



Mere varijabiliteta (disperzije)

predavanja

Značaj statističkih mera varijabiliteta

- Mere varijabilnosti imaju ključnu ulogu u statistici jer nam pružaju uvid u to koliko su podaci u skupu raspršeni ili promenljivi. Dok mere centralne tendencije, kao što su srednja vrednost ili medijana, opisuju "središnju tačku" skupa podataka, mere varijabilnosti nam govore koliko podaci odstupaju od te tačke.
- Zašto su značajne?
- **Razumevanje distribucije podataka:** Pomažu nam da procenimo da li su podaci grupisani blizu proseka ili su raspršeni.
- **Upoređivanje skupova podataka:** Omogućavaju poređenje stabilnosti ili homogenosti između dva ili više skupova podataka.
- **Predviđanje i analize:** Veće varijabilnosti mogu ukazivati na nepredvidivost u podacima, dok manje varijabilnosti sugerišu veću konzistentnost.
- **Praćenje trendova:** Uočavanje odstupanja ili anomalija koje bi mogle biti značajne za analizu.
- U poslovnom okruženju, razumevanje varijabilnosti u prihodima **pomaže u planiranju i proceni rizika.**

Mere varijabiliteta (disperzije)

- Dve serije sa istom aritmetičkom sredinom mogu imati potpuno različite raspodele frekvencije, tj. njihova raspršenost može biti potpuno različita
- Centralne tendencije pružaju uvid u **lokaciju** ili prosečnu vrednost podataka, dok varijabilnost otkriva **stabilnost** ili odstupanja podataka.



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



3, 3, 3, 3, 3, 3, 3, 3, 3.

Mere varijabiliteta

Osnovna karakteristika vrednosti jednog istog obeležja je, da ta vrednost varira od jedne do druge statističke jedinice osnovnog skupa. Te vrednosti, mere centralne tendencije, a pre svega aritmetička sredina, sažimaju u jednu brojčanu vrednost. Ta brojčana vrednost treba da bude reprezentativna za sve vrednosti. Njena reprezentativnost zavisi od stepena varijabilnosti pojedinačnih vrednosti u odnosu na centralnu vrednost, konkretno u odnosu na aritmetičku sredinu. Ukoliko je varijabilnost manja, utoliko su vrednosti obeležja sabijenije oko aritmetičke sredine (manje odstupaju) i ona je reprezentativnija, a za takav skup kažemo da je homogen. Obrnuto, ako je varijabilnost veća, odstupanje pojedinačnih vrednosti od aritmetičke sredine je veće, a reprezentativnost aritmetičke sredine je manja i za takav skup kažemo da je heterogen.



Mere varijabiliteta

Sledi zaključak da mere varijabilnosti zapravo ukazuju na reprezentativnost mera centralne tendencije. Manja mera varijabilnosti ukazuje na veću reprezentativnost srednje vrednosti i obrnuto. Mere varijabilnosti nas opredeljuju koju od mera centralne tendencije treba da koristimo, aritmetičku sredinu (ukoliko je skup homogen), ili medijanu (ukoliko je skup heterogen).

Mere varijabiliteta

Мере дисперзије указују на две чињенице:

- ✓ Колико су средње вредности обележја добри представници свих вредности на посматраном скупу података;
- ✓ Колико се све вредности датог обележја на посматраном скупу међусобно разликују.

Односно: Што је варијанса мања, то је средња вредност бољи представник, и обрнуто.



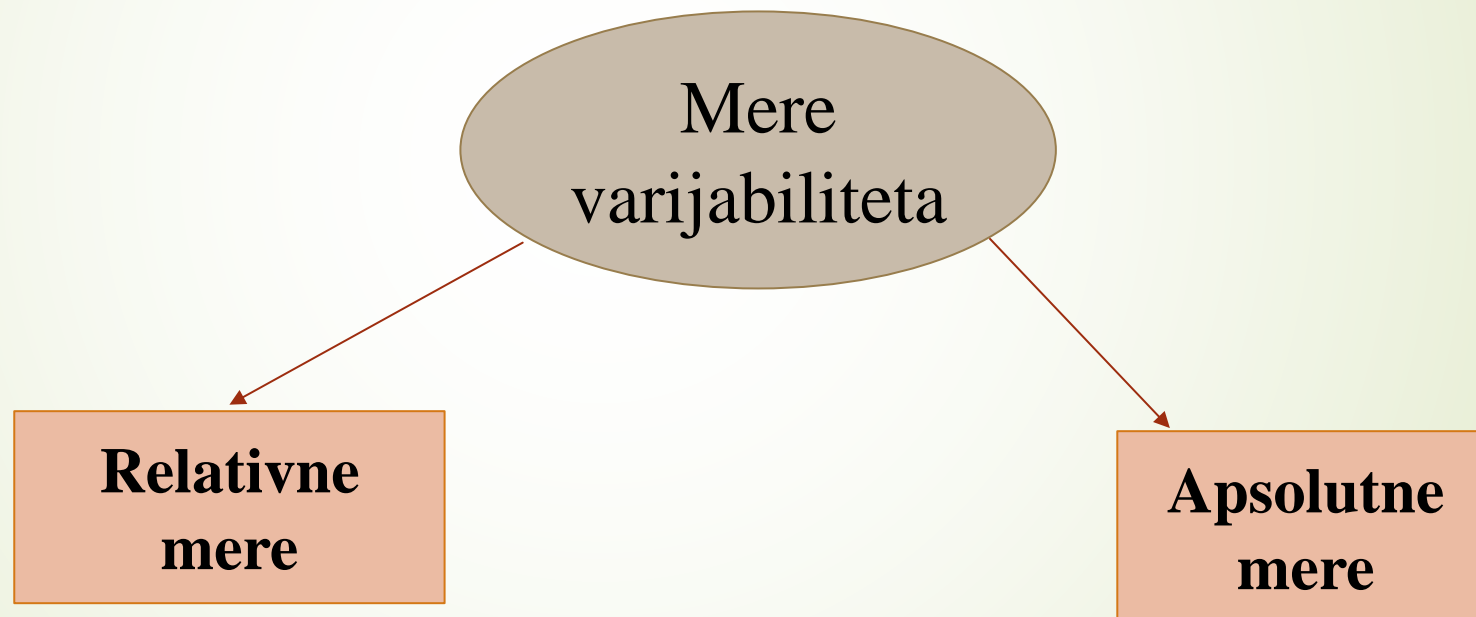
Mere varijabiliteta

➤ Definicija:

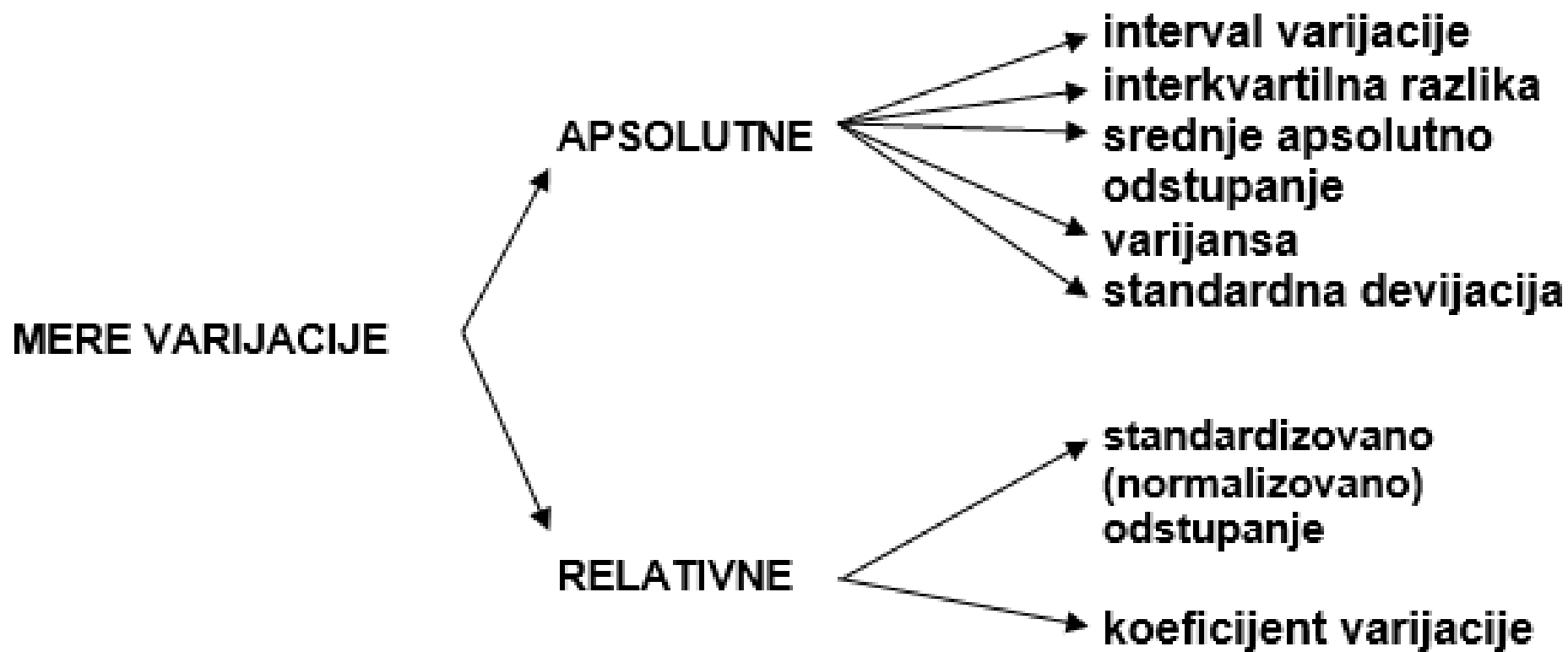
Mere varijabiliteta (disperzije) su numeričke vrednosti koje opisuju u kom stepenu se vrednosti u posmatranoj seriji međusobno razlikuju.

Pomoću ovih mera izražava se stepen varijabilnosti statističkih podataka.

Mere varijabiliteta



Mere varijabiliteta



Apsolutne mere disperzije

Apsolutne mere disperzije su mere koje izražavaju raspršenost ili varijabilnost podataka u apsolutnim jedinicama merenja, tj. u istim jedinicama u kojima su podaci iskazani. One **ne uzimaju** u obzir relativne odnose među podacima i fokusiraju se na apsolutne razlike i odstupanja. Ove mere su korisne kada radimo s podacima u istim jedinicama i želimo da kvantifikujemo ukupnu varijabilnost.

Osnovne apsolutne mere disperzije:

1. Interval varijacije-Raspon (Range):
2. Interkvartilni opseg (Interquartile Range - IQR):
3. Srednje apsolutno odstupanje
4. Varijansa
5. Standardna devijacija

Apsolutne mere disperzije

- **Interval varijacije** je mera raspršenosti koja je jednaka razlici najveće i najmanje vrednosti u seriji podataka

- $I_V = X_{\max} - X_{\min}$

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

- Kod intervalnih distribucija frekvencija, interval varijacije je razlika između gornje granice poslednje grupe i donje granice prve grupe.
- Interval varijacije se lako izračunava, međutim, njegova primena je ograničena jer se zasniva samo na dve vrednosti obeležja.

Interval varijacije

► Primer:

1, 2, 5, 6, 6, 1, 3, 7, 3, 3, 3, 3.

$R_v=6$

Apsolutne mere disperzije

- **Interkvartilna razlika** je razlika između vrednosti obeležja na mestu trećeg i prvog kvartila, tj.

$$i_q = Q_3 - Q_1$$

- Prvi kvartil (Q_1) je vrednost obeležja na mestu $1/4$ sredene serije, tj.

$$Q_1 = x_{\frac{n+1}{4}}$$

- a treći kvartil je vrednost obeležja na mestu $3/4$ sredene serije, odnosno:

$$Q_3 = x_{3\frac{n+1}{4}}$$

- Za intervalne serije kvartili se određuju pomoću sledećih formula:

$$Q_1 = g_1 + \frac{\frac{\sum_{i=1}^k f_i}{4} - K_{i-1}}{f_{q_1}} \cdot d$$

$$Q_3 = g_1 + \frac{3\frac{\sum_{i=1}^k f_i}{4} - K_{i-1}}{f_{q_3}} \cdot d$$

$g_{1(3)}$ - donja granica grupnog intervala u kojem se nalazi mesto prvog (trećeg) kvartila;

f_{q_1} - frekvencija grupnog intervala u kojem se nalazi mesto prvog kvartila;

f_{q_3} - frekvencija grupnog intervala u kojem se nalazi mesto trećeg kvartila;

K_{i-1} - kumulativna frekvencija koja prethodi toj klasi i

d - dužina klasnog intervala.

Apsolutne mere disperzije

► *Srednje apsolutno odstopanje* D (ili SO)

- Srednje apsolutno odstopanje se izračunava tako što se zbir absolutnih odstopanja svih vrednosti obeležja od aritmetičke sredine podeli brojem članova serije.

- Za negrupisane podatke formula je:
$$SO = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

- Za grupisane podatke formula je:
$$SO = \frac{\sum_{i=1}^k f_i |x_i - \bar{x}|}{\sum_{i=1}^k f_i}$$

Apsolutne mere disperzije

- **Varijansa** meri prosečno kvadratno odstopanje vrednosti od aritmetičke sredine. Što je varijansa veća, podaci su više raspršeni; što je manja, podaci su koncentrisaniji oko srednje vrednosti.

Varijansa skupa je

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N}$$

$$\sigma^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N}$$

Varijansa uzorka je

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

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

Apsolutne mere disperzije

Varijansa grupisanog skupa je

$$\sigma^2 = \frac{\sum f_i (x_i' - \mu)^2}{N} = \frac{\sum f_i x_i'^2 - \frac{(\sum f_i x_i')^2}{N}}{N}$$

Varijansa grupisanog uzorka je

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

gde je σ^2 varijansa skupa, s^2 varijansa uzorka,
 x_i' sredina intervala.

Apsolutne mere disperzije

► **Standardna devijacija skupa**

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

► **Standardna devijacija uzorka**

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

Varijansa i standardna devijacija su uvek nenegativne (≥ 0).
Varijansa = 0 akko $x_1 = x_2 = \dots = x_N = \text{const.}$

Izračunavanje apsolutnih mera disperzije u Excelu

- Izračunavanje apsolutnih mera disperzije u Excelu može se obaviti pomoću ugrađenih funkcija i alata.
- **Raspon (Range)**
- Uneti podatke u jednu kolonu (npr. A1:A10).
- Koristiti formulu: $\text{=MAX(A1:A10) - MIN(A1:A10)}$
- **Interquartile Range – IQR**
- Sortirati podatke u rastućem redosledu.
- Ručno pronaći prvi kvartil (Q1) i treći kvartil (Q3) koristeći funkciju $\text{=QUARTILE(A1:A10, 1)}$ za Q1 i $\text{=QUARTILE(A1:A10, 3)}$ za Q3.
- Izračunati IQR: =Q3 - Q1 .

Izračunavanje srednjeg apsolutnog odstupanja u Excelu

- **Srednje apsolutno odstupanje (Mean Absolute Deviation - MAD)**
- Izračunati aritmetičku sredinu koristeći `=AVERAGE(A1:A10)`.
- Za svaku vrednost u skupu podataka, izračunati apsolutno odstupanje od srednje vrednosti: `=ABS(A1 - AVERAGE(A1:A10))`.
- Izračunati prosečnu vrednost apsolutnih odstupanja koristeći `=AVERAGE(B1:B10)` (pretpostavljajući da su odstupanja u koloni B).

Izračunavanje varijanse u Excelu

- Izračunavanje varijanse u Excelu je jednostavno uz pomoć ugrađenih funkcija.
- **Koraci za izračunavanje varijanse**
- Uneti podatke:- Uneti sve vrednosti u jednu kolonu (npr. A1:A10).
- Koristiti funkciju za varijansu:
 - Za populaciju: Koristiti funkciju VAR.P (npr. =VAR.P(A1:A10)).
 - Za uzorak: Koristiti funkciju VAR.S (npr. =VAR.S(A1:A10)).
- Prikaz rezultata:- Excel automatski izračunava varijansu i prikazuje rezultat u ćeliji gde je uneta formula.



Izračunavanje standardne devijacije u Excelu

- Za uzorak: Koristiti funkciju =STDEV.S(A1:A10).
- Za populaciju: Koristite funkciju =STDEV.P(A1:A10)
- Excel takođe ima alatku "**Data Analysis**" koja omogućava automatsko izračunavanje ovih mera

Izračunavanje apsolutnih mera disperzije primenom MATLABa

1. Raspon (Range)

Koristite funkciju `max` i `min`:

Matlab

```
data = [2, 4, 6, 8, 10];  
range = max(data) - min(data);  
disp(['Raspon: ', num2str(range)]);
```

Izračunavanje apsolutnih mera disperzije primenom MATLABa

► Interquartile Range - IQR

Matlab

```
data = [2, 4, 6, 8, 10];  
iqr_value = iqr(data);  
disp(['Međukvartilski raspon: ', num2str(iqr_value)]);
```


Izračunavanje apsolutnih mera disperzije primenom MATLABa

► Srednje apsolutno odstupanje (Mean Absolute Deviation - MAD)

Koristiti funkciju **mad**

Matlab

```
data = [2, 4, 6, 8, 10];  
mad_value = mad(data, 1); % '1' označava apsolutno odstupanje  
disp(['Srednje apsolutno odstupanje: ', num2str(mad_value)]);
```


Izračunavanje apsolutnih mera disperzije primenom MATLABa

Koristiti funkciju **var**:

Matlab

```
data = [2, 4, 6, 8, 10];  
variance = var(data);  
disp(['Varijansa: ', num2str(variance)]);
```

Izračunavanje apsolutnih mera disperzije primenom MATLABa

5. Standardna devijacija (Standard Deviation)

Koristite funkciju `std`:

Matlab

```
data = [2, 4, 6, 8, 10];  
std_dev = std(data);  
disp(['Standardna devijacija: ', num2str(std_dev)]);
```