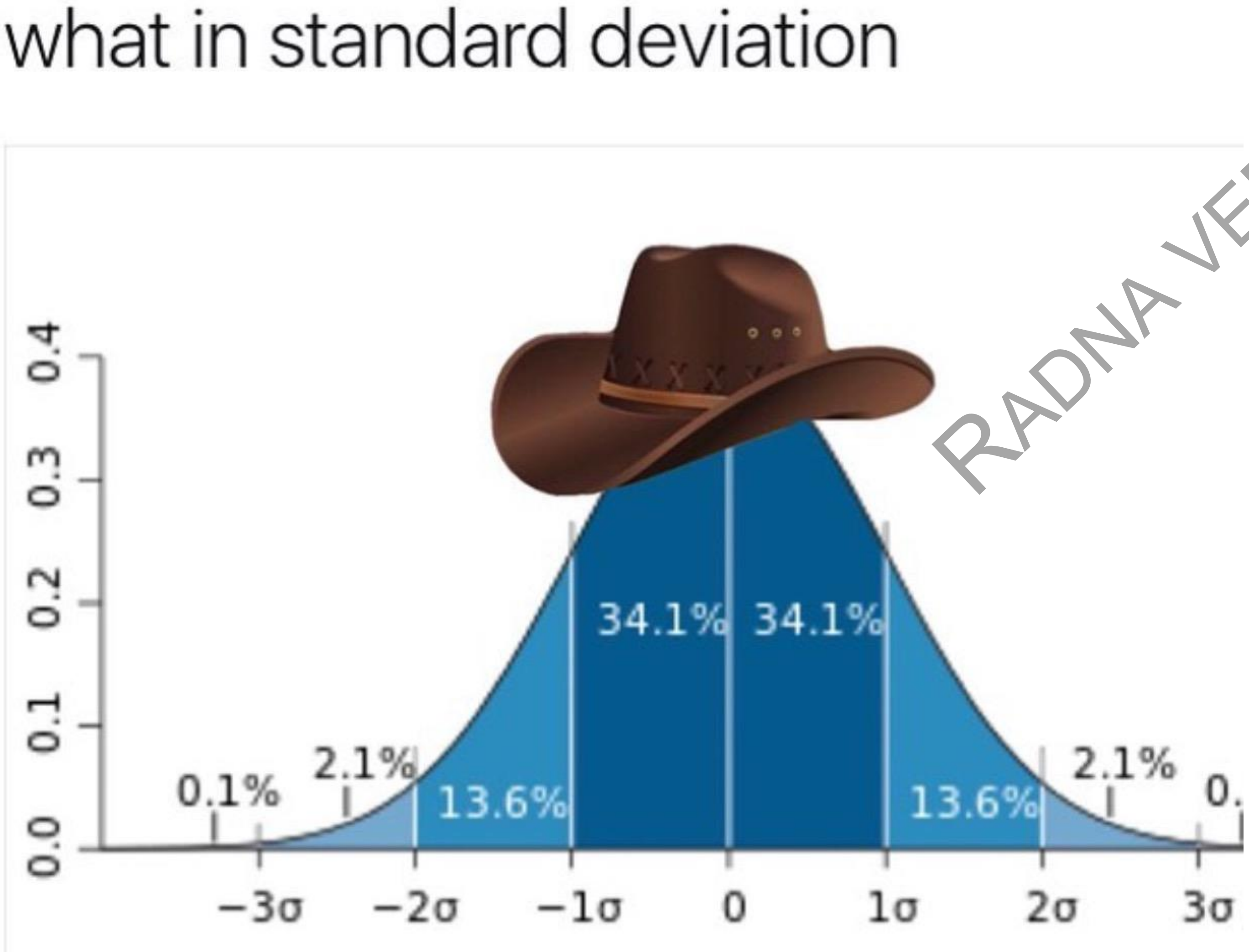
Standardna devijacija **Uzorka** (U):

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

Napomena: Kao ocena Standardne devijacije uyorka uzorka koristi se veličina ***s***

Deljenje sa **(n-1)** u prethodnim formula umesto sa očekivanim **n**, vezano je za broj stepeni slobode (*fd-freedom degree*).

Broj sepeni slobode za disperziju nije **n** (broj slobodnih promenljivih-podataka- posmatranja) već **(n-1)**, zato što se **(n-1)** razlika $(x_i - \bar{x})$ razlika „ponaša" slobodno ali je **n**-ta razlika „neslobodna" jer važi da je $\Sigma(x_i - \bar{x}) = 0$. Dakle, ukida se jedan stepen slobode, te više nije **n** već **n-1** stepen slobode.



RADNA VERZIJA

Deskriptivna Statistika / Mere disperzije

Mere relativne disperzije

U relativne mere disperzije ubrajaju se: koeficijent varijacije, koeficijent interkvartilne varijacije i standardizovano odstupanje.

Koeficijent varijacije

Odnos standardne devijacije i aritmetičke sredine naziva se koeficijentom varijacije.

$$V = \frac{\sigma}{\mu} \quad [\cdot 100\%]$$

Ovaj koeficijent se najčešće se izražava u procentima. Što je koeficijent varijacije veći odstupanje je veće.
Pošto je u pitanju relativna mera disperzije, ovaj pokazatelj se može koristiti za poređenje serija čije merne jedinice nisu iste i u tome je prednost ovog koeficijenta.

Koeficijent Interkvartilne varijacije

Za upoređenje disperzija više skupova ili uzoraka upotrebljava se i koeficijent interkvartilne varijacije čiji obrazac glasi:

$$V = \frac{Q_3 - Q_1}{Q_3 + Q_1} \quad [\cdot 100\%]$$

Ukoliko se vrednost koeficijenta približava nuli disperzija je relativno manja.

Standardizovano odstupanje

Varijacija se može posmatrati ne samo sa gledišta rasporeda frekvencija kao celine, nego i sa gledišta individualnih podataka.
Dakle, pitanje je da li je odstupanje neke vrednosti obeležja $(x_i - \bar{x})$ malo ili veliko. Odgovor se nalazi u poređejnu ovog odstupanja sa standardnom devijacijom.

Obrazac za normalizovano ili standardizovano odstupanje glasi:

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

Standardizovano odstupanje predstavlja opštu meru odstupanja individualnih podataka od aritmetičke sredine.

RADNA VERZIJA

Deskriptivna Statistika / Mere oblika rasporeda

Mere oblika rasporeda

Osnovna dva pokazatelja koja se odnose na oblik rasporeda su koeficijent asimetrije i koeficijent spljoštenosti.
Za određena merenja u statistici koriste se odstupanja vrednosti obeležja od aritmetičke sredine skupa na određeni stepen, tzv. centralni momenti rasporeda:

- **Nulti moment (M_0)** - **jednak je jedinici**
- **Prvi moment (M_1)** - **jednak je nuli**
- **Drugi moment (M_2)** - **jednak je disperziji**
- **Treći moment (M_3)** - **koristi se za meru asimetrije**
- **Četvrti moment (M_4)** - **koristi se za meru spljoštenosti**

Obrasci za treći i četvrti moment **osnovnog skupa** (grupisani podaci) glase:

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

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

U software-skim paketima koriste se korigovane formule:

$$M_3 = \frac{N}{(N-1)\cdot(N-2)}\sum_{i=1}^{k} f_i \cdot (x_i - \mu)^3$$

$$M_4 = \frac{(N+1)}{(N-1)\cdot(N-2)\cdot(N-3)}\sum_{i=1}^{k} f_i \cdot (x_i - \mu)^4 - 3\cdot\frac{(N-1)^2}{(N-2)\cdot(N-3)}\cdot\sigma_0^4$$

Isti oblik imaju formule koje se primenjuju za **uzorak** (grupisani podaci)

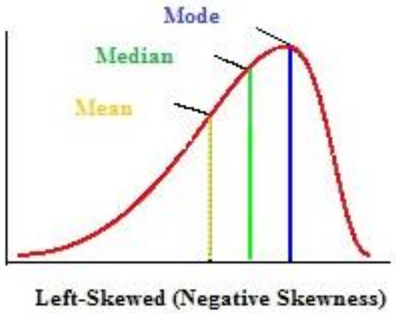
$$M_3 = \frac{1}{n}\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^3$$

$$M_4 = \frac{1}{n}\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^4$$

Odnosno korigovane formule:

$$M_3 = \frac{n}{(n-1)\cdot(n-2)}\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^3$$

$$M_4 = \frac{(n+1)}{(n-1)\cdot(n-2)\cdot(n-3)}\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^4 - 3\cdot\frac{(n-1)^2}{(n-2)\cdot(n-3)}\cdot\sigma^4$$



Ili slično za negrupisane podatke

$$M_3 = \frac{1}{n}\sum_{i=1}^{k} (x_i - \bar{x})^3$$

$$M_4 = \frac{1}{n}\sum_{i=1}^{k} (x_i - \bar{x})^4$$

uz korekcije:

$$M_3 = \frac{n}{(n-1)\cdot(n-2)}\sum_{i=1}^{k} (x_i - \bar{x})^3$$

$$M_4 = \frac{(n+1)}{(n-1)\cdot(n-2)\cdot(n-3)}\sum_{i=1}^{k} (x_i - \bar{x})^4 - 3\cdot\frac{(n-1)^2}{(n-2)\cdot(n-3)}\cdot\sigma^4$$

RADNA VERZIJA

VER. 01

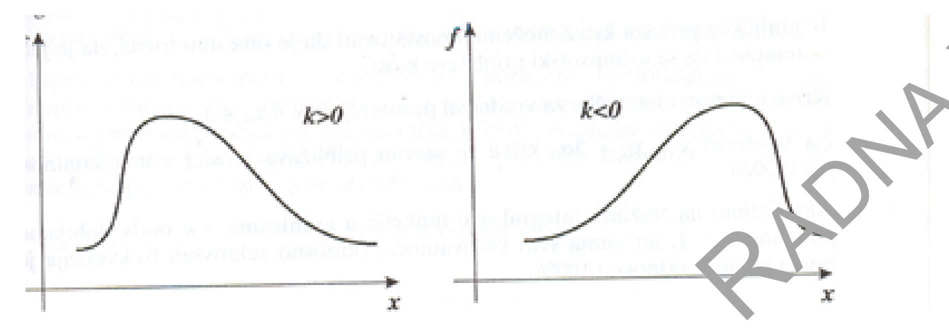
Deskriptivna Statistika / Mere oblika rasporeda

Kako su treći i četvrti moment izraženi u istim jedinicama mere kao i emirijski podaci na osnovu kojih se računaju, to se u cilju poređenja izračunavaju relativni pokazatelji oblika rasporeda

$$\alpha_3 = \frac{M_3}{\sigma^3}$$ / mera asimetrije / *skewness* /

$$\alpha_4 = \frac{M_4}{\sigma^4}$$ / mera spljoštenosti / *kurtosis* /

Kod pozitivne asimetrije (asimetrija „udesno") koeficijent α_3 je veći od nule a kod negativne (asimetrija „ulevo") je manji od nule.
Što se vrednost ovog koeficijenta više razlikuje od nule to je asimetrija rasporada veća.
Smatra se da je rasporeda umereno asimetričan ako je: $-0.5 < \alpha_3 < 0.5$



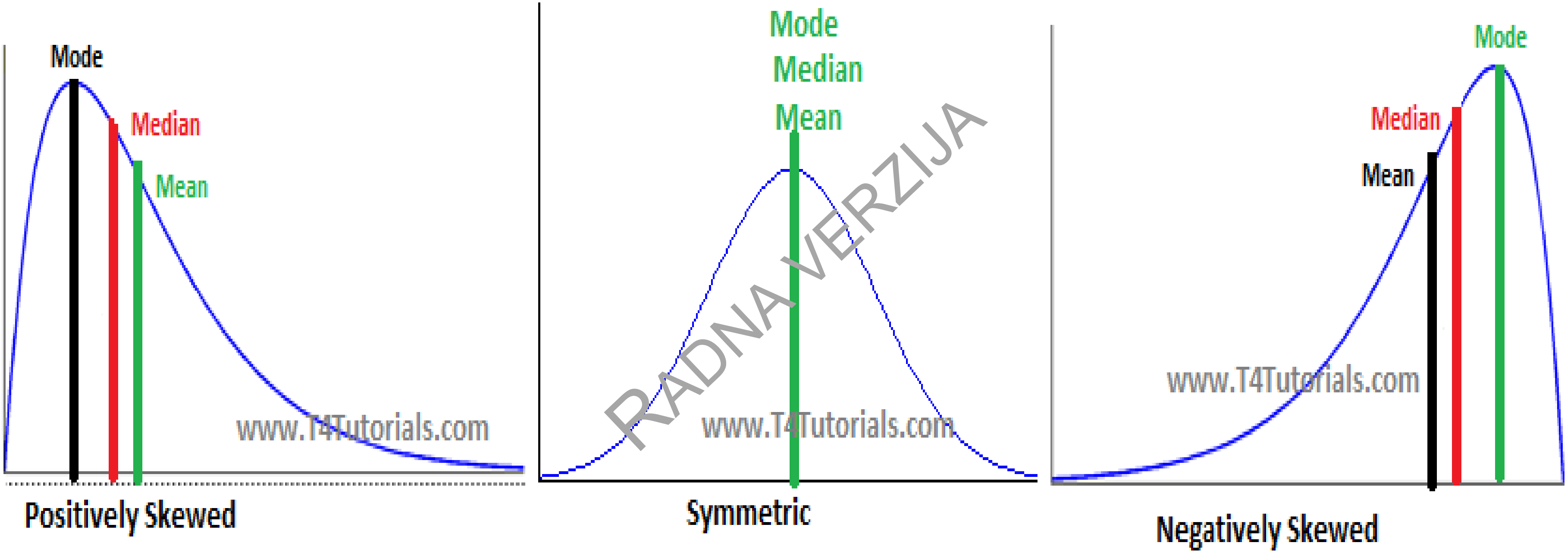
Sl. 4. k = α Mera asimetrije

RADNA VERZIJA

Koeficijent α_4 pedstavlja relativnu meru spljoštenosti.
Okvirna analiza za mere spoljoštenosti ukazuje na sledeće:

$\alpha_4 = 3$	raspored ima normalnu spljoštenost
$\alpha_4 > 3$	raspored je izdužen, ima spljoštenost manju od normalne
$\alpha_4 < 3$	raspored je sabijen, ima spljoštenost veću od normalne

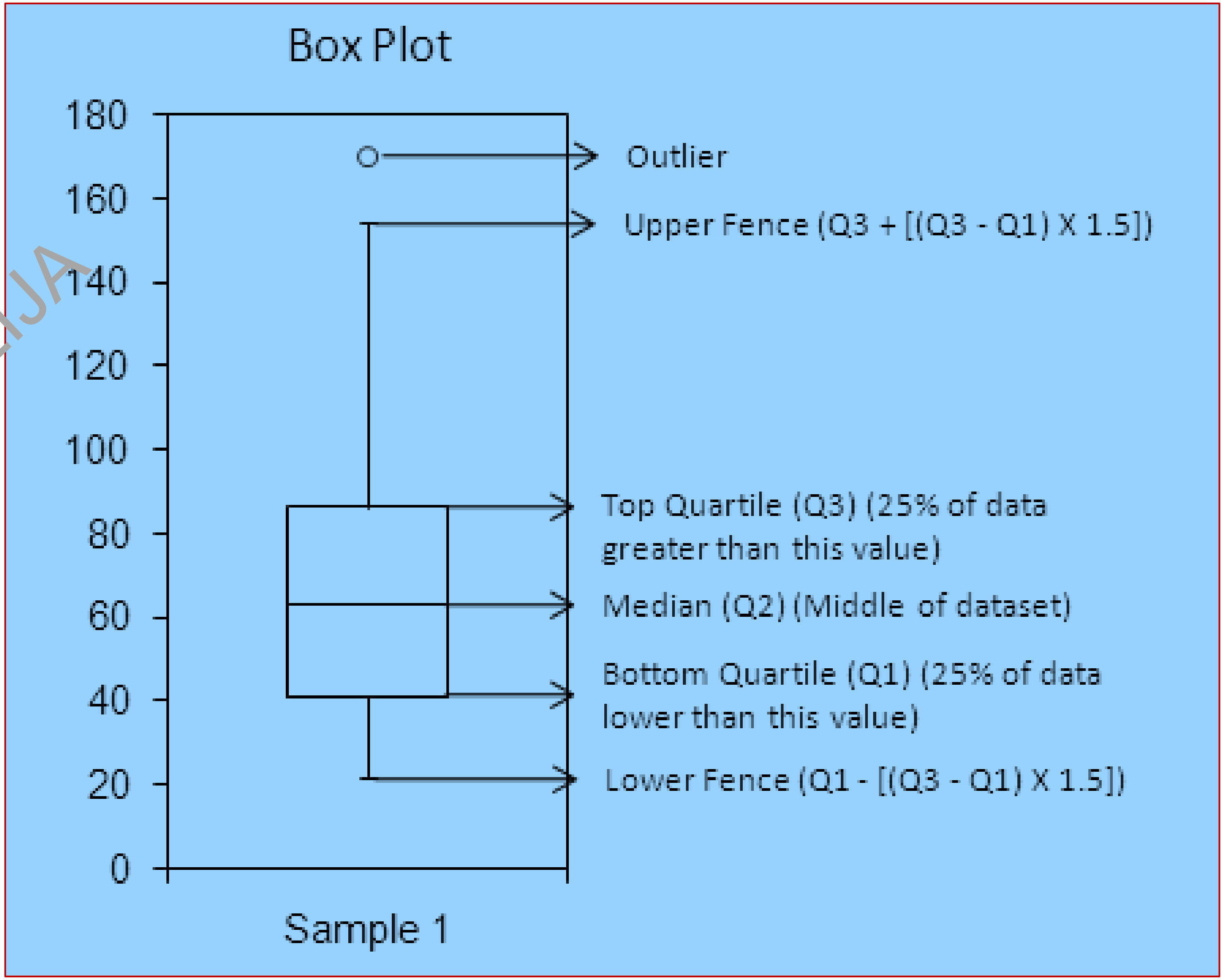
Deskriptivna Statistika / Asimetricnost & Mode, Median, Mean



RADNA VERZIJA

Deskriptivna Statistika / Box box-and-whiskers plot

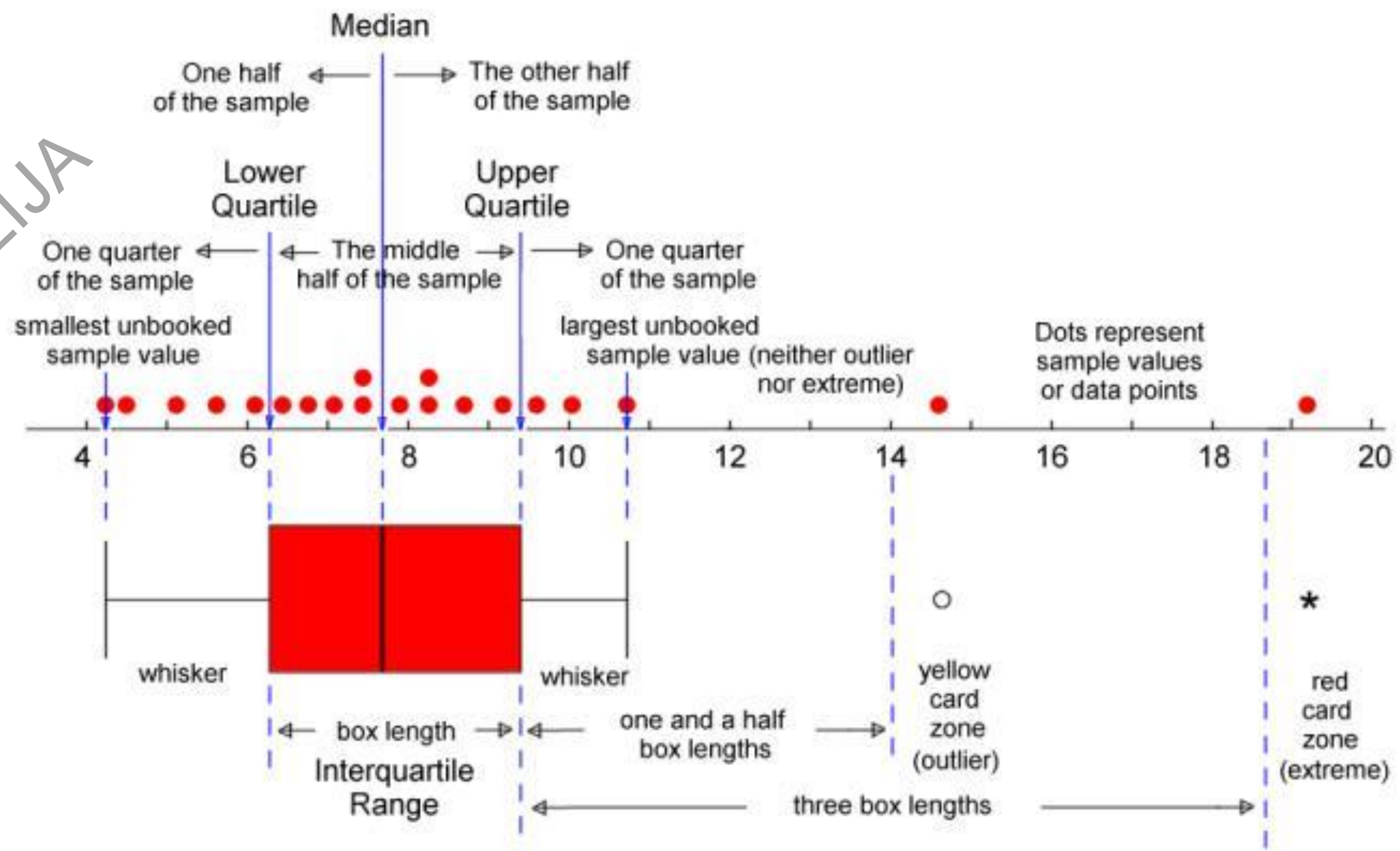
RADNA VERZIJA



VER. 01

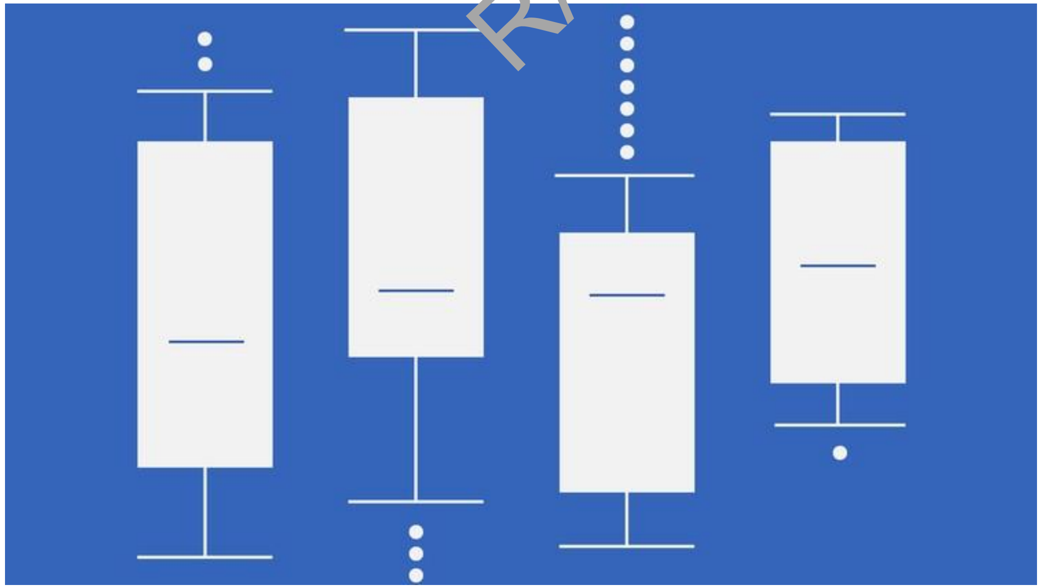
Deskriptivna Statistika / Box box-and-whiskers plot

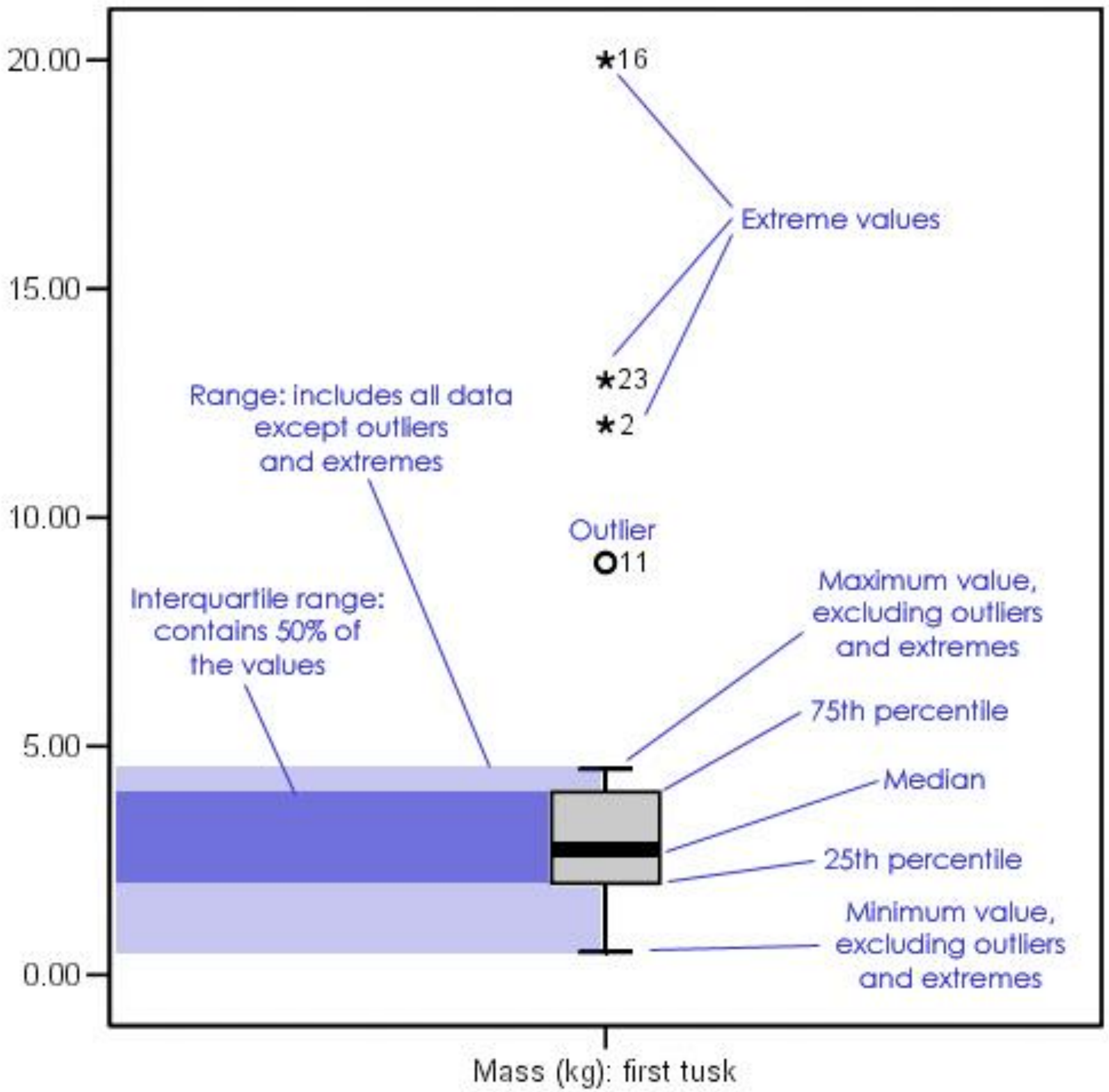
RADNA VERZIJA



Deskriptivna Statistika / Box box-and-whiskers plot

RADNA VERZIJA





Dobrodosli u svet podataka!