

**OSNOVE
ISTRAŽIVAČKOG RADA**

**STATISTIKA U
ISTRAŽIVANJU**

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

MALO UVODA

- npr.
- itd.
- i sl.
- tzv.
- gđa
- kn
- kg

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

MALO UVODA....

$$\sum_{i=1}^3 x_i$$

$$5!$$

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

ZNAK SUME

$$x_1 + x_2 + \dots + x_{n-1} + x_n = \sum_{i=1}^n x_i$$

$$\sum_{n=1}^3 n = 1 + 2 + 3$$

$$\sum_{i=1}^3 x_i = x_1 + x_2 + x_3$$

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

ZNAK SUME

$$\sum_{i=1}^2 x_i y_i = x_1 y_1 + x_2 y_2$$

$$\sum_{i=1}^2 x_i \sum_{i=1}^2 y_i = (x_1 + x_2)(y_1 + y_2) =$$

$$= x_1 y_1 + x_1 y_2 + x_2 y_1 + x_2 y_2$$

$$\sum_{i=1}^n x_i y_i \neq \sum_{i=1}^n x_i \sum_{i=1}^n y_i$$

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

ZNAK SUME

$$\sum_{i=1}^2 x_i^2 = x_1^2 + x_2^2$$

$$\left(\sum_{i=1}^2 x_i \right)^2 = (x_1 + x_2)^2 = x_1^2 + 2x_1 x_2 + x_2^2$$

$$\sum_{i=1}^n x_i^2 \neq \left(\sum_{i=1}^n x_i \right)^2$$

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

ZNAK SUME

$$\sum_{i=1}^3 2x_i = 2x_1 + 2x_2 + 2x_3 =$$

$$= 2(x_1 + x_2 + x_3) =$$

$$= 2 \sum_{i=1}^3 x_i$$

$$\sum_{i=1}^n cx_i = c \sum_{i=1}^n x_i$$

FAKTORIJELE

$$n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 3 \cdot 2 \cdot 1 \quad ; \quad 0! = 1$$

$$4! = 4 \cdot 3 \cdot 2 \cdot 1$$

$$7! = 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$$

$$(n+2)! = (n+2) \cdot (n+1) \cdot n!$$

$$\frac{(n+2)!}{n!} = \frac{(n+2)(n+1)n!}{n!} = (n+2)(n+1)$$

SVRHA ISTRAŽIVANJA

- opis
 - stanje u populaciji
- usporedba
 - novi postupak VS stari postupak
- povezanost
 - rizični čimbenik i bolest/stanje

STATISTIKA?



....neki razlozi - praćenje literature....

“...povećan je rizik koronarnog incidenta kod bolesnika čija je koncentracija CRP u petoj kvantili u odnosu na prve četiri kvantile...”

“...srednja vrijednost dobi 42±8 godina...”

“...značajnost razlike testirana je Studentovim T-testom...”

“...p < 0.05...”

Multiple R	.96764
R Square	.93632
Adjusted R Square	.92883
Standard Error	6.54079

....neki razlozi - deskripcija i analiza rezultata....

Rezultati mjerenja visine studenata prve godine:

188	175	179	179	173	193	183	177	165	170
165	164	168	182	193	183	160	183	165	166
168	172	178	193	167	174	176	176	172	181
169	172	184	190	182	176	176	164	162	167
189	187	168	175	173	182	175	165	167	181

Tko je najviši ? Tko je najniži ?

Koju visinu ima najveći broj studenata?

?????

Kako varira visina studenata?

?

Koja je visina kojoj teži najveći broj rezultata?

?

.neki razlozi - zaključivanje iz pojedinačnog na "opće"..

Kada i pod kojim uvjetima možemo zaključivati o populaciji iz rezultata mjerenja provedenih na nekoj skupini ispitanika?

Kako odabrati ispitanike?

Koliko ispitanika uključiti u promatranje?

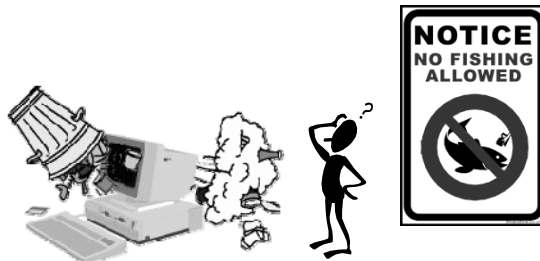
Kolika je pogreška pri zaključivanju?

...



...neki razlozi - planiranje istraživanja i eksperimenta...

- metoda "što ispadne" u istraživanju i eksperimentu može rezultirati nepouzdanim i neinterpretabilnim rezultatima



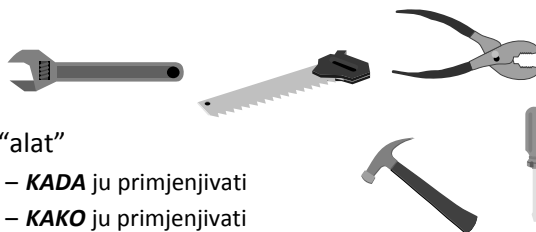
"Savjetovati se sa statističarem **nakon** što je pokus završen često je samo zahtjev da izvrši post-mortem pregled.

On možda može reći od čega je pokus preminuo."

R.A.Fisher, 1938.

Pojam i predmet statistike

- metoda (grupa metoda) razvijena za rad s numeričkim podacima



- "alat"
 - **KADA** ju primjenjivati
 - **KAKO** ju primjenjivati

KADA ?



KAKO ?



...pojam i predmet statistike...

predmet statistike

- metodologija kojom se istražuju **masovne pojave** s ciljem otkrivanja zakonitosti koje u njima vladaju
- **masovne pojave se istražuju na skupu pojedinačnih slučajeva - populaciji**

(osnovni skup, statistička masa)

skup zdravstvenih radnika, skup studenata III godine,

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 19

...pojam i predmet statistike...

(statističke jedinice, entiteti)

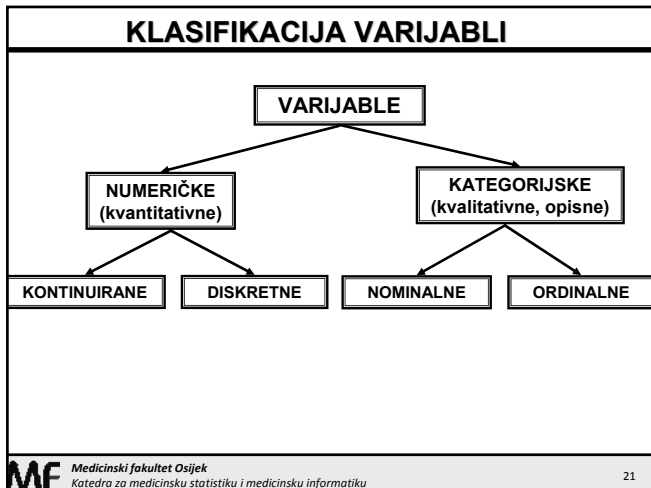
- **populaciju čine osnovne jedinice promatranja**
- **osnovne jedinice promatranja u jednoj populaciji međusobno se razlikuju prema pojedinačnim svojstvima**

zdravstveni radnik u populaciji zdravstvenih radnika

zdravstveni radnici razlikuju se prema stručnoj spremi, specijalizaciji, dobi,

varijabla (obilježje, atribut, karakteristika)

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 20



Numeričke varijable

- kontinuirane varijable
– varijable koje teorijski pretpostavljaju postojanje beskonačnog broja vrijednosti

visina, težina, ...

- diskretne varijable
– varijable čija se vrijednost uvećava ili umanjuje za cijeli broj jedinica

broj pobačaja, broj nesreća u tvornici, ...

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 22

Kategoričke varijable

- nominalne varijable
– varijable čije su vrijednosti određene pripadnošću nekoj kategoriji

bračno stanje, krvna grupa, ...

- dihotomna (binarna) varijabla - samo dvije kategorije

spol (muški/ženski), ishod bolesti (preživljenje/smrtnost) ...

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 23

...kategoričke varijable...

- ordinalne varijable
– varijable čije su vrijednosti određene pripadnošću nekoj od kategorija s poznatim redoslijedom

stupanj opekotine, ocjena na ispitu ...

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 24

Još neke vrste podataka...

- **postotci ili proporcije** (*percentages or proportions*)
 - odnos dijela prema cjelini
 - omjer dva istovrsna podatka

ejekcijska frakcija, ...

- **omjeri** (*ratios*)
 - omjer dva raznovrsna podatka (može biti i omjer dviju varijabli)

ITM

...još neke vrste podataka...

- **stope (rates)**
 - odnos opaženog broja pojave nekog svojstva prema jedinici populacije nekog područja u nekom vremenskom periodu

morbiditet, mortalitet,...

- za izračunavanje stope trebamo:
 - broj pojave svojstva (npr. broj oboljelih)
 - skup u kojem se to obilježje pojavljuje (npr. broj stanovnika nekog područja)
 - **specifikaciju vremena i prostora** (npr. Slavonija, 1992.)

...još neke vrste podataka...

- **skorovi** (*scores*)
 - rezultat zbrajanja vrijednosti dodijeljenih kategorijama varijabli od interesa

Apgar, GCS, ...

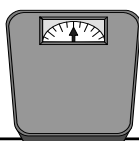
Kriterij	0 bodova	1 bod	2 boda
Boja kože	modra ili blijeda	trup ružičast, okrajine modre	ružičast
Frekvencija srca	nema	<100	>100
Refleksna podražljivost	bez odgovora	grimasa, slab pokret	plač pokret
Mišićni tonus	mlohav	slaba fleksija	jaka fleksija
Disanje	ne diše	grčevito, slab plač	jak, pravilan plač

...još neke vrste podataka...

- **cenzorirani** (*censored*)
 - podatci koji se ne mogu točno izmjeriti, ali je poznato da prelaze neku granicu mjerenja
- najčešće se pojavljuju:
 - u laboratorijskim mjerenjima – vrijednosti ispod/iznad mogućnosti detekcije nekog uređaja
 - u studijama praćenja (follow-up):
 - očekivano svojstvo se nije pojavilo u promatranom periodu
 - ispitanik uključen u istraživanje je iz nekog razloga izuzet prije završetka istraživanja

LJESTVICE MJERENJA

1. nominalna
2. ordinalna
3. intervalna
4. omjerna



1. Nominalna ljestvica

- kategorizacija kojom objektima ili događajima pridružujemo riječi ili simbole
- nema informacije o veličini pojedinačnog rezultata

spol (muški/ženski)

ishod bolesti (preživljavanje/smrt)

boja očiju (plava, crna, smeđa, ...)

2. Ordinalna ljestvica

2 1 3

- ima sve karakteristike nominalne ljestvice i dodatno uključuje **redosljed** (nominalna + rangiranje)
- intervali nisu jednaki, a granice među grupama nisu čvrste
- nema informacije o "jačini" razlike između pojedinih grupa

stupanj opekotine (prvi, drugi, treći)

ocjene na ispitu

31

3. Intervalna ljestvica



- ljestvica sa jednakim intervalima koji imaju definirane granice => **razlike imaju smisla**
- nema apsolutnu nulu (moguće su i negativne vrijednosti) => **omjeri nemaju smisla**

temperatura, kvocijent inteligencije

100°F ≠ 2*50°F jer je 100°F=38°C

50°F=10°C

$$\left(\frac{t^{\circ}\text{C}}{100} = \frac{t^{\circ}\text{F} - 32}{180} \right)$$

Katedra za medicinsku statistiku i medicinsku informatiku

32

4. Omjerna ljestvica



- ima sva svojstva intervalne ljestvice ali i apsolutnu nulu => **omjeri imaju smisla**
- nula znači totalnu odsutnost obilježja

visina, težina, dob, ...

90 kg = 3*30 kg

3174.653 oz = 3*1058.218 oz (1 kg=35.27392 oz)

Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

33

Rezultati mjerenja na ljestvicama višeg reda mogu se prikazati ljestvicama nižeg reda, ali NE I OBRNUTO !

Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

34

NEKOLIKO VAŽNIH PRAVILA ...

- numeričke podatke **UVIJEK** mjeriti ljestvicom najvišeg mogućeg reda (točna vrijednost tlaka, točna vrijednost GUK, ...)
- unaprijed odabrati smislen nivo točnosti (težina u gramima za novorođenčad, ali NE i za odrasle osobe!)
- definirati kategorije za **SVE MOGUĆE** vrijednosti kategoričke varijable

Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

35

Unos podataka

VARIJABLE

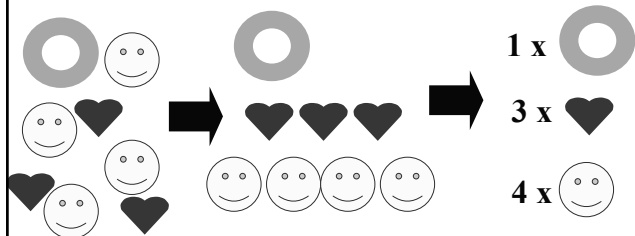
↓ ↓ ↓

	spol	dob	težina
JEDINICE →	bolesnik1			
PROMATRANJA →	bolesnik2			
	bolesnik3			
	.			

Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

36

RAZDIOBA OBILJEŽJA



Primjer

Na jednom čovjeku izvršeno je 50 mjerenja vremena reakcije. Dobiveni su sljedeći podaci (u tisućinkama sekunde):

196	173	186	189	173	165	167	160	140	174
180	151	157	164	154	169	190	180	163	157
169	167	165	160	177	165	157	177	159	175
166	173	185	177	184	183	162	192	174	162
165	172	158	169	146	170	171	169	168	153



TABLICA FREKVENCIJA

... tablica frekvencija ...

- tablica u kojoj su originalni podaci sažeti u određeni broj kategorija (*razreda*) koje su opisane numerički izraženim granicama

raspon (interval) razreda

- razlika granica razreda

... tablica frekvencija ...

sredina razreda

- broj koji najbolje reprezentira dani razred

računanje sredine razreda:

- *diskretne varijable:*

suma granica razreda / 2

- *kontinuirane varijable:*

suma donjih granica razreda / 2

apsolutna frekvencija razreda(f)

- broj podataka koji pripadaju intervalu tog razreda

kumulativna frekvencija razreda(cf)

- broj podataka čija je vrijednost manja ili jednaka gornjoj granici razreda

relativna frekvencija razreda(rf)

- apsolutna frekvencija razreda podijeljena s ukupnim brojem podataka

kumulativna relativna frekvencija razreda (crf)

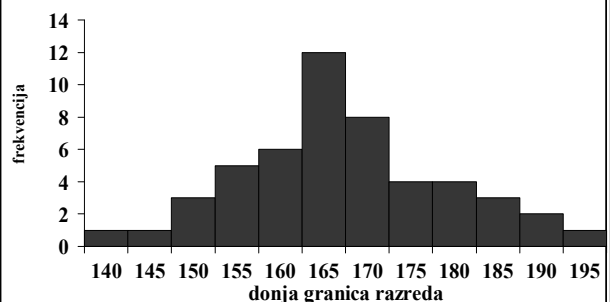
- kumulativna frekvencija razreda podijeljena s ukupnim brojem podataka

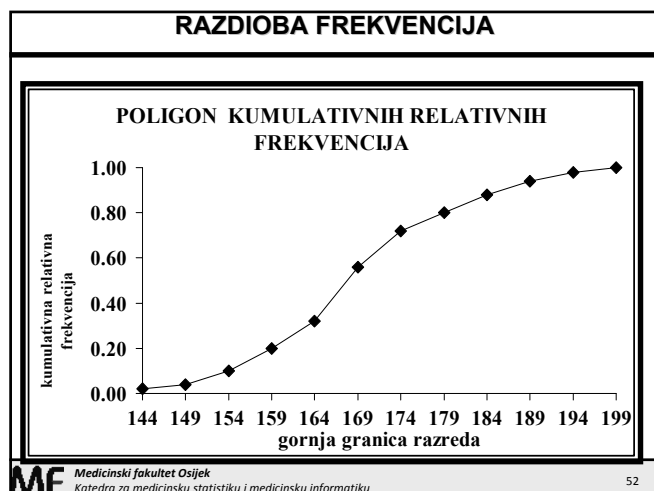
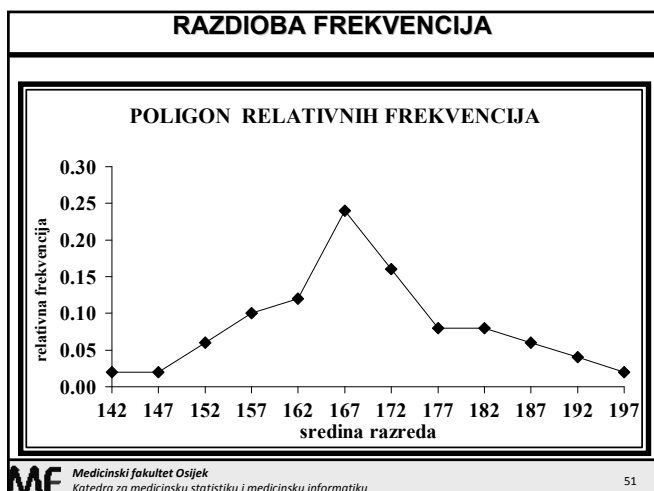
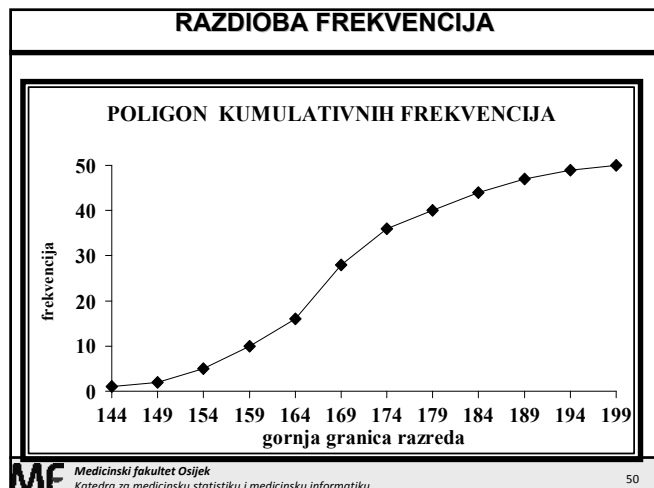
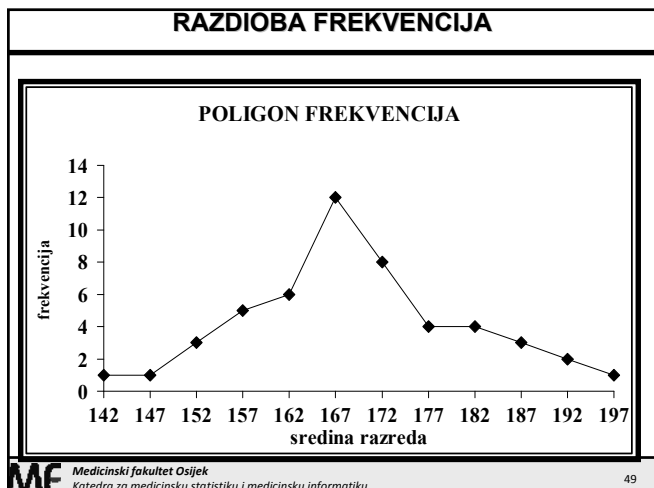
za dani razred:

- apsolutna frekvencija:
 - koliko mjerenja ima vrijednosti iz intervala tog razreda
- apsolutna kumulativna frekvencija:
 - koliko mjerenja ima vrijednost manju ili jednaku gornjoj granici tog razreda
- relativna frekvencija:
 - koliki postotak mjerenja od ukupnog broja mjerenja ima vrijednost iz intervala tog razreda
- kumulativna relativna frekvencija:
 - koliki postotak mjerenja od ukupnog broja mjerenja ima vrijednost manju ili jednaku gornjoj granici tog razreda

Tablica frekvencija podataka iz primjera

Broj r.	Granice r.	Sredina r.	f	cf	rf	crf
1	140 – 144	142	1	1	0,02	0,02
2	145 – 149	147	1	2	0,02	0,04
3	150 – 154	152	3	5	0,06	0,10
4	155 – 159	157	5	10	0,10	0,20
5	160 – 164	162	6	16	0,12	0,32
6	165 – 169	167	12	28	0,24	0,56
7	170 – 174	172	8	36	0,16	0,72
8	175 – 179	177	4	40	0,08	0,80
9	180 – 184	182	4	44	0,08	0,88
10	185 – 189	187	3	47	0,06	0,94
11	190 – 194	192	2	49	0,04	0,98
12	195 – 199	197	1	50	0,02	1,00
UKUPNO :			50		1,00	

RAZDIoba FREKVENCIJA**HISTOGRAM FREKVENCIJA**



Stablo i list (stem-and-leaf)

f	stablo	list
1.00	14	0
1.00	14	6
3.00	15	134
5.00	15	77789
6.00	16	002234
12.00	16	555567789999
8.00	17	01233344
4.00	17	5777
4.00	18	0034
3.00	18	569
2.00	19	02
1.00	19	6

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 53

- PAŽNJA !**
- broj razreda
 - preveliki broj razreda => male frekvencije ili prazni razredi
 - premali broj razreda => razredi jako sažeti => izgubljeno puno informacija
 - uobičajeno: 10-20 razreda (ovisno o broju i prirodi podataka)
 - kod nominalnih varijabli:
 - broj kategorija=broj razreda
- MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 54

PAŽNJA !

– granice razreda

- najmanje na onoj točnosti na kojoj je izvršeno mjerenje
- određene tako da SVAKI PODATAK PADNE U SAMO JEDAN OD RAZREDA!

OPISIVANJE RAZDIOBE PODATAKA

PARAMETAR I STATISTIKA

POPULACIJA



1.UZORAK



2.UZORAK

⋮



n-ti UZORAK

PARAMETAR I STATISTIKA



aritmetička sredina
visine populacije
= 175.4



aritmetička sredina
visine 1. uzorka
= 172.2



aritmetička sredina
visine 2. uzorka
= 178.1



aritmetička sredina
visine n-tog uzorka
= 173.7

PARAMETAR I STATISTIKA

- parametar:
 - vrijednost (obično nepoznata) koja predstavlja neku karakteristiku populacije
 - unutar populacije, parametar je nepromjenljiva vrijednost koja NE VARIRA
- statistika:
 - veličina izračunata iz podataka izmjerenih na uzorku
 - vrijednost statistike MIJENJA SE od uzorka do uzorka

Uobičajene oznake:

	OCJENA PARAMETRA (STATISTIKA)	PARAMETAR POPULACIJE
ARITMETIČKA SREDINA	\bar{X}	μ
STANDARDNA DEVIJACIJA	S	σ
PROPORCIJA	ρ	π

OPISIVANJE RAZDIOBE PODATAKA



- sredina
- varijabilnost
- oblik



MJERE SREDINE (centralne tendencije)

(srednje vrijednosti, prosjeci, mjere lokacije)

- karakteristične vrijednosti oko kojih se grupiraju mjereni podaci

ARITMETIČKA SREDINA

MEDIJAN (središnja vrijednost)

MOD (dominantna ili tipična vrijednost)

GEOMETRIJSKA SREDINA

HARMONIJSKA SREDINA

ARITMETIČKA SREDINA

- oznake: \bar{X} uzorak μ populacija

ARITMETIČKA SREDINA INDIVIDUALNIH PODATAKA

$$\mu = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{\sum_{i=1}^N x_i}{N}$$

x_i ... vrijednosti mjenog obilježja
N ... ukupan broj podataka

Primjer

Kolika je aritmetička sredina niza podataka:

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

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6}{6} = \frac{1 + 2 + 3 + 3 + 4 + 5}{6} = \frac{18}{6} = 3$$

Kolika je aritmetička sredina niza podataka:

1, 1, 1, 1, 2, 12?

$$\bar{x} = \frac{1 + 1 + 1 + 1 + 2 + 12}{6} = \frac{18}{6} = 3$$

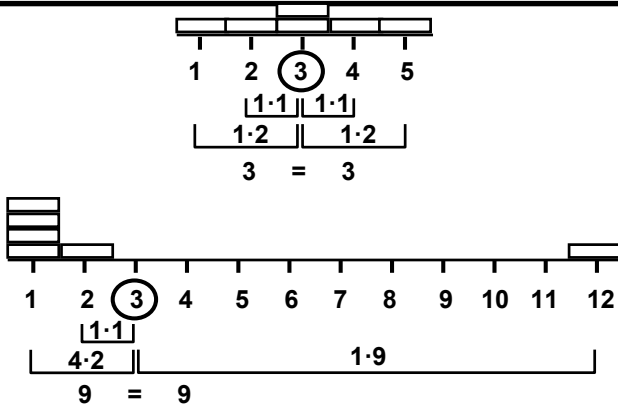
Imamo dva niza podataka:

A: 1 2 3 3 4 5
B: 1 1 1 1 2 12

aritmetička sredina niza A = 3
aritmetička sredina niza B = 3

- loše opisuje niz B
- veliki utjecaj ekstremne vrijednosti (12)

• aritmetička sredina je težište distribucije



ARITMETIČKA SREDINA GRUPIRANIH PODATAKA

$$\mu = \frac{f_1 x_{S1} + f_2 x_{S2} + \dots + f_k x_{Sk}}{N} = \frac{\sum_{i=1}^k f_i x_{Si}}{N}$$

f_i ... frekvencija i-tog razreda
 x_{Si} ... sredina i-tog razreda
 k ... broj razreda
 N ... ukupan broj podataka

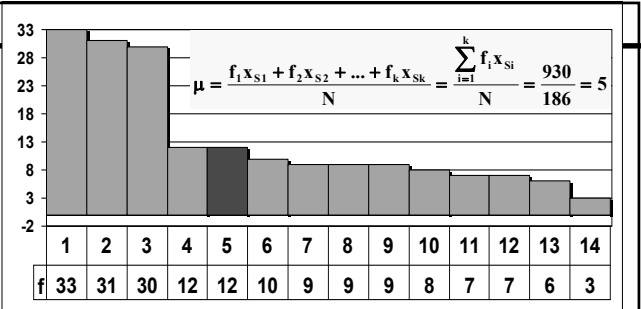
Kolika je aritmetička sredina podataka u sljedećoj tablici:

x	1	2	3	4	5	6	7	8	9	10	11	12	13	14
f	33	31	30	12	12	10	9	9	9	8	7	7	6	3
f·x	33	62	90	48	60	60	63	72	81	80	77	84	78	42

$$\sum_{i=1}^{14} f_i = 186$$

$$\sum_{i=1}^{14} f_i \cdot x_i = 930$$

$$\bar{x} = \frac{930}{186} = 5$$



➤ razdioba frekvencija nije simetrična
 => primjer lošeg opisivanja rezultata aritmetičkom sredinom

ZAJEDNIČKA ARITMETIČKA SREDINA
 (aritmetička sredina aritmetičkih sredina,
 ponderirana aritmetička sredina)

$$\bar{x}_{zaj} = \frac{\bar{x}_1 N_1 + \bar{x}_2 N_2 + \dots + \bar{x}_n N_n}{\sum_{i=1}^n N_i} = \frac{\sum_{i=1}^n \bar{x}_i N_i}{\sum_{i=1}^n N_i}$$

x_i ... aritmetička sredina (vrijednost) dobivena iz
 N_i ("ponder", težina) mjerenja
 n ... broj skupina mjerenja (broj mjerenja)

Dva studenta studija sestrinstva postigla su sljedeći uspjeh u V semestru studija:

PREDMET	OCJENA		ECTS
	STUDENT 1	STUDENT 2	
Osnove istraživačkog rada	2	5	4
Zdravstvena njega odraslih II	5	2	9
Zdravstvena njega psihijatrijskih bolesnika	5	3	8
Klinička medicina III	3	5	4

Koji je student postigao bolji uspjeh u V semestru?

PREDMET	OCJENA		ECTS
	STUDENT 1	STUDENT 2	
Osnove istraživačkog rada	2	5	4
Zdravstvena njega odraslih II	5	2	9
Zdravstvena njega psihijatrijskih bolesnika	5	3	8
Klinička medicina III	3	5	4
S	15	15	25

$$\text{uspjeh}_{s_1} = \frac{2 \cdot 4 + 5 \cdot 9 + 5 \cdot 8 + 3 \cdot 4}{25} = \frac{105}{25} = 4.20$$


$$\text{uspjeh}_{s_2} = \frac{5 \cdot 4 + 2 \cdot 9 + 3 \cdot 8 + 5 \cdot 4}{25} = \frac{82}{25} = 3.28$$

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku


U dva navrata vršeno je mjerenje neke dužine i dobiveni su slijedeći rezultati:
 $\bar{x}_1 = 20\text{cm}$; $N_1 = 15$
 $\bar{x}_2 = 23\text{cm}$; $N_2 = 60$

a) Kolika je zajednička aritmetička sredina?
b) Kolika je zajednička aritmetička sredina za $N_1=60; N_2=15$?

a) $\bar{x}_{\text{zaj}} = \frac{20 \cdot 15 + 23 \cdot 60}{15 + 60} = \frac{300 + 1380}{75} = 22.4\text{cm}$
b) $\bar{x}_{\text{zaj}} = \frac{20 \cdot 60 + 23 \cdot 15}{15 + 60} = \frac{1200 + 354}{75} = 20.6\text{cm}$

 aritmetička sredina osjetljiva je **ne samo na vrijednost nego i na broj podataka**

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku



Aritmetička sredina nema smisla, tj. nije dobar reprezentant podataka ako je:

- razdioba asimetrična
- broj podataka mali, a varijabilnost velika (velike razlike u vrijednostima podataka)

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 75

Imamo niz podataka:
A: 2 2.5 3.5 3 4


Kolika je suma odstupanja pojedinačnih vrijednosti od aritmetičke sredine? Kolika je suma kvadrata odstupanja pojedinačnih vrijednosti od aritmetičke sredine te od vrijednosti 2; 4; 5?

$(2-3)+(2.5-3)+(3.5-3)+(3-3)+(4-3) = -1-0.5+0.5+0+1 = 0$

x_i	$(x_i-3)^2$	$(x_i-2)^2$	$(x_i-4)^2$	$(x_i-5)^2$
2	1	0	4	9
2.5	0.25	0.25	2.25	6.25
3	0	1	1	4
3.5	0.25	2.25	0.25	2.25
4	1	4	0	1
Σ	2.5	7.5	7.5	22.5

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

SVOJSTVA ARITMETIČKE SREDINE



- $\sum_{i=1}^N (x_i - \mu) = 0$
- $\sum_{i=1}^N (x_i - \mu)^2 < \sum_{i=1}^N (x_i - a)^2, \forall a \neq \mu$

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 77

MEDIJAN (središnja vrijednost)

- vrijednost koja se u nizu podataka **poredanih po veličini** nalazi **točno u sredini** – središnja vrijednost po položaju
- vrijednost medijana:
 - za **neparan N**: vrijednost koja se nalazi na $(N+1)/2$ mjestu
 - za **paran N**: sredina vrijednosti podataka koji se nalaze na $N/2$ i $(N+2)/2$ mjestu

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 78

MEDIJAN (središnja vrijednost)

- prednosti:
 - na vrijednost medijana *ne utječu ekstremne vrijednosti*
- ⇒ **pogodan** kao mjera centralne tendencije kod **asimetričnih raspodjela**

oznaka: M_e (C , M_d)

PRIMJER

Za nizove podataka iz primjera:

A: 1 2 3 3 4 5

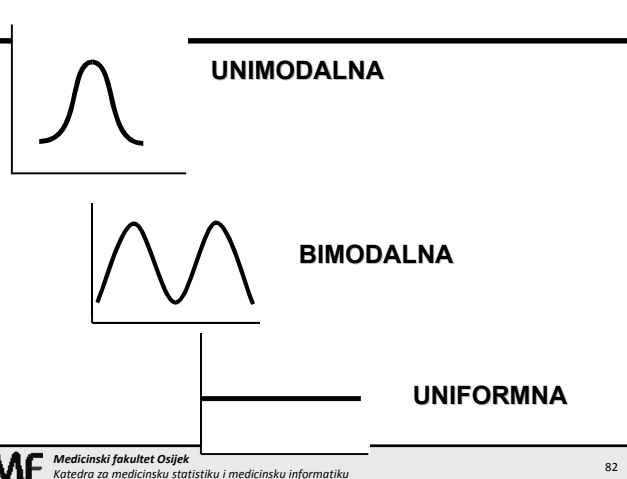
B: 1 1 1 1 2 12

niz A: medijan.... **Me=3** arit. sred. $\bar{X} = 3$

niz B: medijan.... **Me=1** arit. sred. $\bar{X} = 3$

MOD (dominantna vrijednost)

- vrijednost koja se u nizu mjerenja **najčešće** javlja (dominira svojom frekvencijom)
 - na mod ne utječu ni broj ni veličina podataka, već **samo frekvencija**
- oznaka: M_o



MJERE SREDINE PREMA LJESTVICI MJERENJA

LJESTVICA MJERENJA	MJERA SREDINE
NOMINALNA	MOD
ORDINALNA	MOD MEDIJAN
INTERVALNA/OMJERNA (asimetrična raspodjela)	MOD MEDIJAN
INTERVALNA/OMJERNA (simetrična raspodjela)	MOD MEDIJAN ARITMETIČKA SREDINA

MJERE RASPRŠENJA (VARIJABILNOSTI)

(mjere disperzije)

RASPON
KVANTILE
VARIJANCA
STANDARDNA DEVIJACIJA
KOEFICIJENT VARIJABILNOSTI

RASPON

$$R = \max - \min$$

- nedostaci:
 - uzima u obzir samo dvije ekstremne vrijednosti koje uopće ne moraju biti karakteristične za promatranu varijablu
 - ovisi o broju opažanja (veći broj opažanja => veći raspon)

KVANTILE

- mjere varijabilnosti po položaju
- kvartile, decile, centile
- donja kvartila (Q_1 ili 25%)
 - vrijednost podatka koji stoji na centralnom mjestu polovice podataka nižih od medijana
- gornja kvartila (Q_3 ili 75%)
 - vrijednost podatka koji stoji na centralnom mjestu polovice podataka viših od medijana
- Q_2 - medijan

centila	obuhvat jedinica promatranja	decila	obuhvat jedinica promatranja
prva	1%	prva	10%
druga	2%	druga	20%
treća	3%	treća	30%
....		

VARIJANCA

- prosječno kvadratno odstupanje od aritmetičke sredine

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

**VARIJANCA
POPULACIJE**

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

**VARIJANCA
UZORKA**

STANDARDNA DEVIJACIJA

ZA POPULACIJU

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

ZA UZORAK

$$s = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}}$$

- služi za ocjenu pojedinih rezultata oko aritmetičke sredine
- izražava se uz aritmetičku sredinu
- obično je: $2s < \text{raspon} < 6s$

KOEFICIJENT VARIJABILNOSTI

ZA POPULACIJU

$$K.V. = \frac{\sigma}{\mu} \cdot 100$$

ZA UZORAK

$$K.V. = \frac{s}{\bar{x}} \cdot 100$$

- relativna standardna devijacija
- govori o HOMOGENOSTI promatranog obilježja
- koristan je ako želimo znati:
 - a) razlike u varijabilnosti različitih svojstava neke grupe ispitanika
 - b) razlike u varijabilnosti istog svojstva u različitim grupama ispitanika

OPIS VARIJABLE

MJERA SREDINE	MJERA VARIJABILNOSTI
ARITMETIČKA SREDINA	STANDARDNA DEVIJACIJA
MEDIJAN	GRANICE INTERKVARTILNOG RASPONA (25%-75%)

MJERE ZA OCJENU OBLIKA RAZDIOBE

MOMENTI RAZDIOBE

- uzastopne mjere prosječnih odstupanja od aritmetičke sredine nultog, prvog, drugog, trećeg i višeg reda

MOMENT n-tog REDA (n-ti moment)

$$\mu_n = \frac{\sum_{i=1}^N (x_i - \mu)^n}{N}$$

NULTI I PRVI MOMENT

$$\mu_0 = \frac{\sum_{i=1}^N (x_i - \mu)^0}{N} = \frac{\sum_{i=1}^N 1}{N} = \frac{N}{N} = 1$$

prvo svojstvo aritmetičke sredine

$$\mu_1 = \frac{\sum_{i=1}^N (x_i - \mu)^1}{N} = \frac{0}{N} = 0$$

DRUGI MOMENT

$$\mu_2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} = \sigma^2 \quad \text{VARIJANCA}$$

TREĆI MOMENT

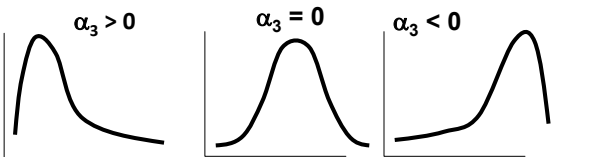
$$\mu_3 = \frac{\sum_{i=1}^N (x_i - \mu)^3}{N}$$

- za simetrične raspodjele $\mu_3 = 0$

KOEFIČIJENT ASIMETRIJE (coefficient of skewness)

$$\alpha_3 = \frac{\mu_3}{\sigma^3}$$

- $\alpha_3 > 0$ asimetrija udesno (pozitivna asimetrija)
- $\alpha_3 < 0$ asimetrija ulijevo (negativna asimetrija)



$\alpha_3 > 0$ $\alpha_3 = 0$ $\alpha_3 < 0$

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 97

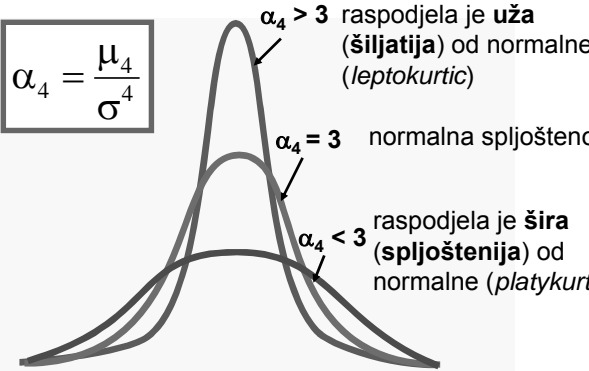
ČETVRTI MOMENT

$$\mu_4 = \frac{\sum_{i=1}^N (x_i - \mu)^4}{N}$$

- koristi se za mjeru spljoštenosti

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 98

KOEFIČIJENT SPLJOŠTENOSTI (coefficient of kurtosis)

$$\alpha_4 = \frac{\mu_4}{\sigma^4}$$


- $\alpha_4 > 3$ raspodjela je uža (šiljatija) od normalne (leptokurtic)
- $\alpha_4 = 3$ normalna spljoštenost
- $\alpha_4 < 3$ raspodjela je šira (spljoštenija) od normalne (platykurtic)

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 99

KOEFIČIJENT SPLJOŠTENOSTI (coefficient of kurtosis)

- statistički programi koeficijent spljoštenosti prikazuju kao **eksces spljoštenosti** (kurtosis excess)

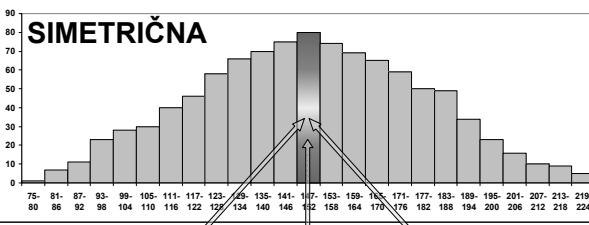
$$\text{eksces spljoštenosti} = \frac{\mu_4}{\sigma^4} - 3$$

- za normalnu raspodjelu eksces spljoštenosti = 0

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 100

ODNOS MJERA SREDINE I ASIMETRIJE

SIMETRIČNA



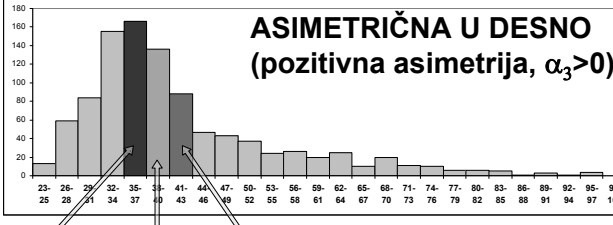
$\bar{x} = 150$ $Me = 150$ $Mo = 150$

$$\bar{x} = Me = Mo$$

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 101

ODNOS MJERA SREDINE I ASIMETRIJE

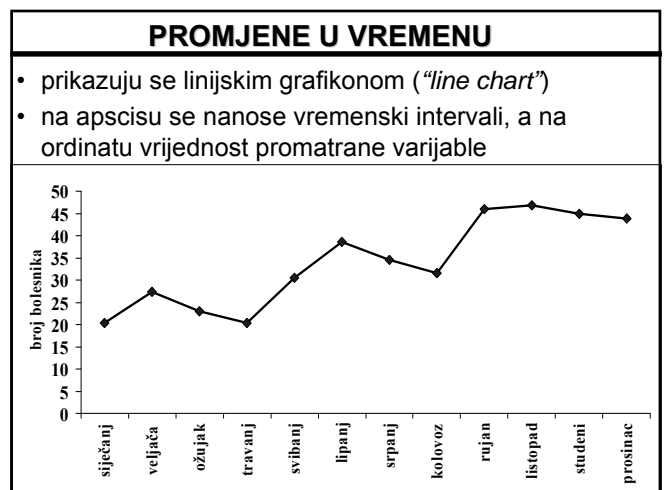
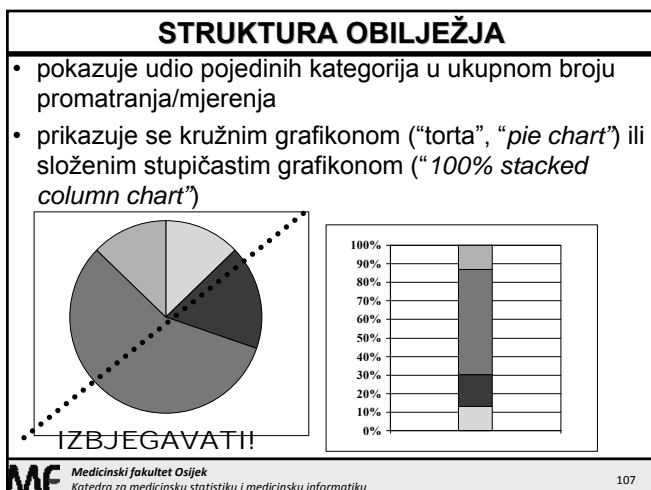
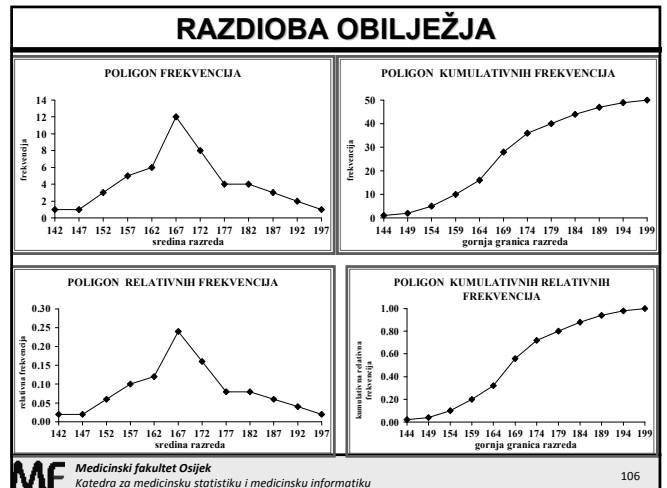
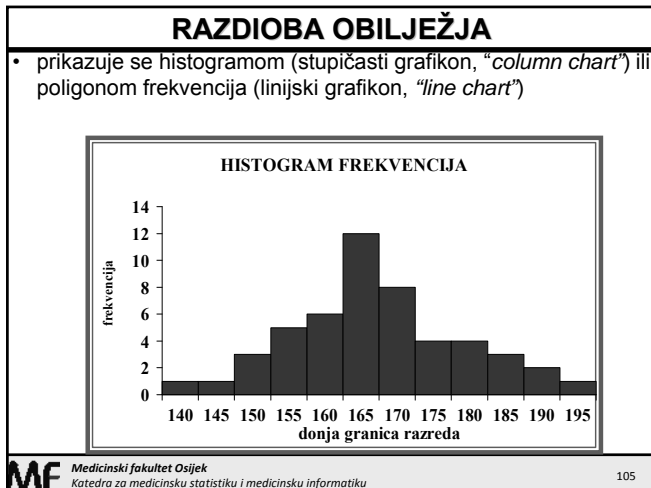
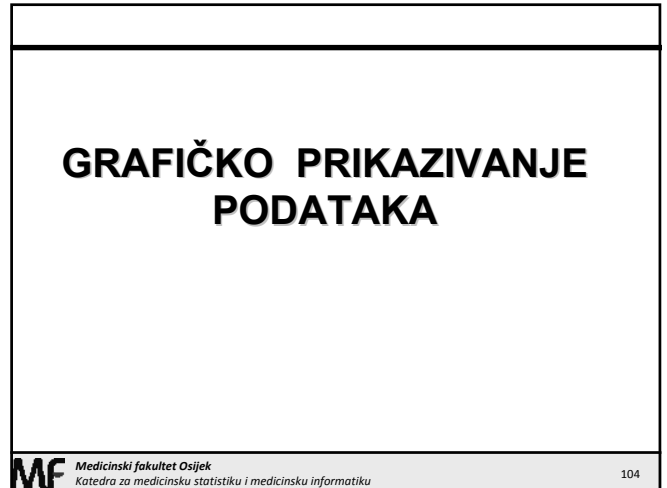
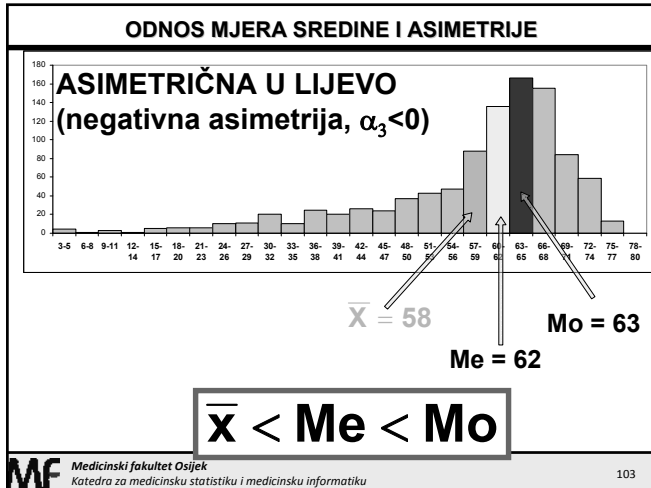
ASIMETRIČNA U DESNO (pozitivna asimetrija, $\alpha_3 > 0$)



$Mo = 37$ $Me = 38$ $\bar{x} = 42$

$$Mo < Me < \bar{x}$$

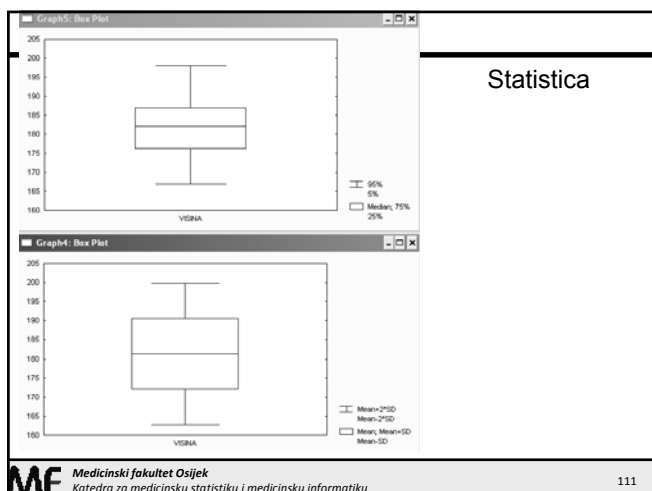
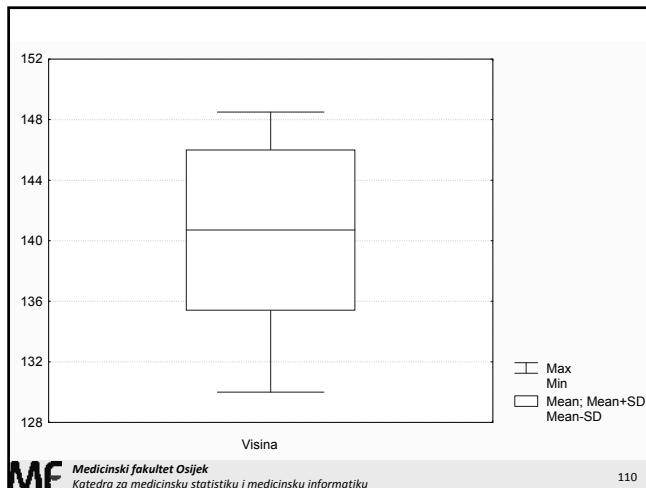
MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 102



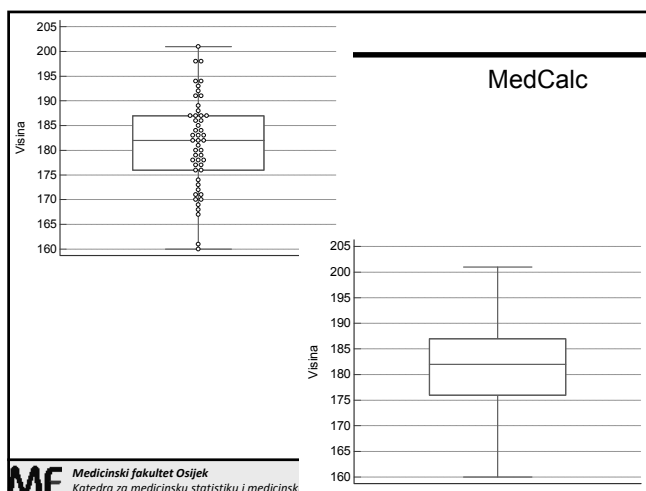
OSNOVNE MJERE SREDINE I RASPRŠENJA VARIJABLE

- “kutija i brkovi” grafikon (“Box-and-Whisker” plot)
- najčešće prikazuje kombinacije:

sredina	kutija	brkovi
aritmetička sredina	standardna devijacija	2 SD ili raspon
medijan	25% - 75%	5% - 95% ili raspon



Statistica



MedCalc

POVEZANOST DVIJU VARIJABLI

- raspršni grafikon (korelacijski, “scatter graph”)
- svaka točka predstavlja par vrijednosti promatranih varijabli

