Deskriptivne mjere

- Numeričko opisivanje podataka
  - Centralna tendencija
    - Aritmetička sredina
    - Medijana
    - Modus
    - Geometrijska sredina
  - Varijacija (Disperzija)
    - Raspon
    - Varijansa
    - Standardna devijacija
    - Koeficijent varijacije
  - Oblik
    - Spljoštenost

Mjere centralne tendencije

- Centralna tendencija – tendencija koja se ispoljava kao koncentrisanje podatka oko određenih numeričkih vrijednosti u statističkoj seriji
- Mjere lokacije (modus i medijana)
- Izračunate mjere centralne tendencije (aritmetička i geometrijska sredina)

Mjere centralne tendencije

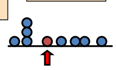
Pregled

Centralna tendencija

Aritmetička sredina

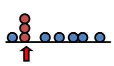
$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Medijana



Središna vrijednost opservacije (centralna opservacija)

Modus



Opservirana vrijednost sa najvećom frekvencijom

Geometrijska sredina

$$\overline{X}_G = (X_1 \times X_2 \times \cdots \times X_n)^{1/n}$$

Mjere disperzije

Srednja vrijednost karakteriše, u izvjesnom smislu, dati raspored kao mjera centralne tendencije vrijednosti obilježja, ali ona nije dovoljna karakteristika, jer drugi rasporedi mogu imati istu srednju vrijednost a različite varijacije (raspršenost ili disperziju)

Mjere disperzije

Statistički opis skupa kao i uzorka iziskuje zato pored mjera lokacije centralne tendencije i odgovarajuće mjere varijacije ili disperzije.

Za mjerenje disperzije jedne serije koristi se više mjera od kojih neke imaju apsolutni a neke relativni izraz.

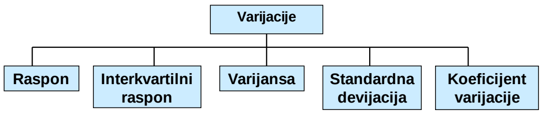
Apsolutne mjere disperzije

Apsolutne mjere disperzije iskazuju varijabilitet u apsolutnim iznosima onih mjernih jedinica u kojima su dati modaliteti posmatranog obilježja: u milionima dinara, hiljadama tona, kilometrima, komadima i sl.

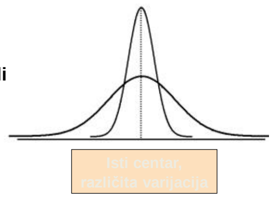
Ove mjere kao i mjere lokacije centralne tendencije mogu biti:

- ❑ **pozicione**
- ❑ **izračunate**

Mjere disperzije



- Mjere varijacije daju informaciju o **disperziji ili varijabilnosti** vrijednosti obilježja, odnosno podataka.



Apsolutne mjere disperzije

Od pozicionih mjera najčešće se koristi **razmak** ili **interval varijacije.**

Razmak ili interval varijacije predstavlja razliku između najviše i najniže vrijednosti obilježja u seriji.

$$\text{Interval varijacije} = x_{\max} - x_{\min}$$

8

Apsolutne mjere disperzije

Ova mjera (koja ima smisla samo za konačne razlike) izračunava se jednostavno, ali ona daje približnu informaciju o disperziji serije, jer na nju utiče samo krajnje vrijednosti posmatranog obilježja, koje se mogu znatno razlikovati od ostalih vrijednosti.

Drugi nedostatak je u tome što je razlika minimalne i maksimalne vrijednosti obilježja kod većih serija (čak i ako one pokazuju velike varijabilitete) po pravilu veća nego kod malih serija.

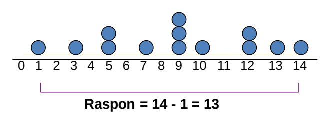
9

Razmak (raspon)

- Najjednostavnija mjera varijacije
- Razlika između najveće i najmanje vrijednosti obilježja u statističkoj seriji:

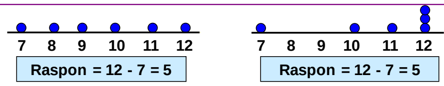
$$\text{Raspon} = X_{max} - X_{min}$$

Primjer:



Nedostaci raspona

- Ignoriše se raspodjelapodataka



- Osjetljiv na ekstremne vrijednosti

1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,4,5

Raspon = 5 - 1 = 4

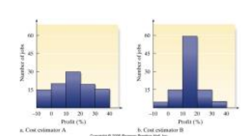
1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,4,120

Raspon = 120 - 1 = 119

Nedostaci raspona

- Gubi smisao kod jako velikih uzoraka

- Ove 2 raspodjelecije imaju isti raspon.
- Koliko vam raspon govori o varijabilnosti podataka?



Apsolutne mjere disperzije

Da bi se eliminisao uticaj ekstremnih vrijednosti na iznos razmaka, odnosno intervala varijacije, izračunava se kao dopunska mjera **interkvartilna razlika**, to jest razlika između trećeg i prvog kvartila.

Apsolutne mjere disperzije

Prvi kvartil Q1 – prvih 25% elemenata skupa

Drugi kvartil Q2 - prvih 50% elemenata skupa
= Medijana

Treći kvartil Q3 – prvih 75% elemenata skupa

Apsolutne mjere disperzije

Prvi kvartil Q1 je vrijednost obilježja od koje 25% elemenata skupa uređenih po veličini ima manju ili jednaku vrijednost.

$$Q_1 = L_1 + \frac{\frac{N}{4} - \sum f_1}{f_{Q1}} * i$$

L1- donja granica grupnog intervala za prvi kvartil

∑f1- suma frekvencija do kvartilnog intervala

FQ1- frekvencija kvartilnog intervala

i – dužina intervala

Apsolutne mjere disperzije

Treći kvartil Q3 je vrijednost obilježja od koje 75% elemenata skupa uređenih po veličini ima manju ili jednaku vrijednost.

$$Q_3 = L_1 + \frac{\frac{3N}{4} - \sum f_1}{f_{Q3}} * i$$

L1- donja granica grupnog intervala za prvi kvartil

∑f1- suma frekvencija do kvartilnog intervala

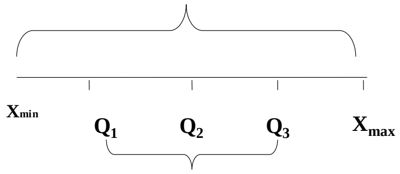
FQ3- frekvencija kvartilnog intervala

i – dužina intervala

Apsolutne mjere disperzije

Interkvartalna razlika $i_q = Q_3 - Q_1$

Razmak – interval varijacije



Interkvartalna razlika

Apsolutne mjere disperzije

Ako je razmak, odnosno interval varijacije veliki a interkvartilna razlika mala, znači da na krajevima rasporeda postoje ekstremne vrijednosti, ali da ostali članovi serije ne pokazuju velike varijabilitete.

Apsolutne mjere disperzije

Preciznije podatke o varijabilitetu posmatrane serije daju pokazatelji čije se izračunavanje zasniva na odstupanju srednje vrijednosti, najčešće aritmetičke sredine, od svih vrijednosti obilježja koje ta serija sadrži.

Odstupanje pojedinih vrijednosti obilježja od aritmetičke sredine biće:

$$d_1 = x_1 - \bar{x}; \ d_2 = x_2 - \bar{x}; \ \cdots d_n = x_n - \bar{x}$$

19

Apsolutne mjere disperzije

Algebarski zbir odstupanja aritmetičke sredine od vrijednosti obilježja jednak je 0.

Zato se umjesto algebarskih polazi od apsolutnih odstupanja aritmetičke sredine od vrijednosti obilježja, čiji prosjek predstavlja mjeru varijabiliteta poznatu pod nazivom **srednje apsolutno odstupanje.**

20

Apsolutne mjere disperzije

Za negrupisane podatke izračunava se po formuli:

$$\bar{d} = \frac{1}{N}\sum_{i=1}^{N}\left|x_i - \bar{x}\right|$$

A za grupisane podatke po formuli:

$$\bar{d} = \frac{1}{N}\sum_{i=1}^{k} f_i\left|x_i - \bar{x}\right|$$

gdje je:

$$N = f_1 + f_2 + \ldots + f_k = \sum_{i=1}^{k} f_i$$

21

Apsolutne mjere disperzije

Pošto je prosjek odstupanja pojedinih vrijednosti obilježja od aritmetičke sredine jednak nuli, možemo uzeti kao mjeru disperzije prosjek kvadrata odstupanja, koja se naziva **varijansom**.

Za seriju negrupisanih podataka izračunava se kao:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2$$

A za grupisane podatke:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{k} f_i\left(x_i - \bar{x}\right)^2$$

22

Apsolutne mjere disperzije

Obrazac za izračunavanje varijanse za grupisane podatke daljim raščlanjivanjem svodimo na sledeći obrazac:

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

23

Apsolutne mjere disperzije

Pošto je varijansa iskazana u mjernim jedinicama na kvadrat, uzima se njen pozitivan kvadratni korjen i dobija najčešće korišćena apsolutna mjera disperzije – **standardna devijacija.**

24

Apsolutne mjere disperzije

Standardna devijacija za negrupisane podatke dobija se kao:

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

Standardna devijacija za grupisane podatke izračunava se kao:

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

25

Apsolutne mjere disperzije

Varijansa uzorka predstavlja porosjek sume kvadrata odstupanja vrijednosti od aritmetičke sredine uzorka. Za grupisane podatke izračunava se kao:

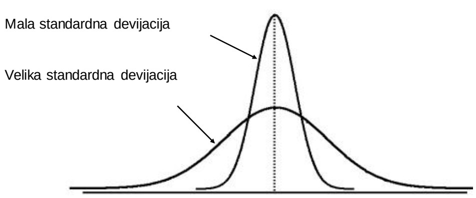
$$s^2 = \frac{\sum f_i (x_i - \bar{x})^2}{n-1}$$

Standardna devijacija uzorka izračunava se kao pozitivan kvadratni korjen iz varijanse uzorka:

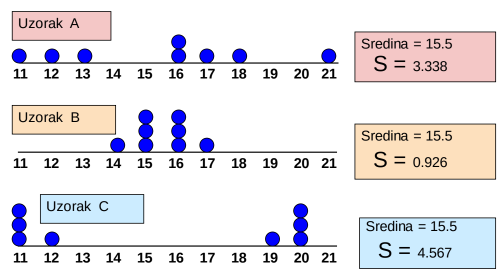
$$s = \sqrt{s^2} = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum f_i x_i^2 - n\bar{x}^2}{n-1}}$$

26

Mjerenje varijacije



Upoređivanje standardnih devijacija



Prednosti varijanse i standardne devijacije

- Za njihovo izračunavanje koristi se svaki podatak iz skupa
- Vrijednosti opservacija koje su daleko od aritmetičke sredine imaju ekstra težinu
  (jer se odstupanja od sredine kvadriraju)