

1838

УНИВЕРЗИТЕТ У БЕОГРАДУ  
ФИЛОЗОФСКИ ФАКУЛТЕТ

Статистика у психологији 1

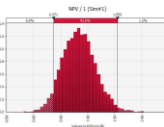
Статистика у истраживању образовања

доц. др Анђела Шошкић, 19.10.2024.

# Опис узорка у погледу једне квантитативне варијабле

## Графичко представљање квантитативних података

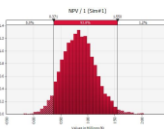
Пре него што почнемо, да се упознамо 😊



# А сад, назад на квантитативне варијабле...

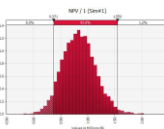
- **Подсетник:**

- Шта су варијабле?
- Шта су подаци?
- Варијабле могу бити квантитативне и ...
- Какве су то квантитативне варијабле?
- Дајте ми неки пример!
- Ком типу статистике припада опис узорка у погледу једне квантитативне варијабле – дескриптивној или инференцијалној?



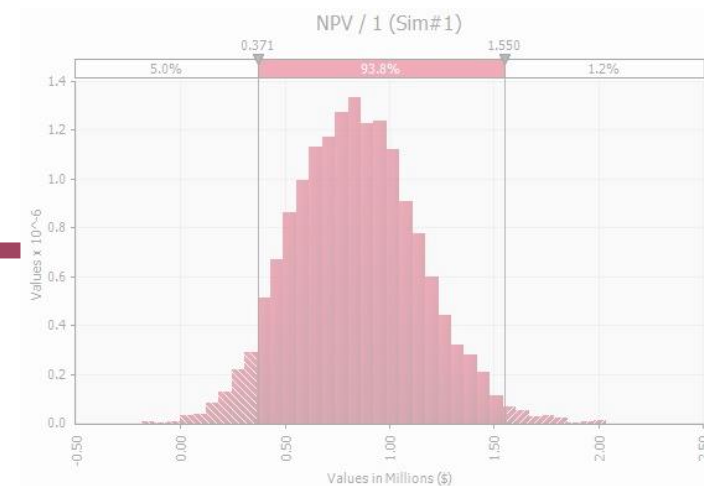
# Теме за данас

1. Шта су резултати на квантитативној варијабли, математички гледано?
2. Уређивање квантитативних резултата:
  - сортирање, рангирање, пребројавање (и израчунавање расподеле вредности), груписање
3. Графички приказ расподеле квантитативних резултата
4. Облик расподеле нумеричких резултата
5. Нумерички опис расподеле квантитативних резултата:
  - Мере локације и централне тенденције, варијабилности, облика расподеле
6. Графички приказ расподеле са мерама ЦТ и варијабилности и упутства за прављење графика
7. Закључак: шта треба да садржи опис једне квантитативне варијабле

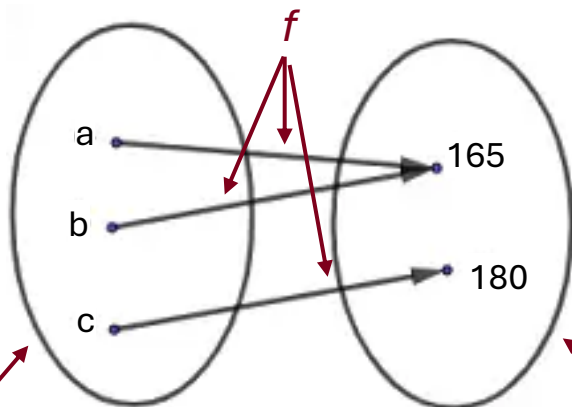


# 1. Квантитативни резултати описани математичким терминима

---



# Шта су резултати на квантитативној варијабли, математички гледано?

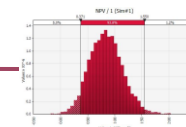


скуп јединица посматрања  
(обично испитаника)  
= узорак

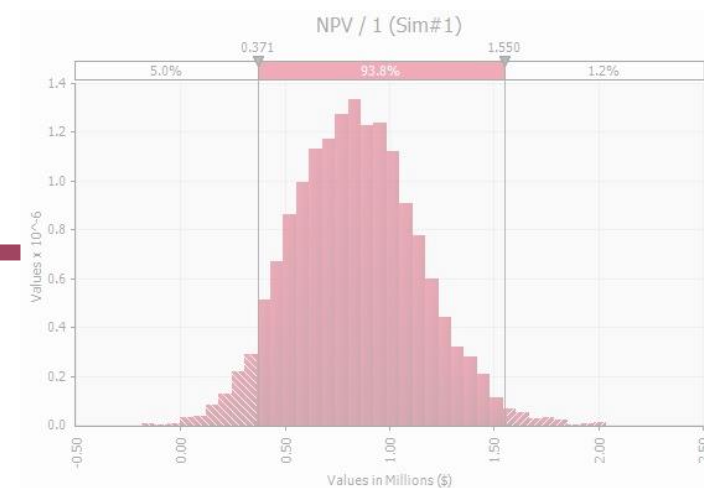
скуп могућих вредности  
нумеричке варијабле

Испитаник	Варијабла V
a	165
b	180
c	180

вектор  $x$  са резултатима  
варијабле V



## 2. Уређивање квантитативних резултата



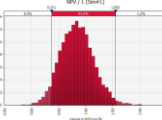
# Сортирање резултата

Табела сортирана по узлазном редоследу за висину

Испитаник (јединица посматрања)	Висина (квант. варијабла)
E1	180,1
E2	150,3
E3	165,0
E4	162,2
E5	171,6
E6	175,5
E7	192,4
E8	189,0
E9	175,5

Испитаник (јединица посматрања)	Висина (квант. варијабла)
E2	150,3
E4	162,2
E3	165,0
E5	171,6
E6	175,5
E9	175,5
E1	180,1
E8	189,0
E7	192,4

Редоследни статистици  $X_1, X_2, \dots$





# Рангирање резултата

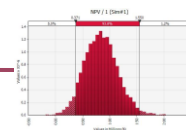
Сортирано по узлазном редоследу за висину

Испитаник (јединица посматрања)	Висина (квант. варијабла)	Ранг
E2	150,3	1
E4	162,2	2
E3	165,0	3
E5	171,6	4
E6	175,5	5,5
E9	175,5	5,5
E1	180,1	7
E8	189,0	8
E7	192,4	9

Оригинални редослед испитаника

Испитаник (јединица посматрања)	Висина (квант. варијабла)	Ранг
E1	180,1	7
E2	150,3	1
E3	165,0	3
E4	162,2	2
E5	171,6	4
E6	175,5	5,5
E7	192,4	9
E8	189,0	8
E9	175,5	5,5

дељени рангови

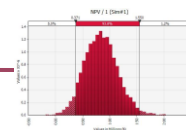


# Јединична расподела (релативних) учестаности

Испитаник (јединица посматрања)	Висина (квант. варијабла)
E1	180,1
E2	150,3
E3	165,0
E4	162,2
E5	171,6
E6	175,5
E7	192,4
E8	189,0
E9	175,5

Вредности висине (варијабле)	Учестаност/ фреквенција = f
150,3	1
162,2	1
165,0	1
171,6	1
175,5	2
180,1	1
189,0	1
192,4	1

$$p_k = f_k / n$$



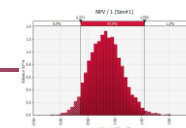
# Расподела (релативних) учестаности са групним интервалима

Вредности висине (варијабле)	Учестаност/ фреквенција = f	Релативна учестаност = p
150,3	1	0.11 (11.1%)
162,2	1	0.11 (11.1%)
165,0	1	0.11 (11.1%)
171,6	1	0.11 (11.1%)
175,5	2	0.22 (22.2%)
180,1	1	0.11 (11.1%)
189,0	1	0.11 (11.1%)
192,4	1	0.11(11.1%)

Интервали вредности висине (варијабле)	Учестаност/ фреквенција = f	Релативна учестаност = p
150-160,0	1	0,11 (11,1%)
160,1-170,0	2	0,22 (22.2%)
170,1-180,0	3	0,33 (33.3%)
180,1-190,0	2	0,22 (22.2%)
190,1-200,0	1	0,11 (11,1%)

Опис интервала:

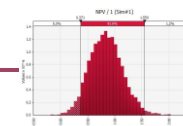
- горња и доња егзактна граница
- средишње место



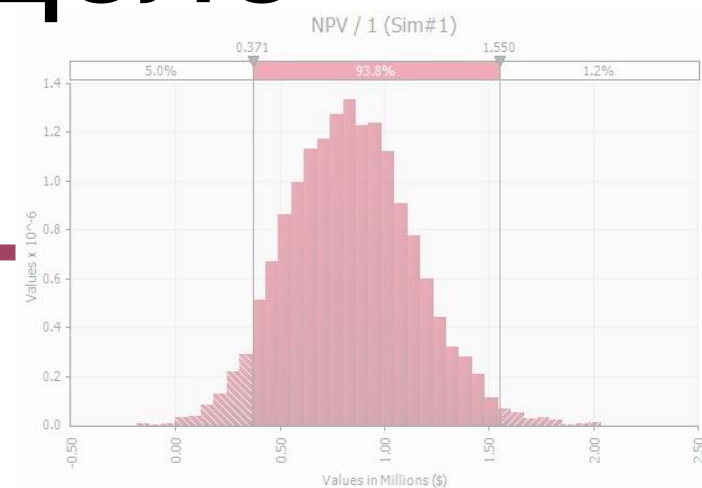
# Расподела (релативних) кумулативних учестаности

Интервали вредности висине (варијабле)	Учестаност/ фреквенција = f	Кумулативна фреквенција = cf	Релативна кумулативна фреквенција = rcf (пропорције)
150-160,0	1	1	0.11
160,1-170,0	2	3	0.33
170,1-180,0	3	6	0.66
180,1-190,0	2	8	0.88
190,1-200,0	1	9	1

$$cf_k = \sum_{i=1}^k f_i \quad rcf_k = cf_k / n$$



# 3. Графички приказ распределе квантитативних резултата

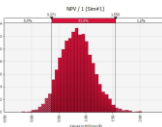
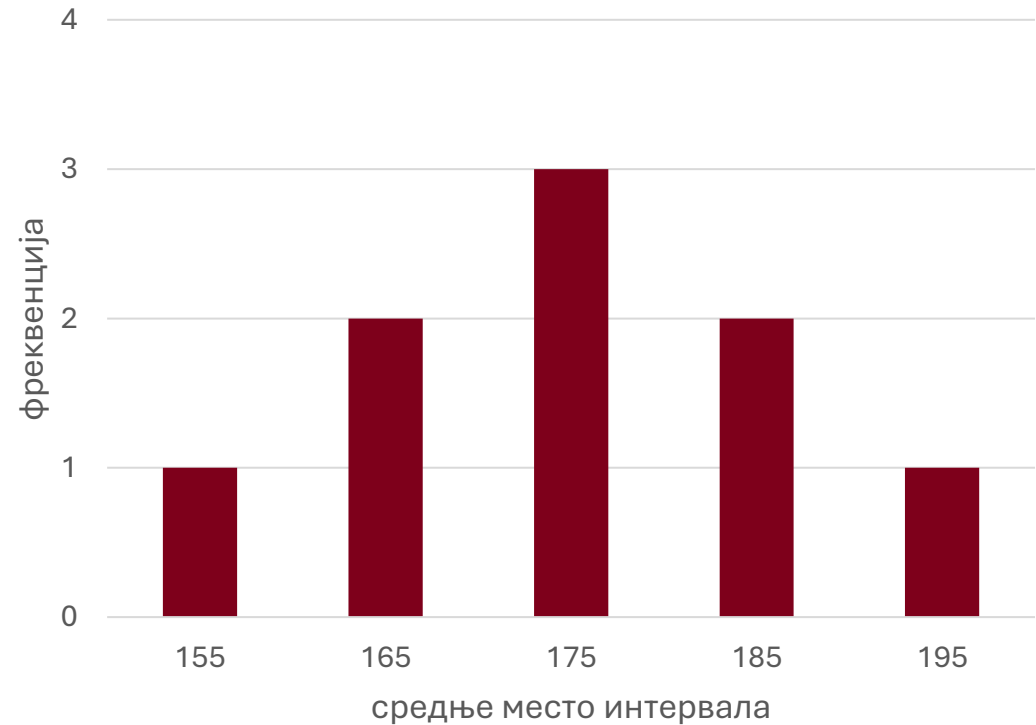


# Графички приказ расподеле учестаности

## Полигон фреквенција

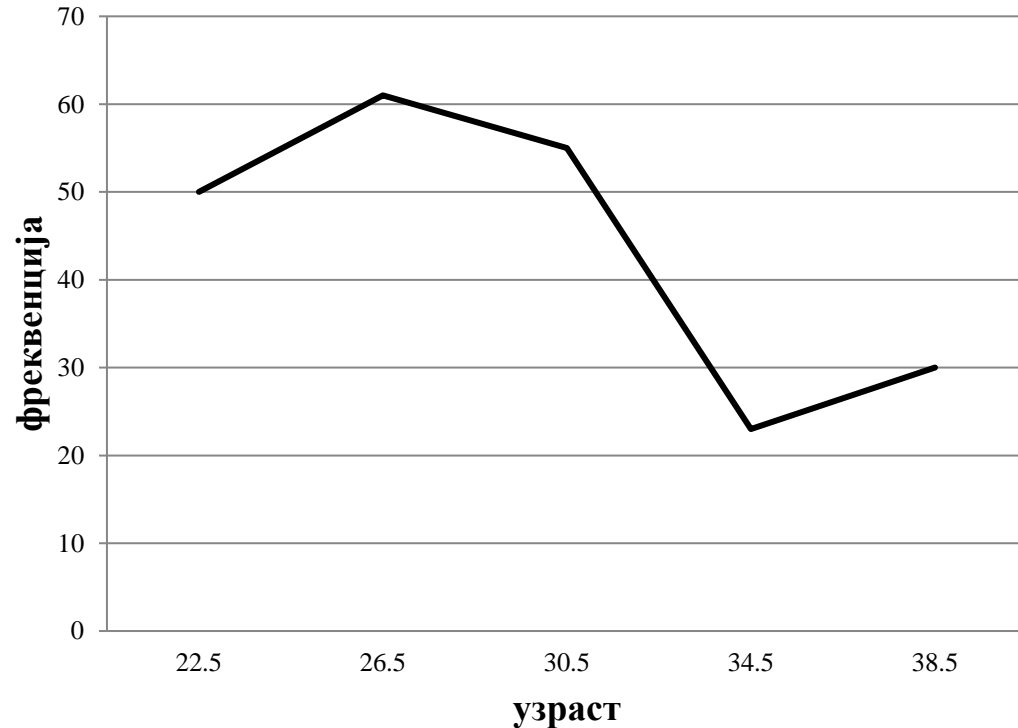


## Хистограм

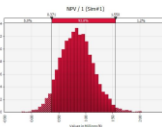
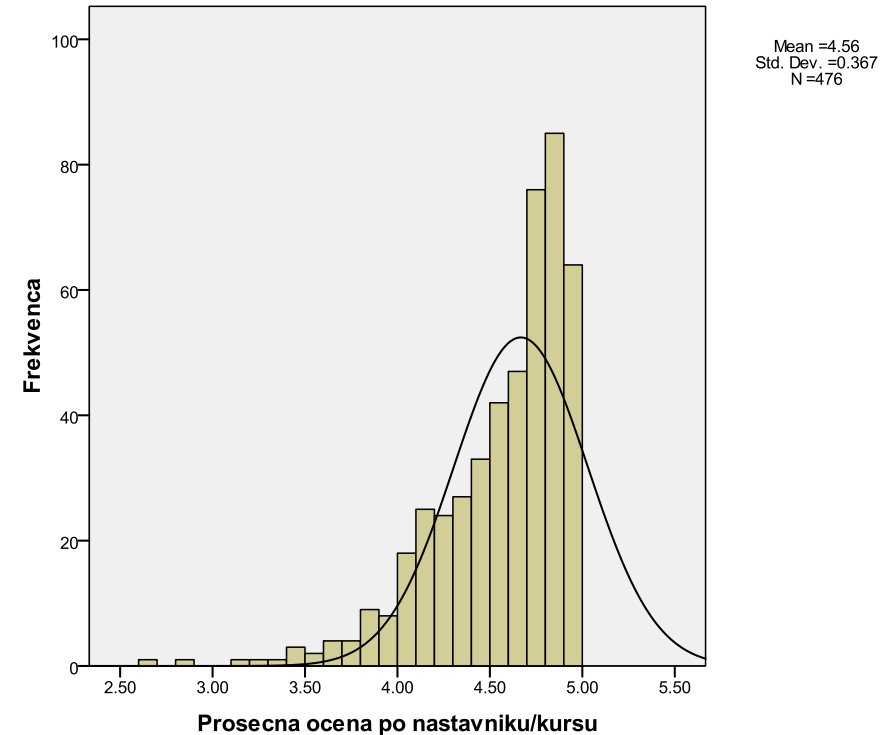


# Графички приказ расподеле учестаности

## Полигон фреквенција

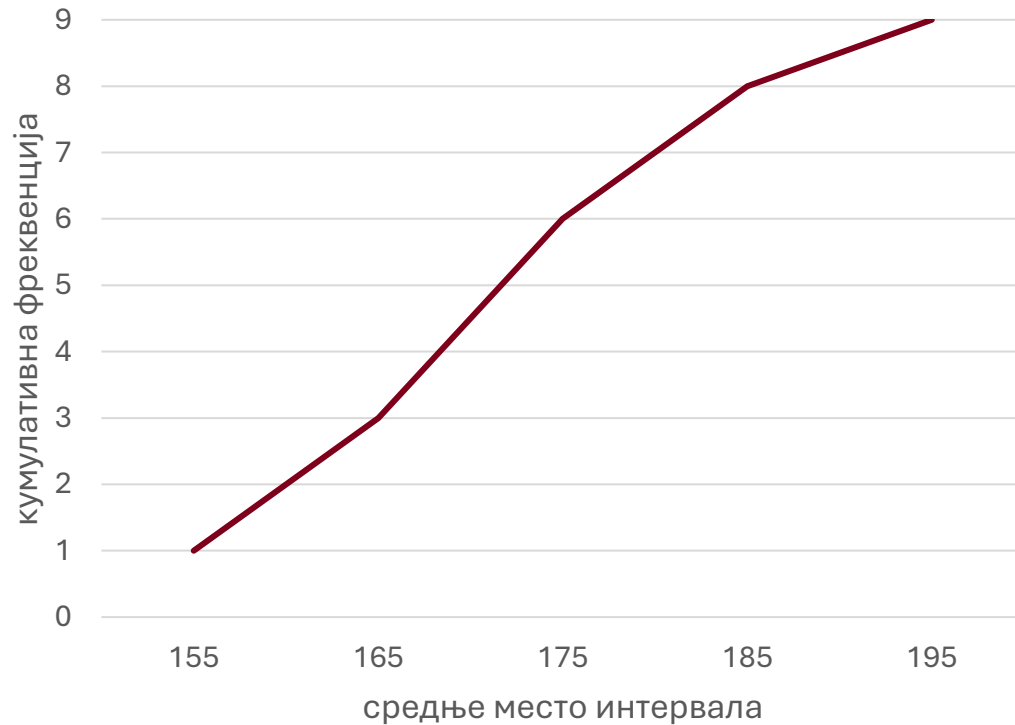


## Хистограм

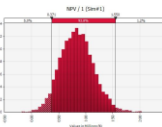
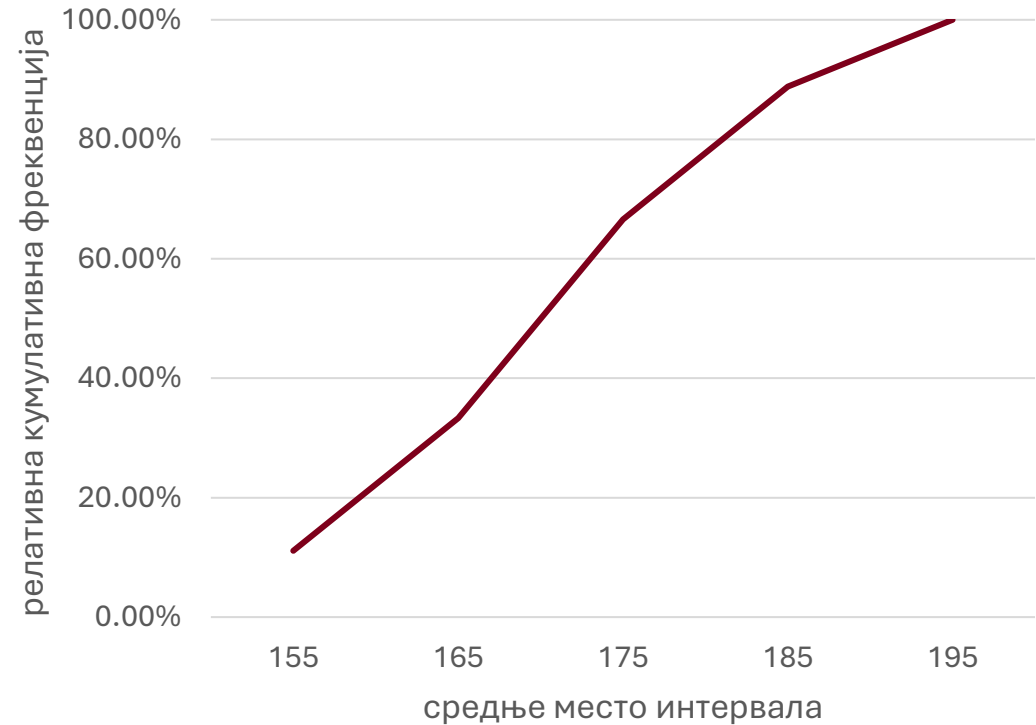


# Графички приказ расподеле кумулативне учестаности

## Огива



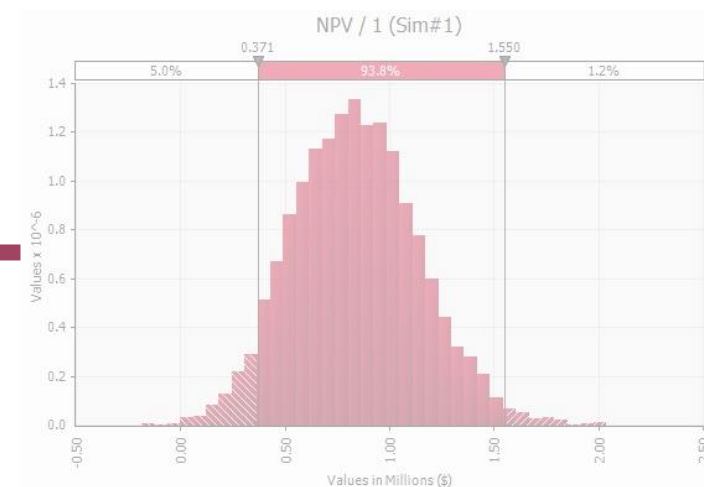
## Огива – релативне фреквенције





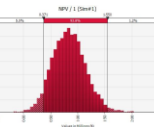
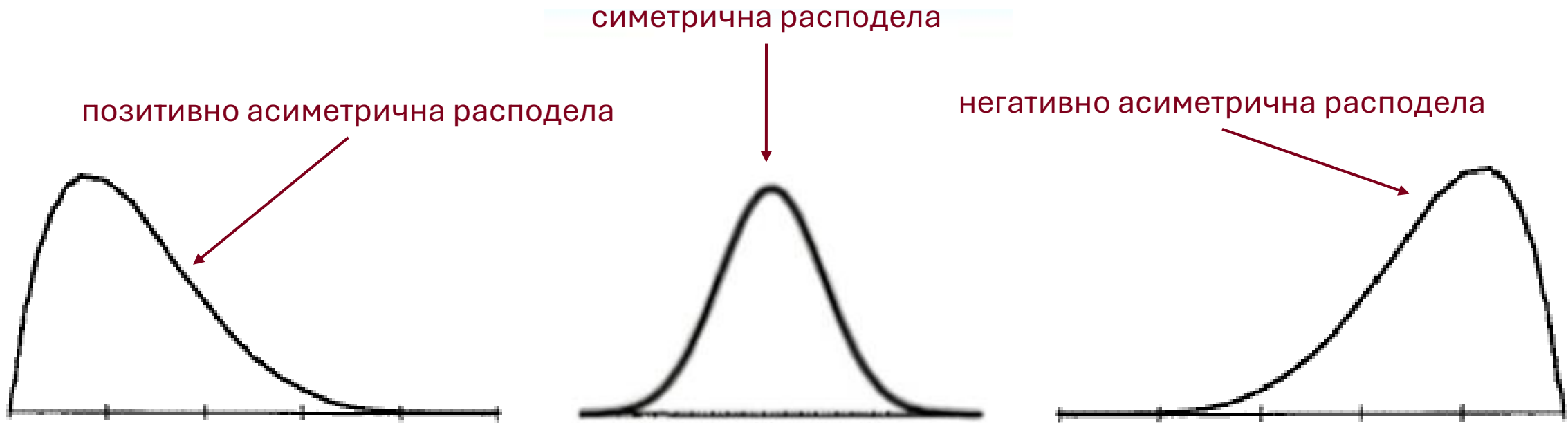
# 4. Облици расправе квантитативних резултата

---



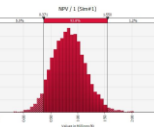
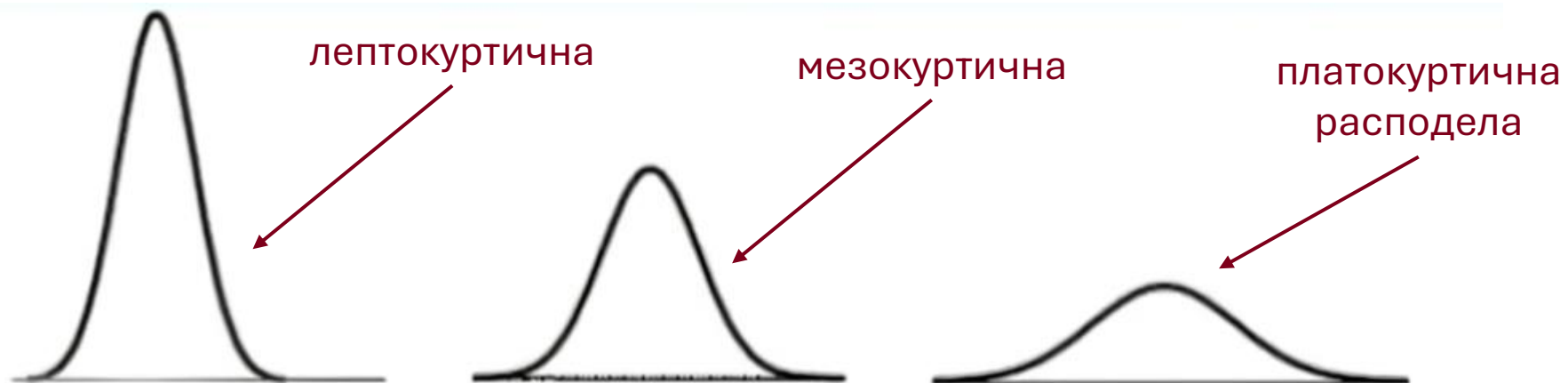
# Унимодалне расподеле

- имају само један брег
- симетричне и асиметричне



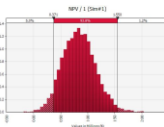
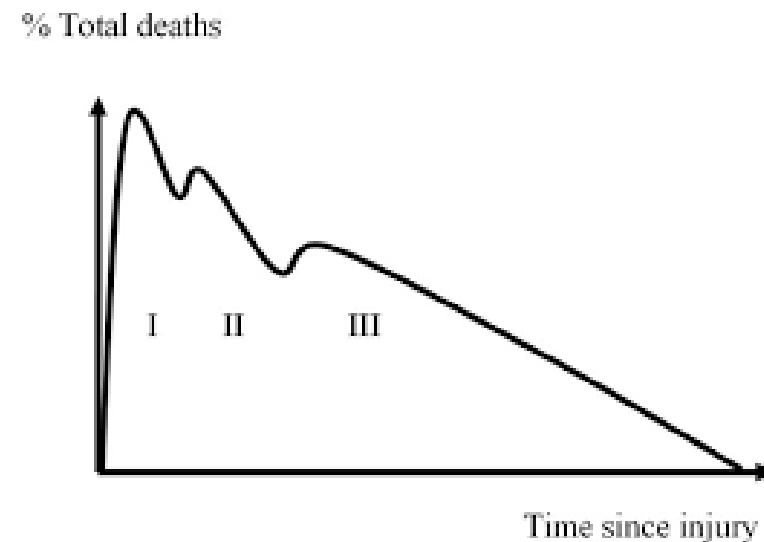
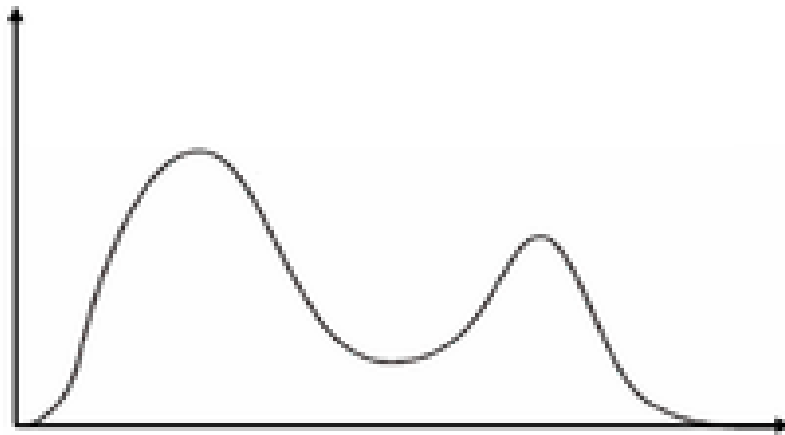
# Унимодалне расподеле

- „спљоштене и шиљасте“ – лептокуртичне, мезокуртичне, платокуртичне
- нормална расподела – унимодална, мезокуртична, симетрична – о детаљима више касније



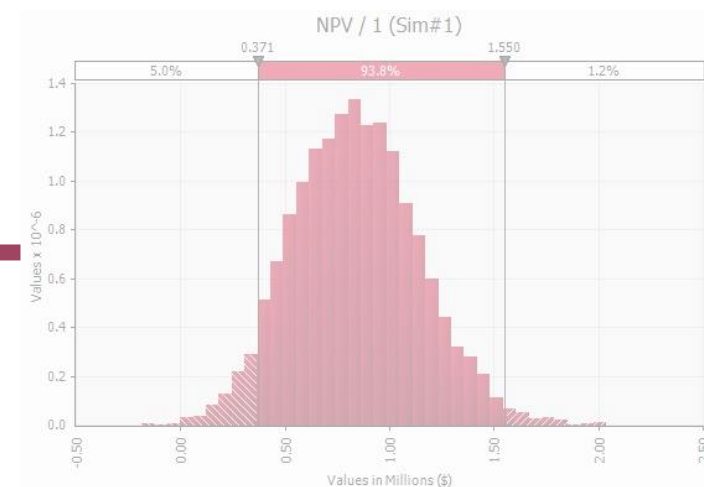
# Мултимодалне расподеле

- више брегова
  - ако су два – бимодалне
  - један чест пример у психологији: U расподеле



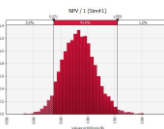
# 5. Нумерички опис квантитативних резултата

---

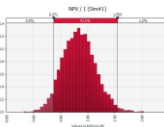
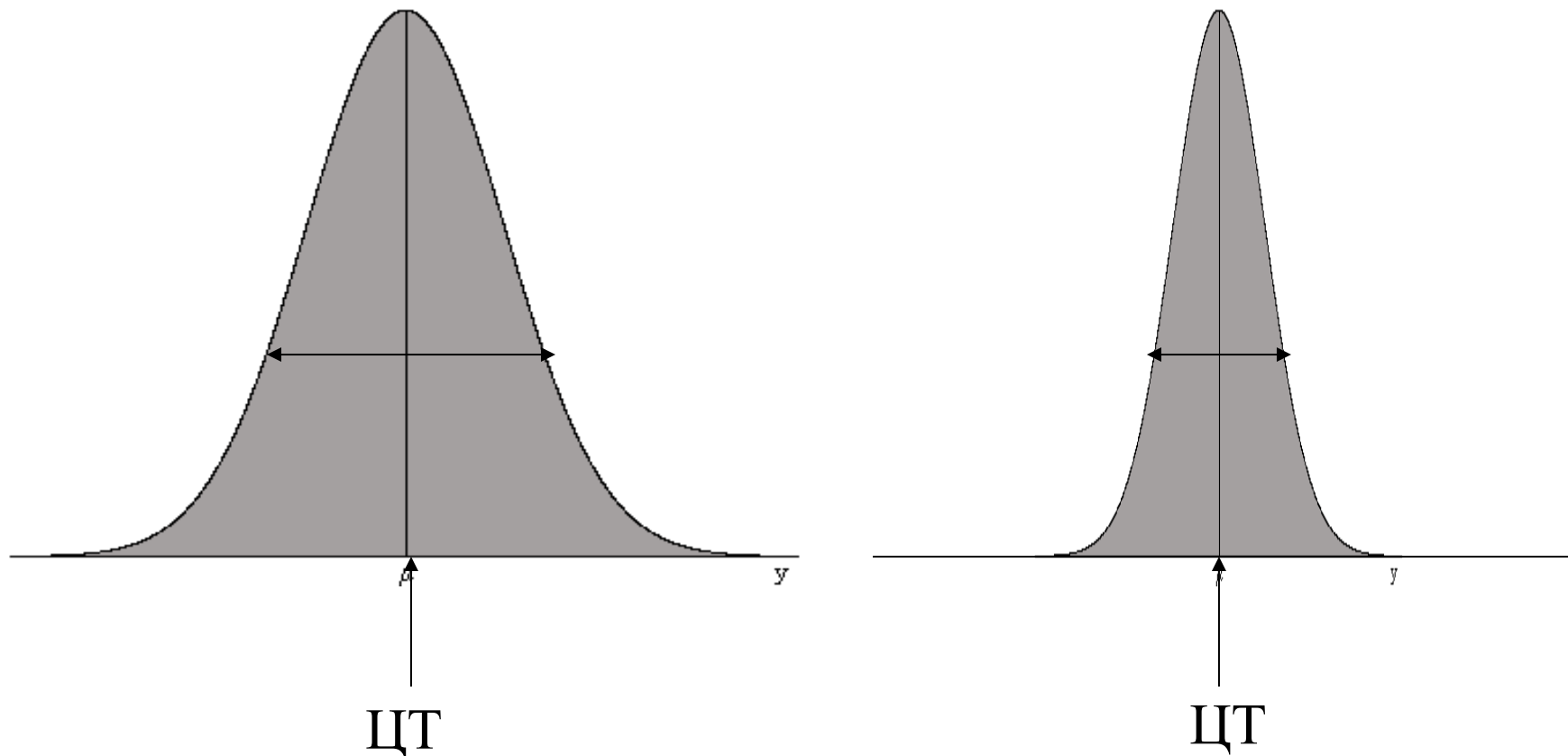


# Мере локације

- мере варијабле које указују на неку локацију (место) у расподели која је од интереса
- јединица мерења: углавном иста као и за саму варијаблу
- често:
  - места око којих се „нагомилавају“ подаци
  - мере средине расподеле (= мере *централне тенденције*)



# Централна тенденција



# Мере централне тенденције

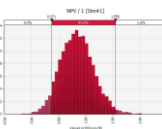
- **Аритметичка средина (M, некад и AS)**

- просек
- збир свих појединачних мера, подељен бројем мера

$$M = \frac{\sum_{i=1}^n x_i}{n}$$

- $M = (180,1 + 150,3 + 165,0 + 162,2 + 171,6 + 175,5 + 192,4 + 189,0 + 175,5) / 9 = 173,5 \text{ cm}$

Испитаник	Висина
E1	180,1
E2	150,3
E3	165,0
E4	162,2
E5	171,6
E6	175,5
E7	192,4
E8	189,0
E9	175,5





# Мере централне тенденције

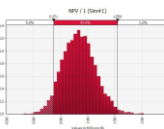
## • Медијана (Mdn)

- вредност изнад и испод које се налази половина резултата
  - за непаран број резултата – вредност тачно у средини
  - за паран број – просек две вредности које су у средини
- рачуна се из редоследних статистика (тј. кад сортирамо низ

$$\text{Mdn} = X_{(m)}; \quad m = \frac{n+1}{2}; \quad \text{Mdn} = \frac{1}{2}(X_{(m)} + X_{(m+1)}); \quad m = \left[ \frac{n+1}{2} \right]$$

- **Mdn = 175,5 cm**

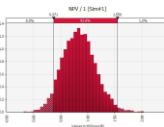
Испитаник	Висина
E2	150,3
E4	162,2
E3	165,0
E5	171,6
E6	175,5
E9	175,5
E1	180,1
E8	189,0
E7	192,4



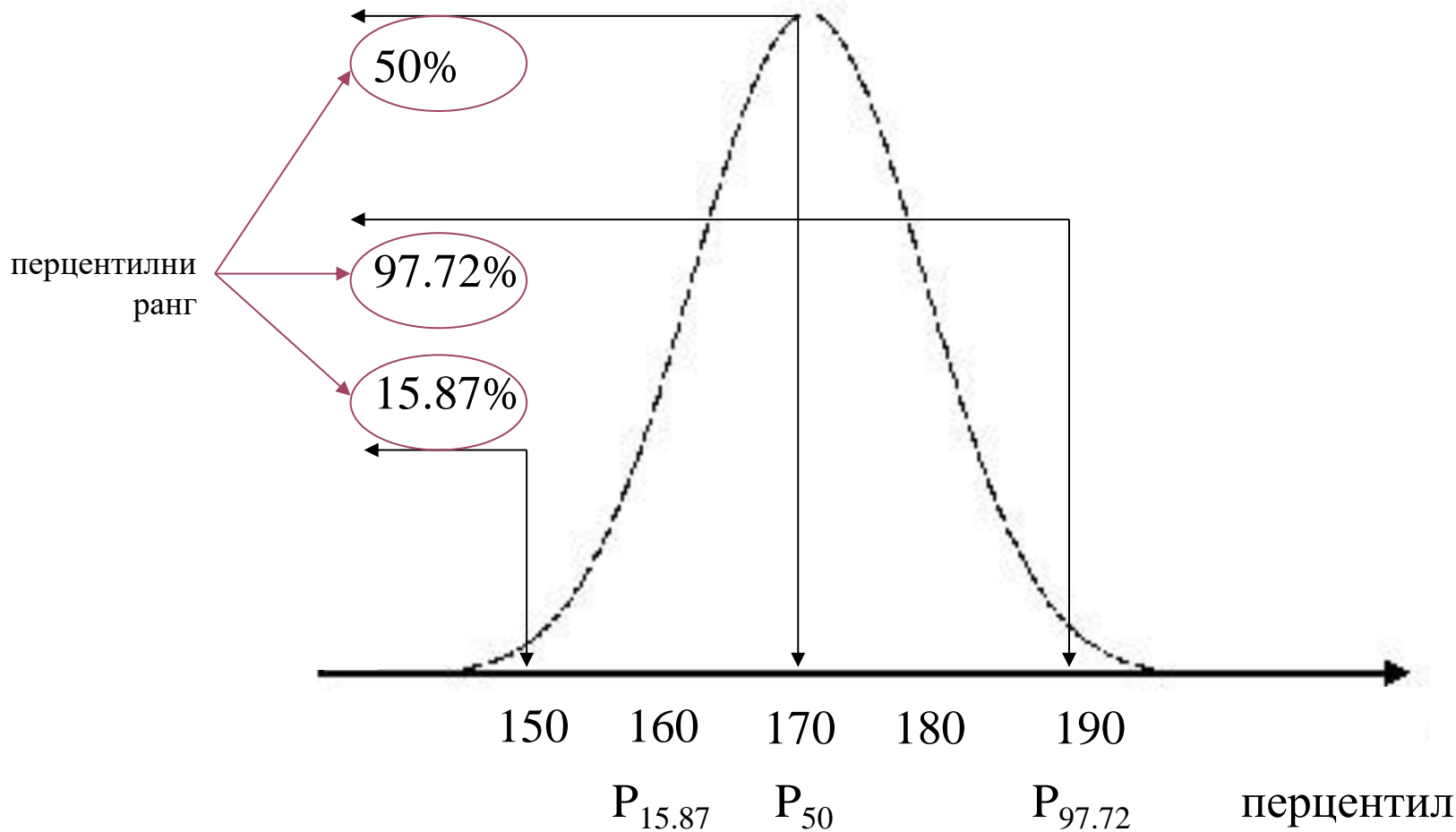
# Мере централне тенденције

- Мод (Mod)
  - најчесталија мера у узорку
  - **Mdn = 175,5**

Испитаник	Висина
E1	180,1
E2	150,3
E3	165,0
E4	162,2
E5	171,6
E6	175,5
E7	192,4
E8	189,0
E9	175,5



# Друге мере локације: перцентили

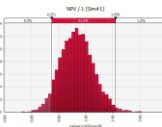


## • Перцентил

- $P_n$ : вредност варијабле испод које је  $n\%$  резултата у узорку
- $P_{88.8} = 190 \text{ cm}$

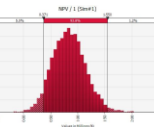
## • Перцентилни ранг

- $PR_x$ : проценат резултата који се налазе испод неке вредности  $x$  на варијабли
- $PR_{190} = 88.8\%$

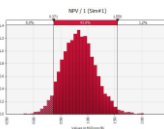
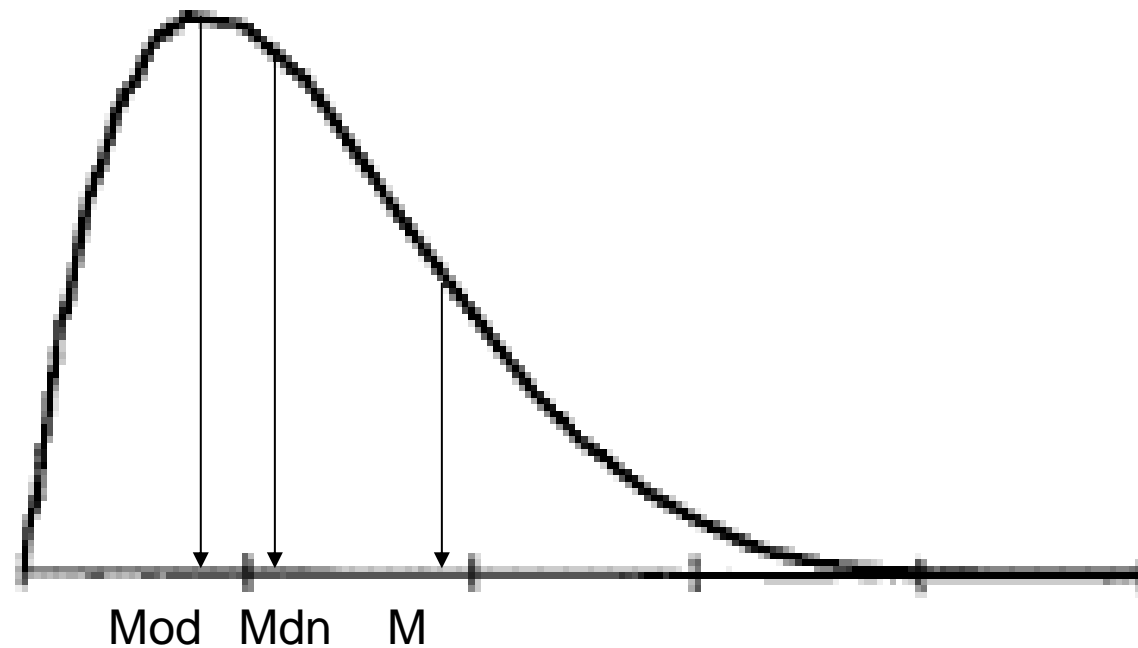


# Друге мере локације: квартили, децили, квинтили, квантили

- **квартили:** четвртине расподеле
  - $Q_1 = P_{25}$
  - $Q_2 = P_{50}$
  - $Q_3 = P_{75}$
  - $Q_4 = P_{100}$
- **квинтили:** петине расподеле
- **децили:** десетине расподеле
- **квантили:** исто као перцентили, али изражени пропорцијом

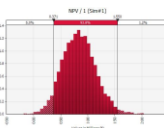


# Облик расподеле и мере централне тенденције



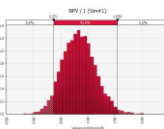
# Како изабрати меру централне тенденције?

- **Аритметичка средина** најбоља кад имамо приближно нормалну расподелу вредности у узорку
  - тада, а и иначе кад је расподела симетрична и унимодална (има један мод) вредности  $M$ ,  $Mdn$ ,  $Mod$  су исте/приближно исте
  - $M$  је најзгоднија за математичке операције и даје најбољу оцену исте мере ЦТ у популацији
  - **али:** јако осетљива на екстремне вредности (изнимке, аутлајере)
    - $M = (180,1 + 150,3 + 165,0 + 162,2 + 171,6 + 175,5 + 192,4 + 189,0 + 175,5) / 9 = 173,5$
    - $M = (180,1 + 150,3 + 165,0 + 162,2 + 171,6 + 175,5 + 192,4 + 189,0 + \mathbf{1750,5}) / 9 = \mathbf{348,5 \text{ cm}}$

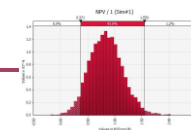
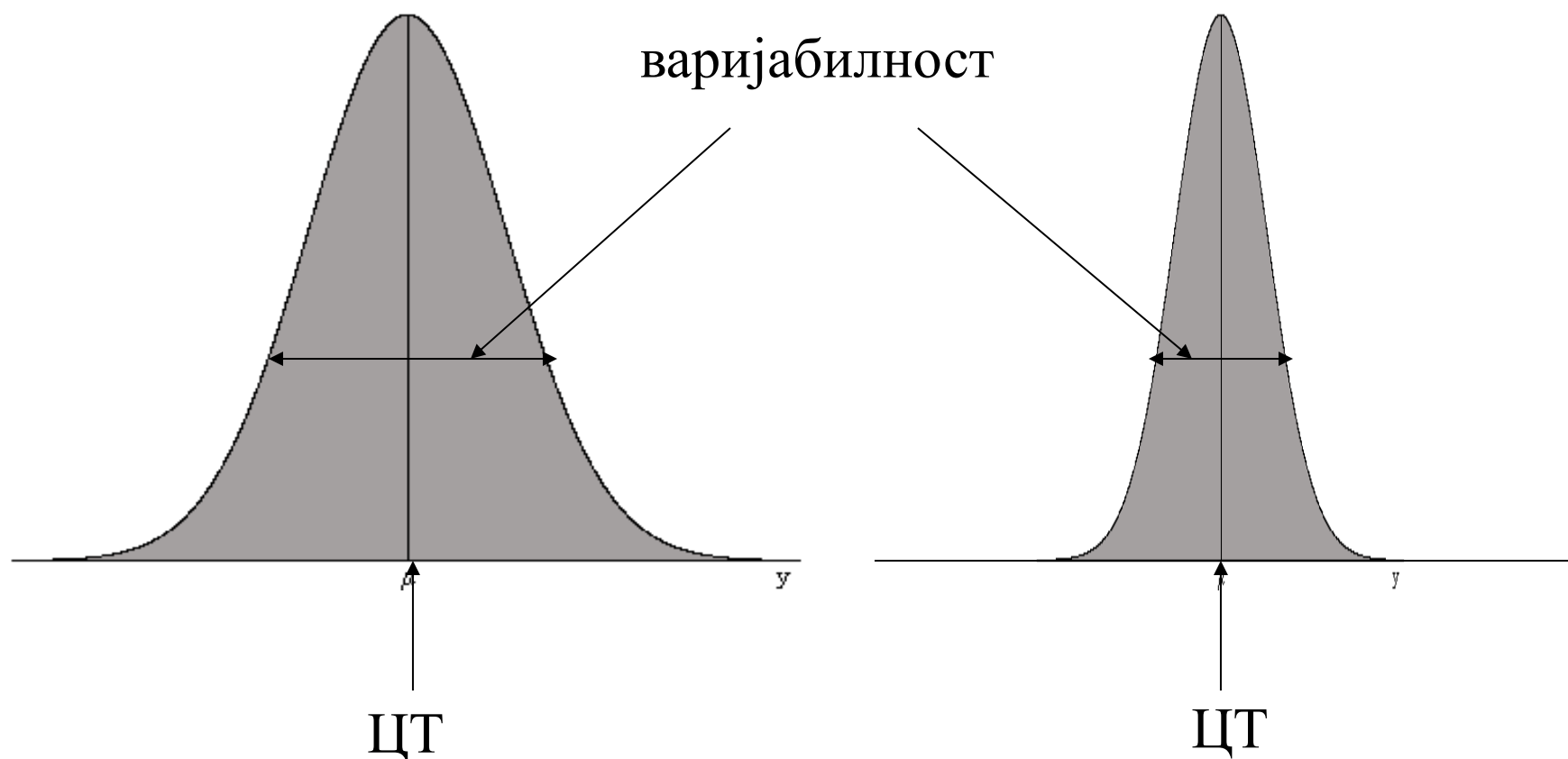


# Како изабрати меру централне тенденције?

- **Медијана** је боља ако:
  - расподела није нормална јер има пуно вредности на екстремима са једне од страна, или са обе стране (асиметрична или платокуртична расподела), јер је мање осетљива на екстреме
  - **али:** и она је погодна само за *унимодалне расподеле* (имају један мод)
  - **Mdn = 175,5 cm** у оба случаја
- **Мод** је врло користан ако имамо мултимодалну расподелу (нпр. U кривуље)
  - **Mod = 175,5 cm** у оба случаја
- Ако је расподела таква да је тешко описати је иједном од мера, онда дати приказ минимума, максимума, медијане, првог и трећег квантила ( $P_0, P_{25}, P_{50}, P_{75}, P_{100}$ ) = **петобројни сажетак**



# Варијабилност (распршење)





# Мере распршења (варијабилности)

- **Распон**

- разлика између најмање и највеће вредности
- $R = x_{\max} - x_{\min}$
- екстремно осетљив на изнимке – сваки га мења

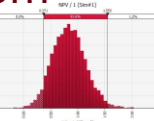
- **Полуинтерквартилни распон (квартилна девијација)**

- половина разлике између трећег и првог перцентила
- $Q = (P_{75} - P_{25}) / 2$
- мање осетљив на изнимке
- обично се користи за расподеле које се много разликују од нормалне заједно са медијаном

Испитаник	Висина
E2	150,3
E4	162,2
E3	165,0
E5	171,6
E6	175,5
E9	175,5
E1	180,1
E8	189,0
E7	192,4

$$R = 192,4 - 150,4 = 42,1 \text{ cm}$$

$$Q = (180,1 - 165) / 2 = 15,1 \text{ cm}$$



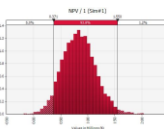
# Мере распршења (варијабилности)

- **Варијанса ( $SD^2$ ,  $S^2$ ,  $\sigma^2$ )**

- мера одступања од аритметичке средине
- просечно *квадрирано одступање* од просека

$$S^2 = \frac{\sum_{i=1}^n (x_i - M)^2}{n}$$

- мера колико су појединачне вредности у просеку далеко од аритметичке средине
- проблем: у којим јединицама мерења?
- много згоднија за математичке операције од интерквартилног распона, и на њој изграђен читав низ статистичких поступака
- $S^2 = 173.85 \text{ cm}^2$



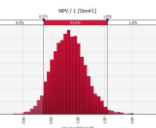
# Мере распршења (варијабилности)

- **Стандардна девијација (SD, S,  $\sigma$ )**

- мера одступања од аритметичке средине
- позитивни корен варијансе

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - M)^2}{n}}$$

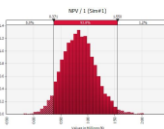
- у истој је јединици мерења као аритметичка средина, па је једноставнија за интерпретацију
- код (приближно) нормалне расподеле се најчешће узима као мера распршења заједно са аритметичком средином као мером ЦТ
- **SD = 13.9 cm**



# Релативне мере варијабилности

- мере колика је варијабилност у односу на ЦТ
- намењене поређењу различитих узорака и варијабли јер их свводе на исту меру
- **коефицијент варијације**
  - однос стандардне девијације и аритметичке средине,  $CV = SD / M$
  - користан када се ове мере користе за опис варијабле
  - вредности 0-1 или 0-100% ако се помножи са 100
- **коефицијент интерквартилне варијације**
  - однос разлике и збира трећег и првог квартила
  - користан када се варијабла описује преко медијане и интерквартилног распона
  - вредности 0-1 или 0-100% ако се помножи са 100

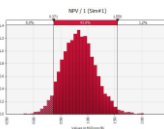
$$V_Q = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$



# Мере облика унимодалне расподеле

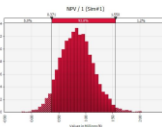
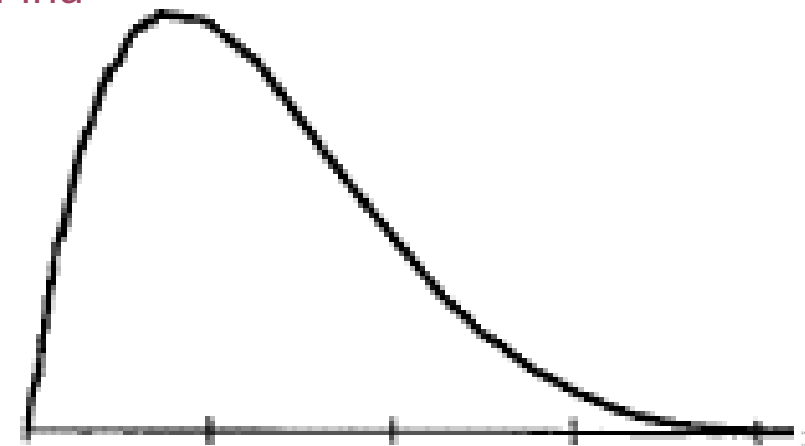
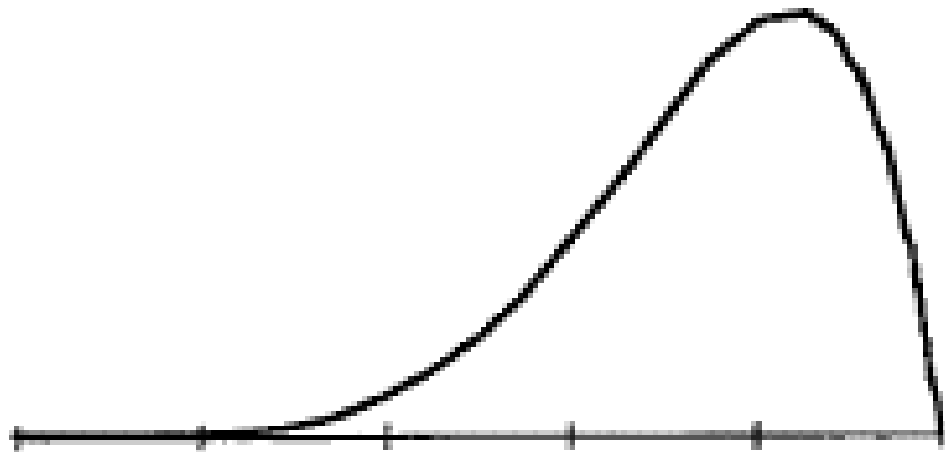
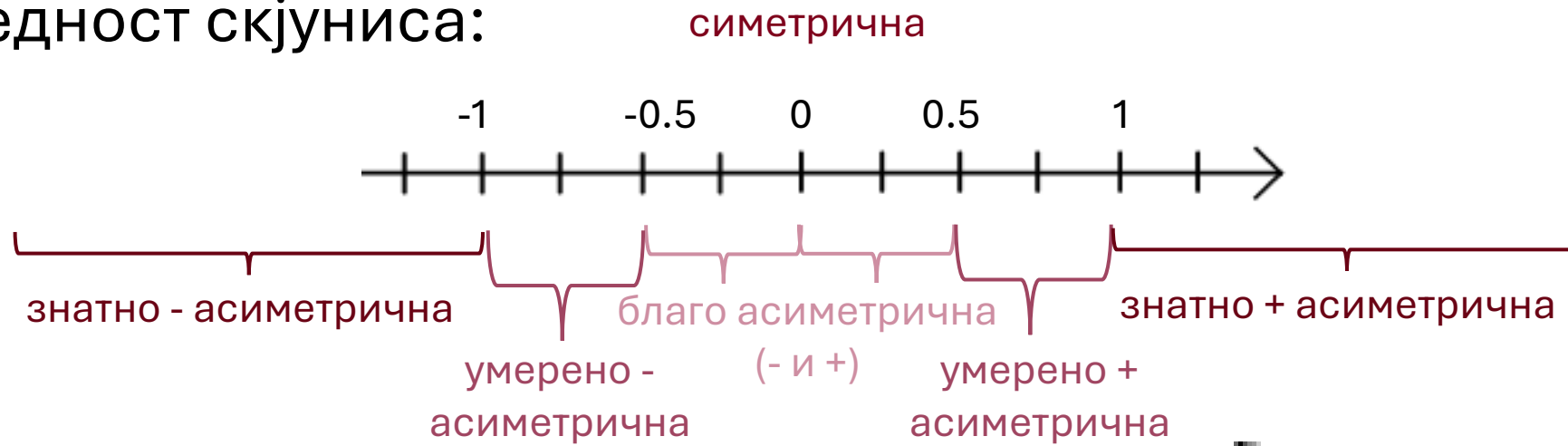
- **Скјунис (Sk)** – мера асиметричности расподеле

$$S_k = \frac{\sum_{i=1}^n (x_i - M)^3}{S * S^2} * \frac{n}{(n-1)(n-2)}$$



# Мере облика унимодалне расподеле

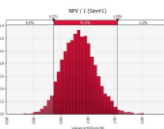
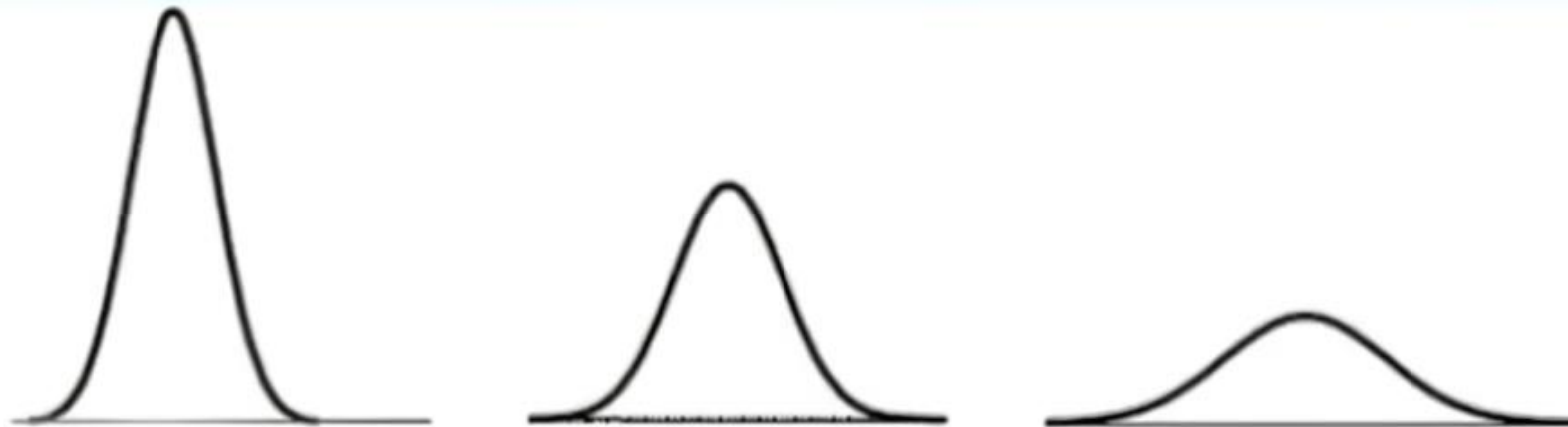
- Вредност скјуниса:



# Мере облика унимодалне расподеле

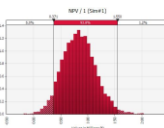
- **Куртозис (Ku)** – мера спљоштености расподеле

$$Ku = \frac{\sum_{i=1}^n (x_i - M)^4}{S^2 * S^2} * \frac{n(n+1)}{(n-1)(n-2)(n-3)} - \frac{3(n-1)^2}{(n-2)(n-3)}$$



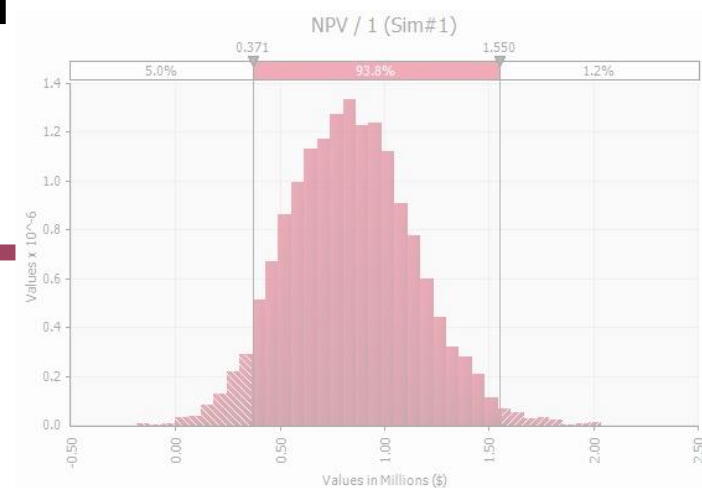
# Мере облика унимодалне расподеле

- Вредност куртозиса, само за унимодалне и релативно симетричне расподеле:
  - $0$  – мезокуртична расподела
  - $> 0$  – платокуртична расподела (веће гомилање на крајевима)
  - $< 0$  – лептокуртична расподела (веће гомилање око централне тачке)
- Нормална расподела:  $S_k = K_u = 0$

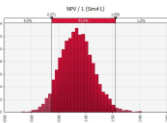
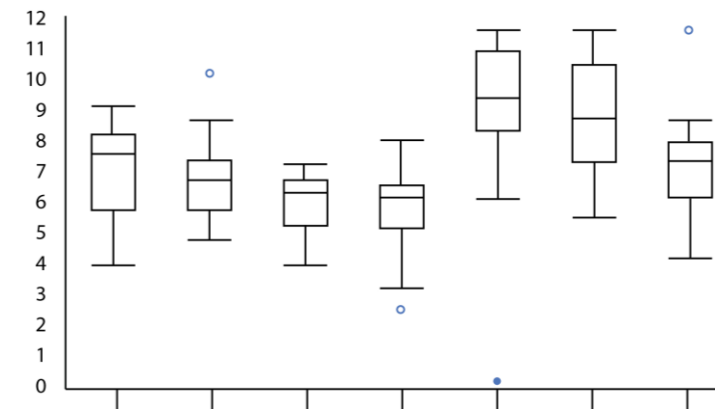
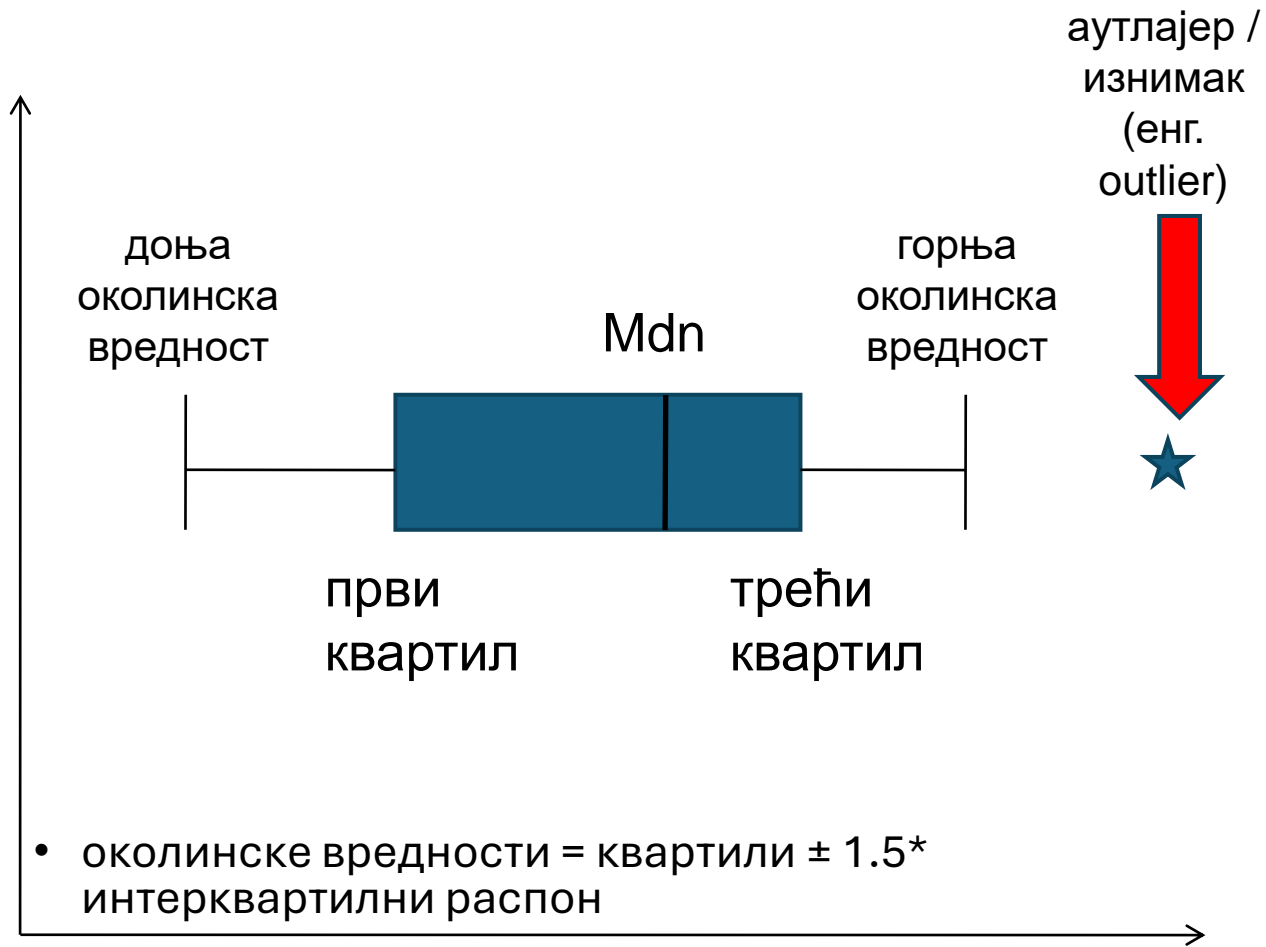




# 6. Графички приказ квантитативних резултата са мерама ЦТ и варијабилности и упутства за графике



# Кутијаста дијаграм 😊

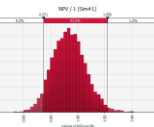
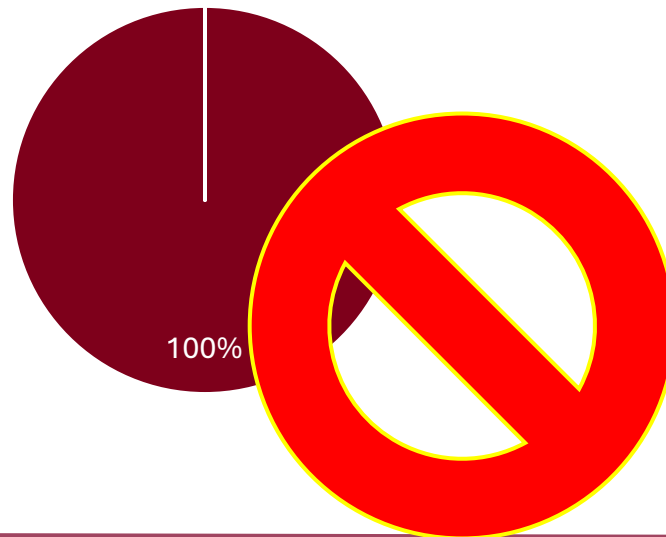


# Савети за осмишљавање графика

- Ствари се често много лакше разумеју када се и виде, али графике прескочите када су ствари сасвим јасне и без њих

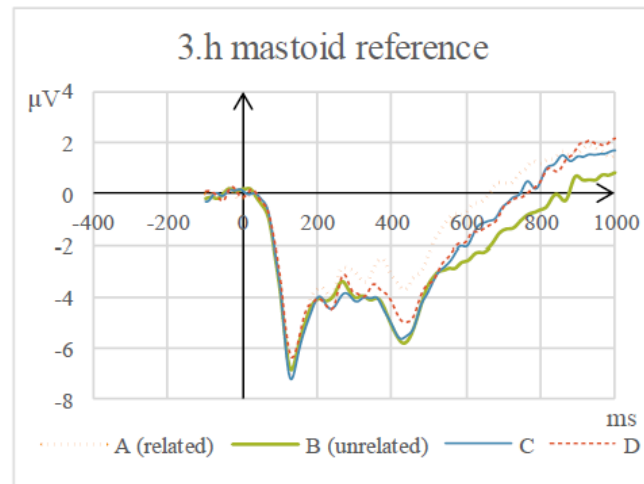
Типови компликација

■ Тип А (n = 1)

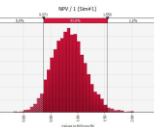


# Савети за осмишљавање графика

- Да би график био користан, потребно је да буде јасно означен (осе, јединице мерења, подеоци на осама)
- График треба да прате легенда и појашњење испод тако да може да се „чита“ сам без консултовања текста

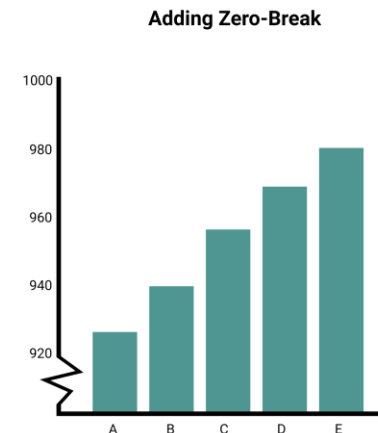
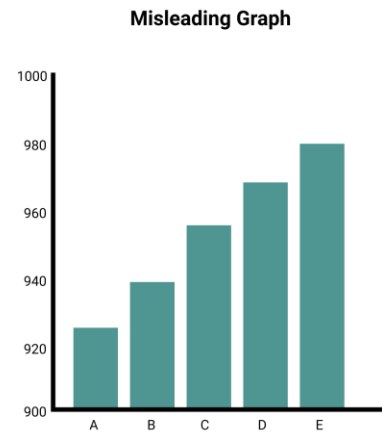
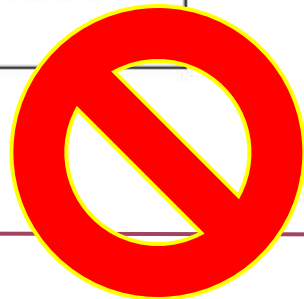
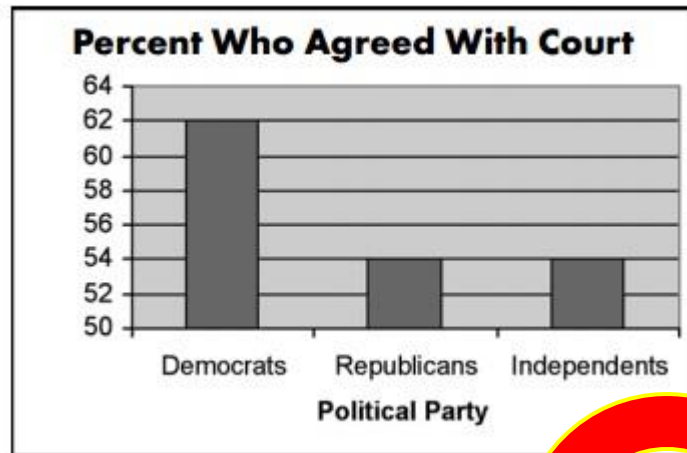


Слика 3.h ...

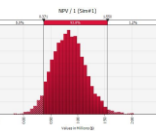
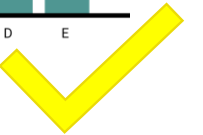
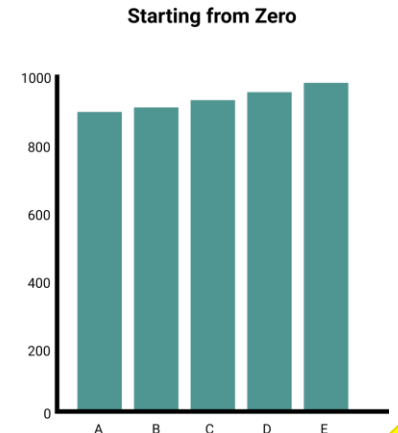


# Савети за осмишљавање графика

- Пажљиво изаберите од које вредности почињу осе и који су подеоци, и јасно прикажите
- Означите прекиде скала ако их има

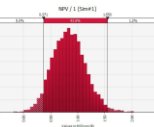
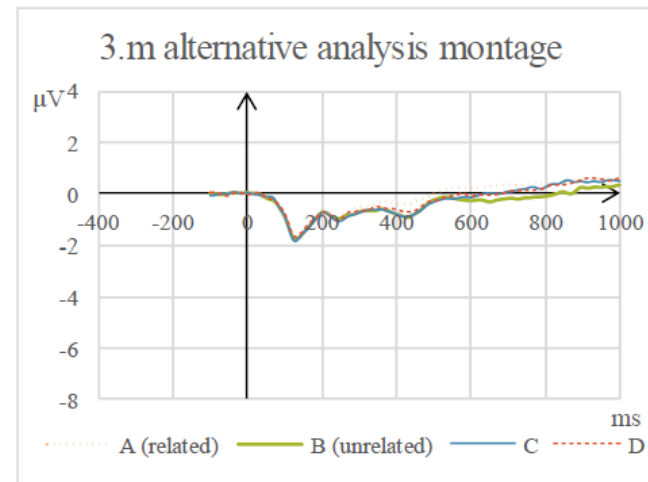
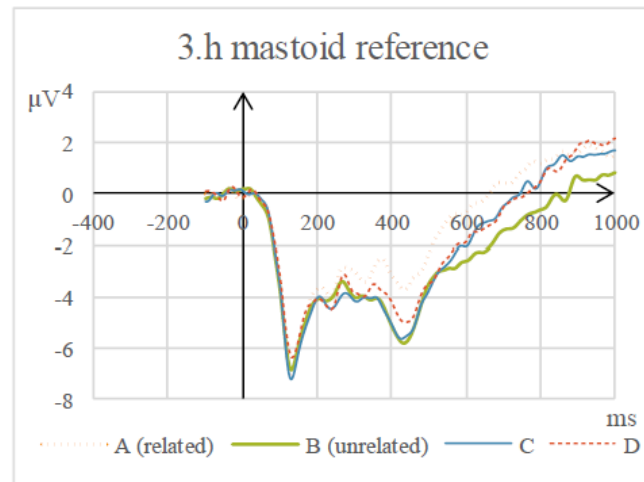


or



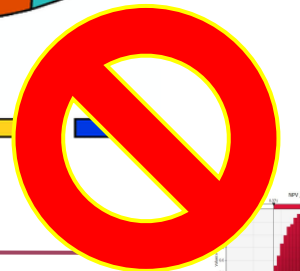
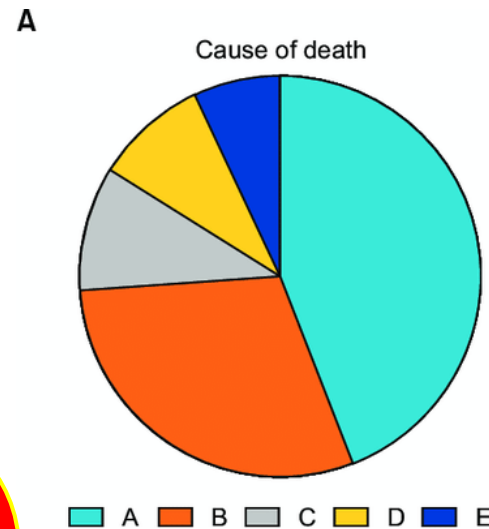
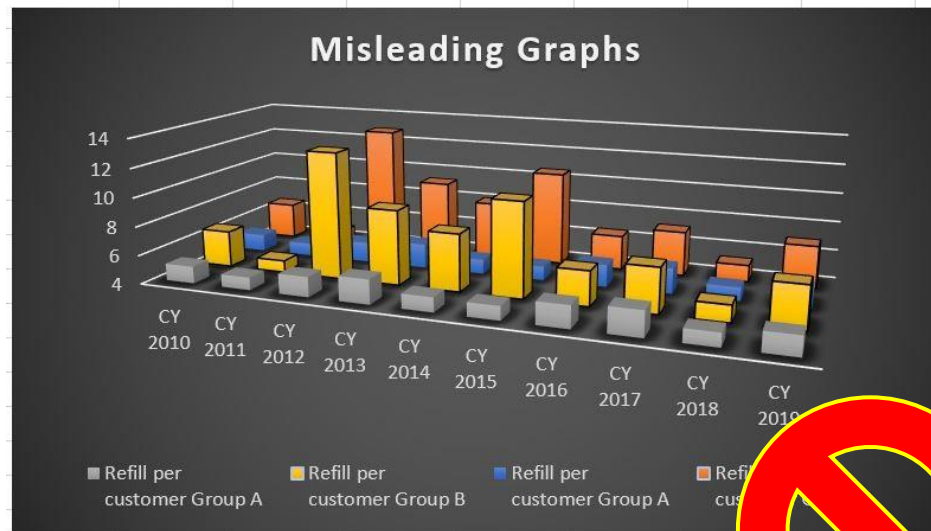
# Савети за осмишљавање графика

- Повезани графици треба да буду приказани на истој скали мерења



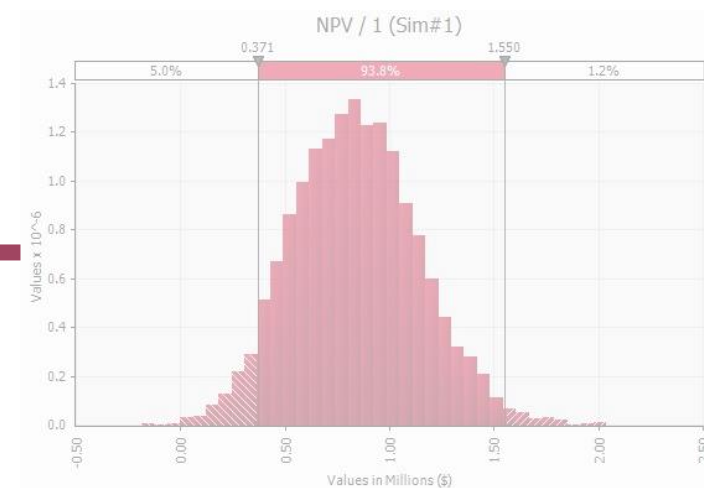
# Савети за осмишљавање графика

- Не стављајте „украсе“ који немају информативну вредност, а отежавају да се уочи суштина (нпр. непотребна трећа димензија или боје, сенке)



# 7. Да сумирамо...

---





# Да сумирамо...

Статистички опис узорка у погледу једне квантитативне варијабле треба да садржи:

- *Прегледне и јасне информације о:*
  - Вредности унутар којих се резултати налазе
    - Одређење најмањег и највећег резултата
  - Учестаности јављања сваког резултата или групе резултата
    - Утврђивање јединичне или груписане расподеле учестаности резултата
  - Вредностима око којих се групишу резултати
    - Утврђивање мера централне тенденције и евентуално других мера локације
  - Степену у коме се резултати међусобно разликују
    - Утврђивање мера варијабилности
  - Облику расподеле
    - Утврђивање мера облика расподеле
- Јасноћи и прегледности често (али не увек) веома помаже графички приказ

