

# MALO UVODA

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

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

5!

# ZNAK SUME

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

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

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

# ZNAK SUME

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

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

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

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

# ZNAK SUME

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

$$\left( \sum_{i=1}^2 x_i \right)^2 = (x_1 + x_2)^2 = x_1^2 + 2x_1x_2 + x_2^2$$

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

# ZNAK SUME

$$\begin{aligned}\sum_{i=1}^3 2x_i &= 2x_1 + 2x_2 + 2x_3 = \\ &= 2(x_1 + x_2 + x_3) = \\ &= 2\sum_{i=1}^3 x_i\end{aligned}$$

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

# FAKTORIJEI

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

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

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

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

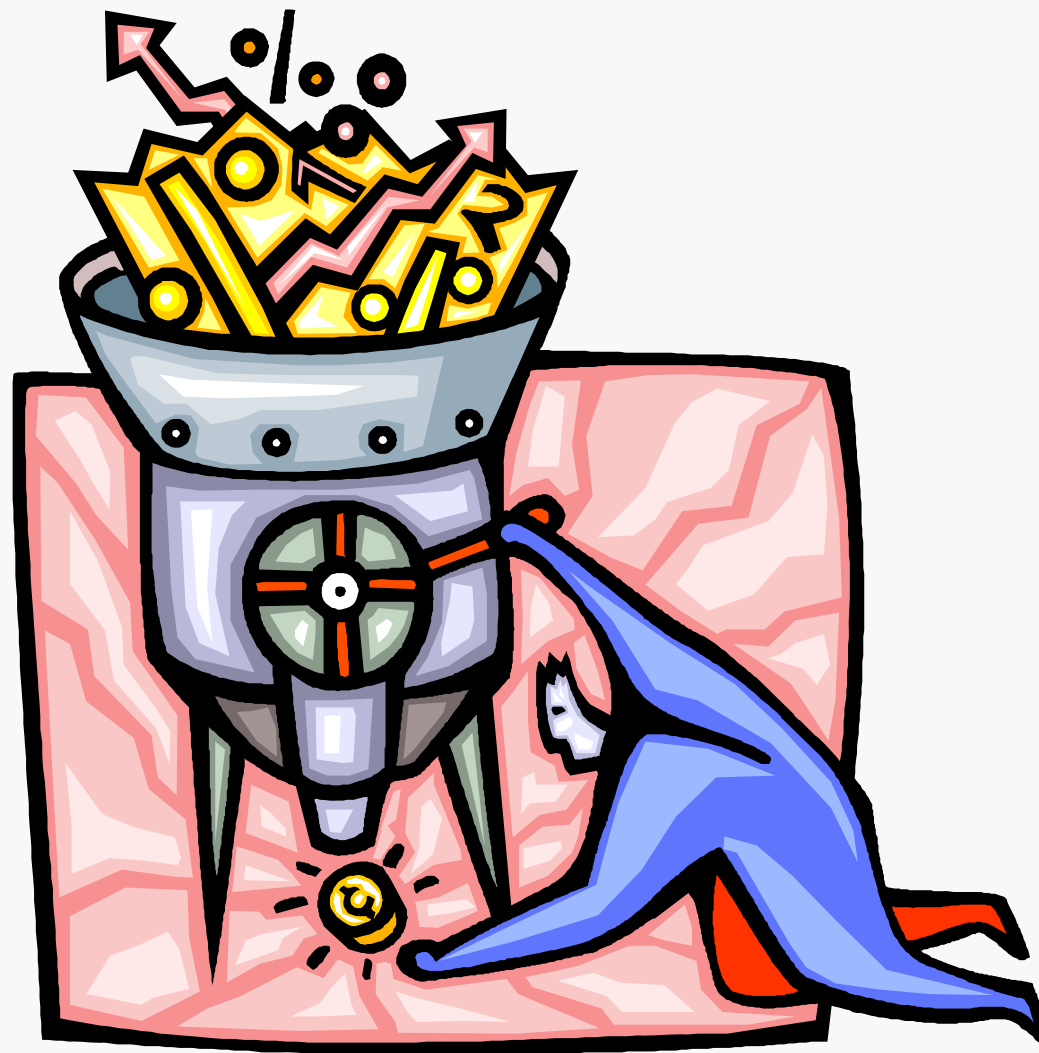
$$\frac{(n+2)!}{n!} = \frac{(n+2)(n+1)n!}{n!} = (n+2)(n+1)$$

# 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...”

“ ...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?**

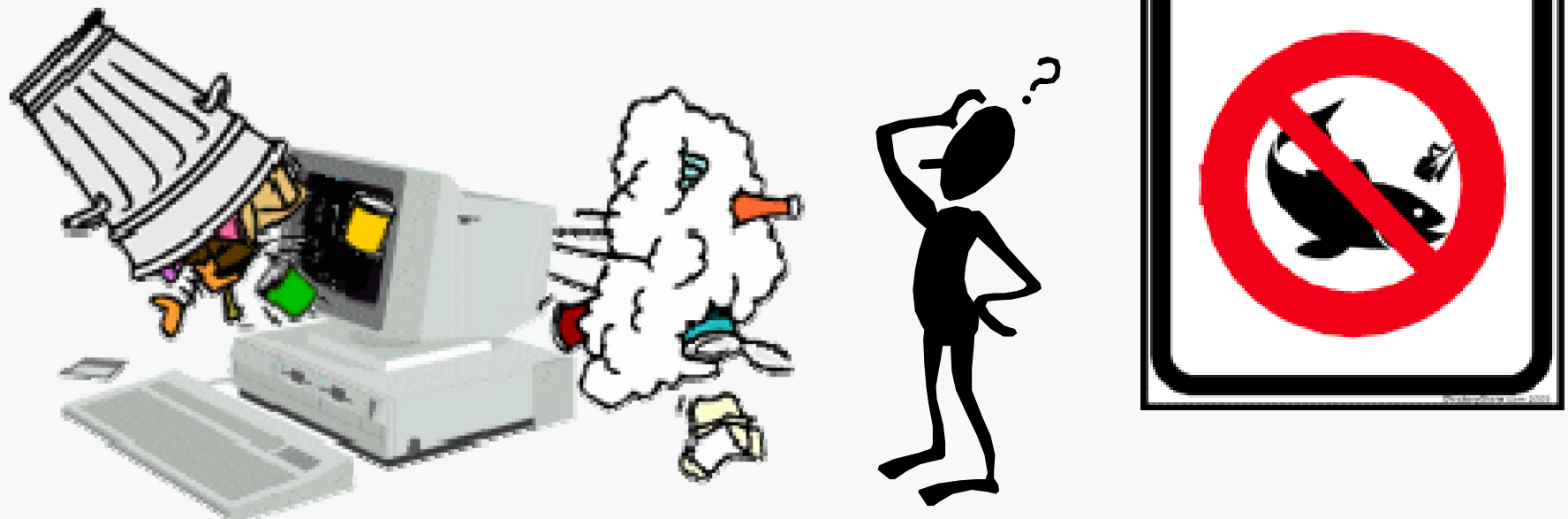
**Kolika je pogreška pri zaključivanju?**

...



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

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



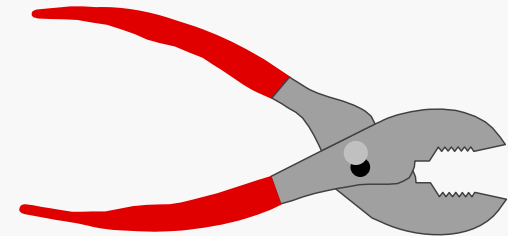
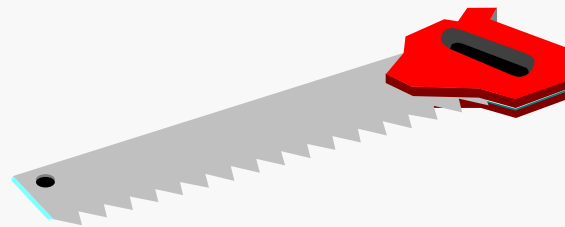
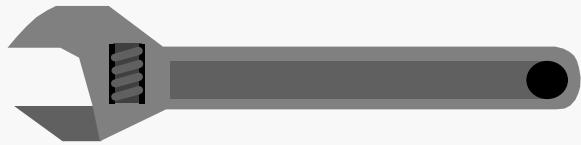
“Savjetovati se sa statističarem **nakon** što je pokus završen često je samo zahtjev da izvrši post-mortem pregled.

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 ?



?

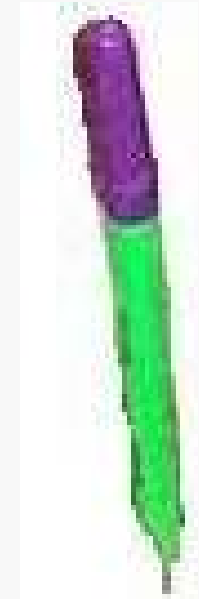




# KAKO ?



?



# ...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**

skup zdravstvenih radnika, skup studenata III godine, .....

(osnovni skup,  
statistička  
masa)

# ...pojam i predmet statistike...

(statističke jedinice,  
entiteti)

- populaciju čine **osnovne jedinice promatranja**

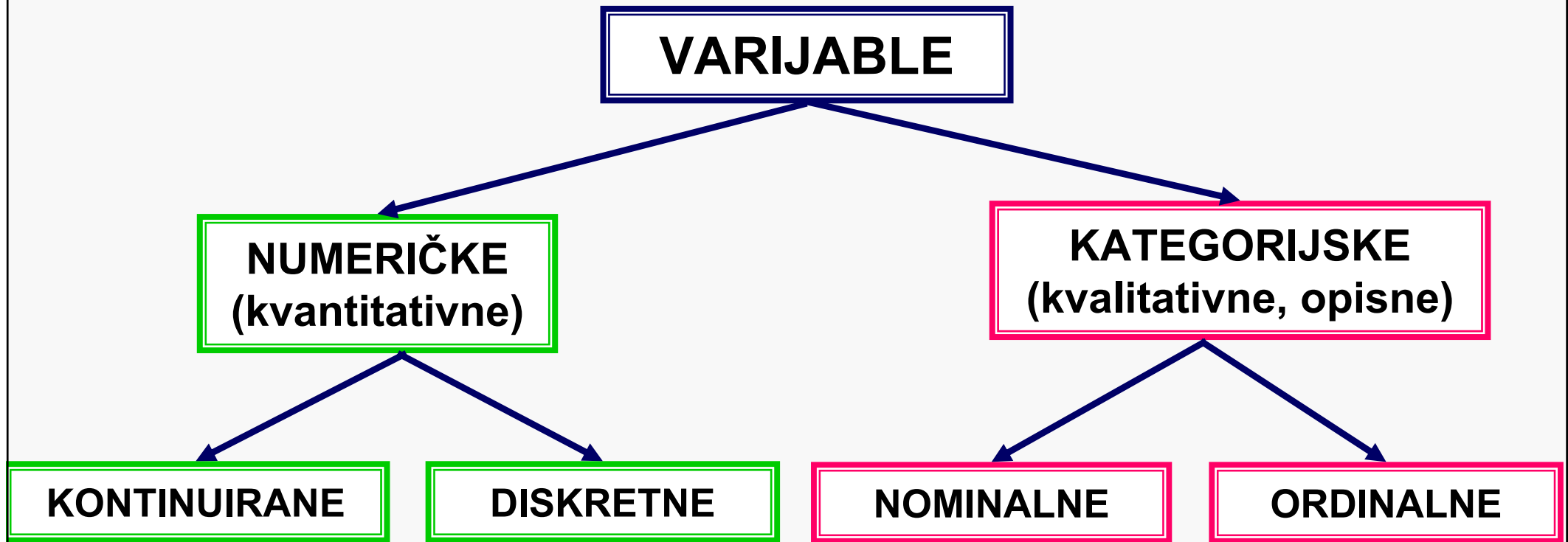
zdravstveni radnik u populaciji  
zdravstvenih radnika

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

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

**varijabla**  
(obilježje, atribut, značajka)

# KLASIFIKACIJA VARIJABLI



# 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, ...**

# Kategorijske 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/smrt) ...**

## ...kategorijske varijable...

- ordinalne varijable

- varijable čije su vrijednosti određene pripadnošću nekoj od kategorija s poznatim redosljedom

**stupanj opekotine, ocjena na ispitu ...**

# 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, ...**

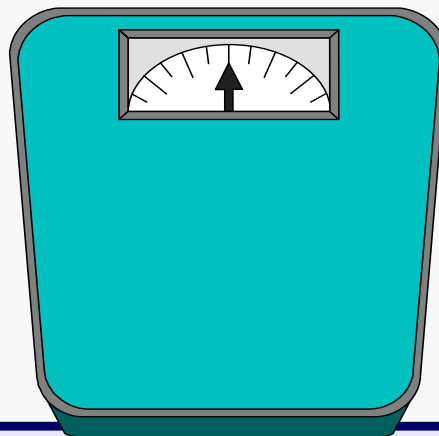
Kriterij	0 bodova	1 bod	2 boda
Boja kože	modra ili blijeda	trup ružičast, okrajine modre	ružičast
Frekvencija srca	nema	<100	>100
Refleksna podražljivost	bez odgovora	grimasa, slab pokret	plač pokret
Mišićni tonus	mlohav	slaba fleksija	jaka fleksija
Disanje	ne diše	grčevito, slab plač	jak, pravilan plač

## ...još neke vrste podataka...

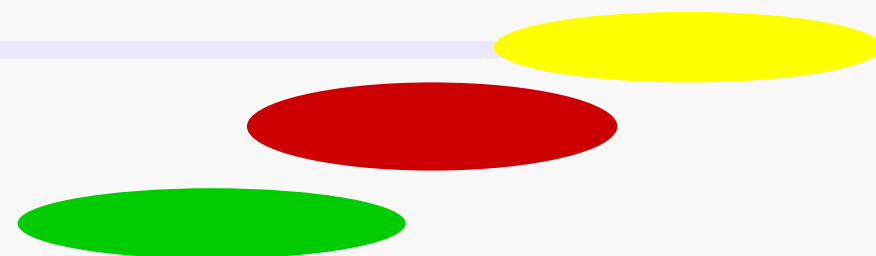
- **cenzorirani** (censored)
  - podatci koji se ne mogu točno izmjeriti, ali je poznato da prelaze neku granicu mjerenja
- najčešće se pojavljuju:
  - u laboratorijskim mjerenjima – vrijednosti ispod/iznad mogućnosti detekcije nekog uređaja
  - u studijama praćenja (follow-up):
    - očekivano svojstvo 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



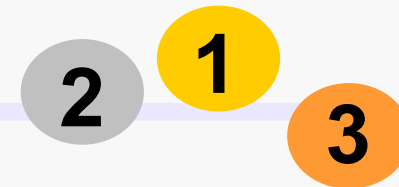
- kategorizacija kojom objektima ili događajima pridružujemo riječi ili simbole
- nema informacije o veličini pojedinačnog rezultata

**spol (muški/ženski)**

**ishod bolesti (preživljavanje/smrt)**

**boja očiju (plava, crna, smeđa, ....)**

## 2. Ordinalna ljestvica

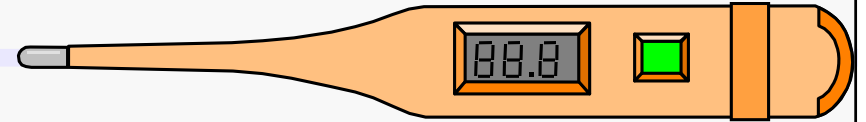


- ima sve karakteristike nominalne ljestvice i dodatno uključuje **redosljed** (nominalna + rangiranje)
- intervali nisu jednaki, a granice među grupama nisu čvrste
- nema informacije o "jačini" razlike između pojedinih grupa

**stupanj opekotine (prvi, drugi, treći)**

**ocjene na ispitu**

### 3. Intervalna ljestvica



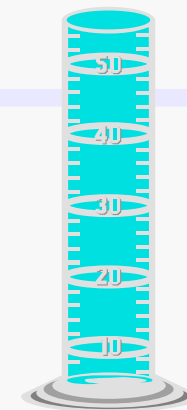
- ljestvica sa jednakim intervalima koji imaju definirane granice => **razlike imaju smisla**
- nema apsolutnu nulu (moguće su i negativne vrijednosti) => **omjeri nemaju smisla**

temperatura, kvocijent inteligencije

$100^{\circ}\text{F} \neq 2 \cdot 50^{\circ}\text{F}$       jer je     $100^{\circ}\text{F} = 38^{\circ}\text{C}$   
 $50^{\circ}\text{F} = 10^{\circ}\text{C}$

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

# 4. Omjerna ljestvica



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

visina, težina, dob, ...

$$90 \text{ kg} = 3 \cdot 30 \text{ kg}$$

$$3174.653 \text{ oz} = 3 \cdot 1058.218 \text{ oz} \quad (1 \text{ kg} = 35.27392 \text{ oz})$$

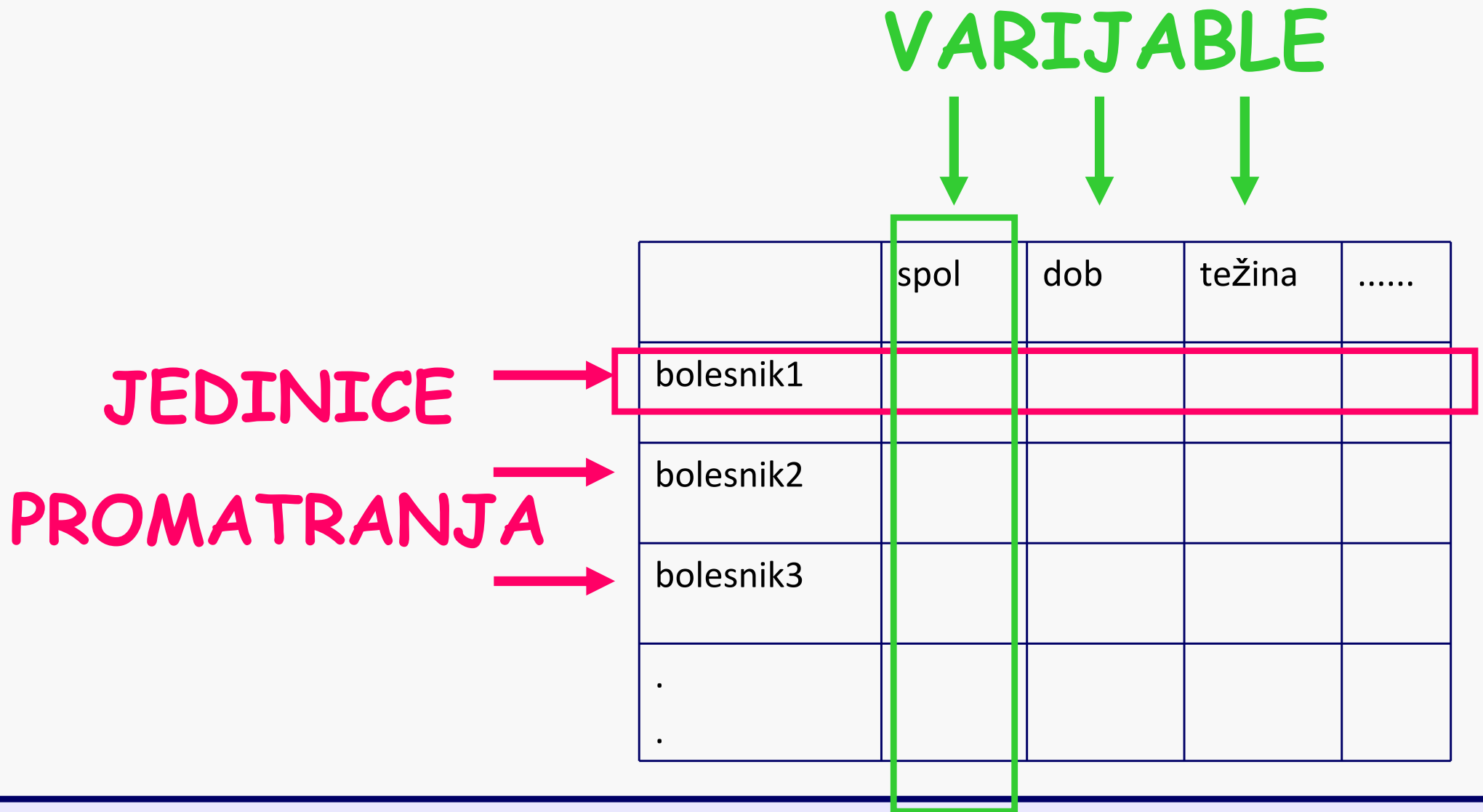


**Rezultati mjerenja na ljestvicama višeg reda mogu se prikazati ljestvicama nižeg reda, ali NE I OBRNUTO !**

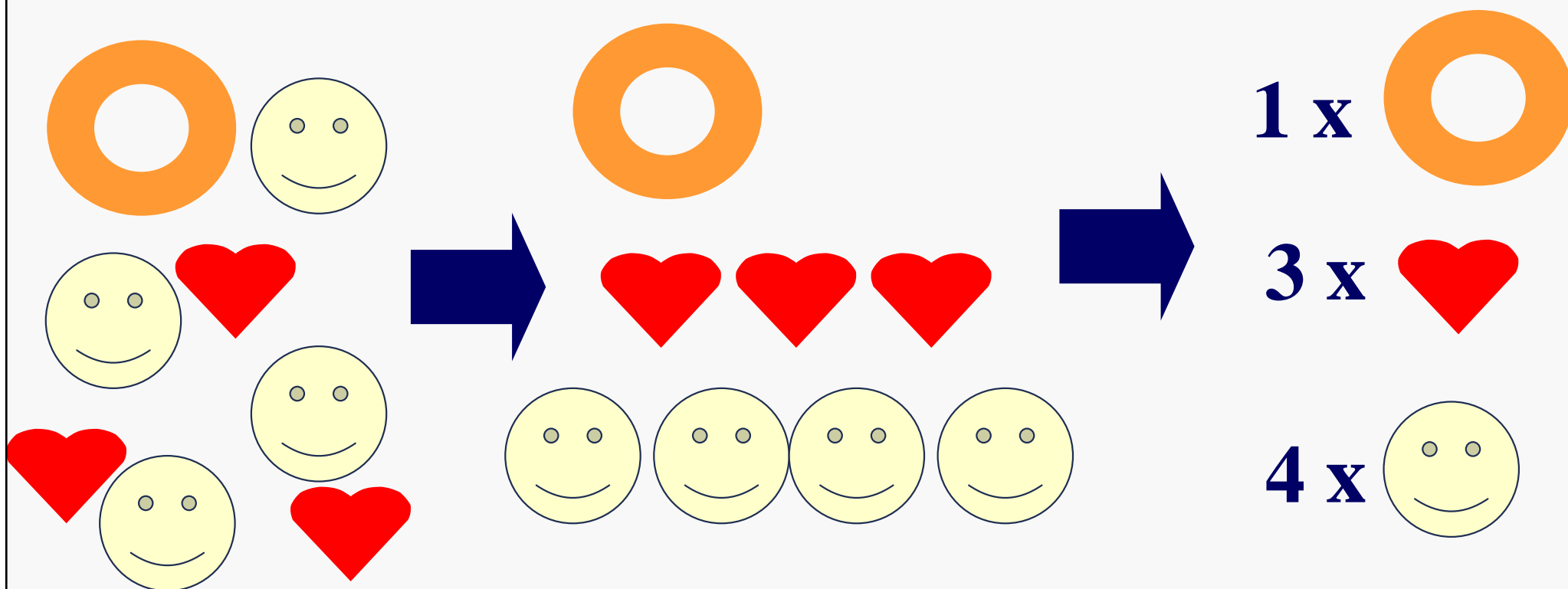
# NEKOLIKO VAŽNIH PRAVILA ...

- 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

# Unos podataka



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

**raspon (interval) razreda**

– **razlika granica razreda**



## 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

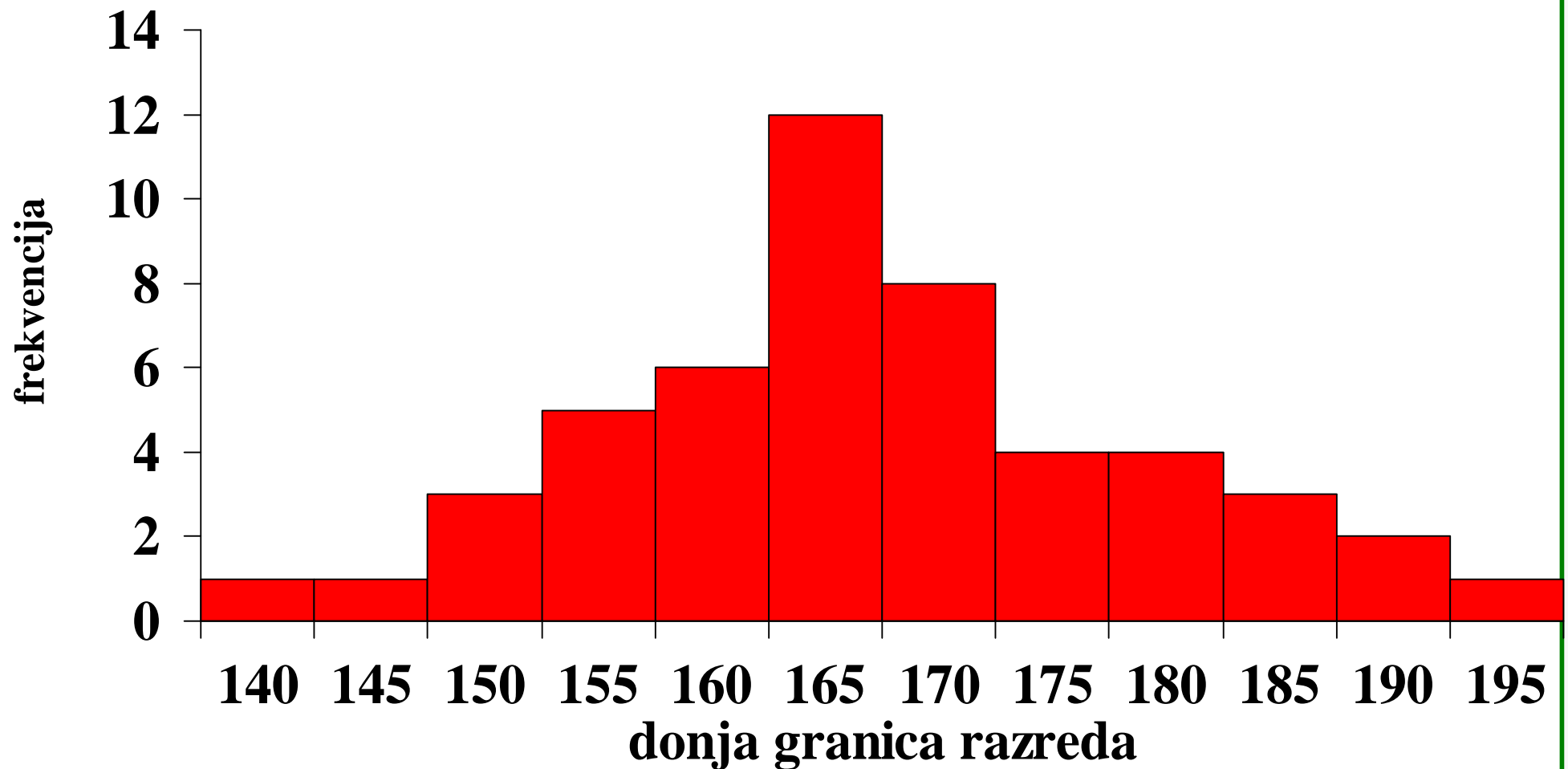
# Tablica frekvencija podataka iz primjera

**min = 140****max = 196**

<b>Broj r.</b>	<b>Granice r.</b>	<b>Sredina r.</b>	<b>f</b>	<b>cf</b>	<b>rf</b>	<b>crf</b>
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
<b>UKUPNO :</b>			<b>50</b>		<b>1,00</b>	

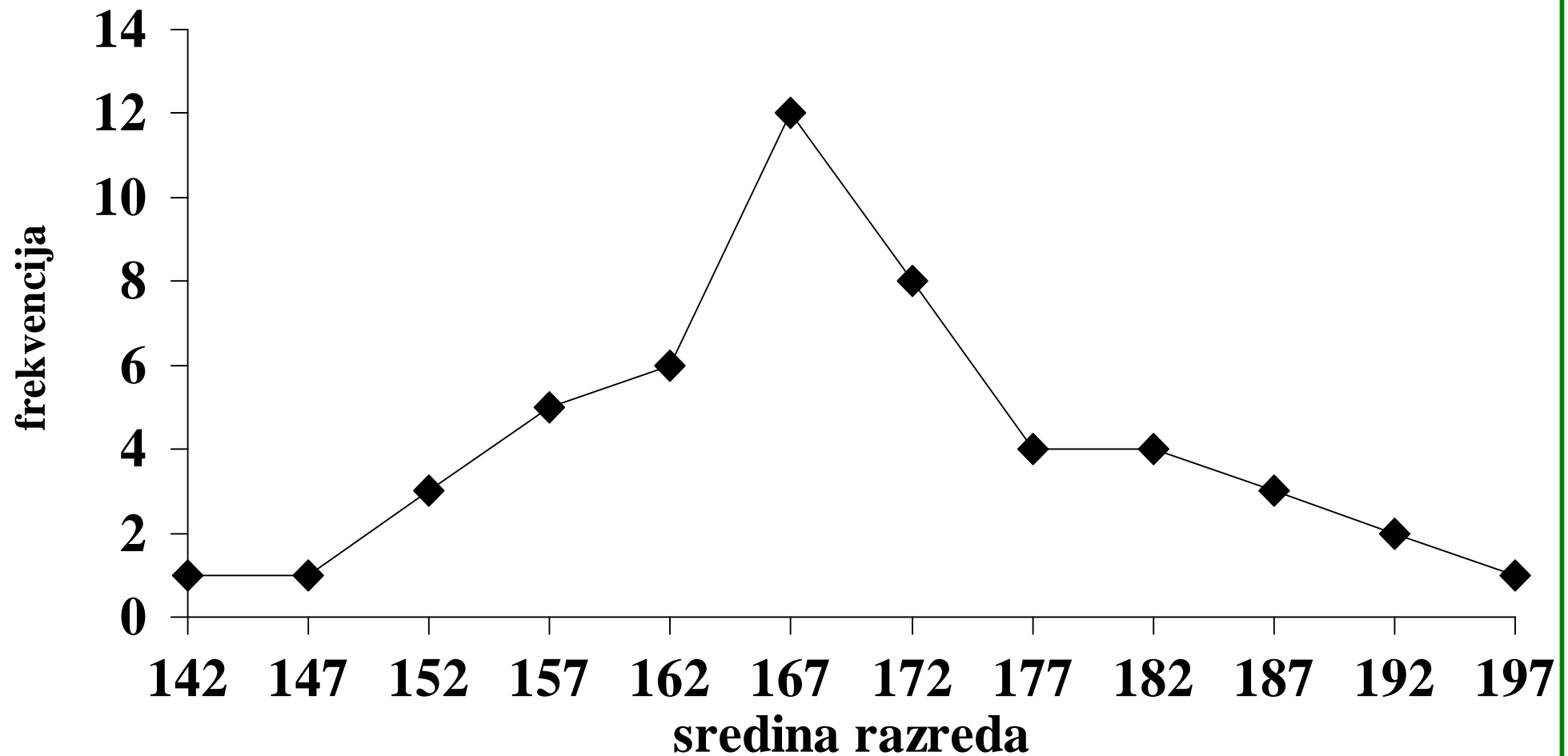
# RAZDIOBA FREKVENCIJA

## HISTOGRAM FREKVENCIJA



# RAZDIoba FREKVENCIJA

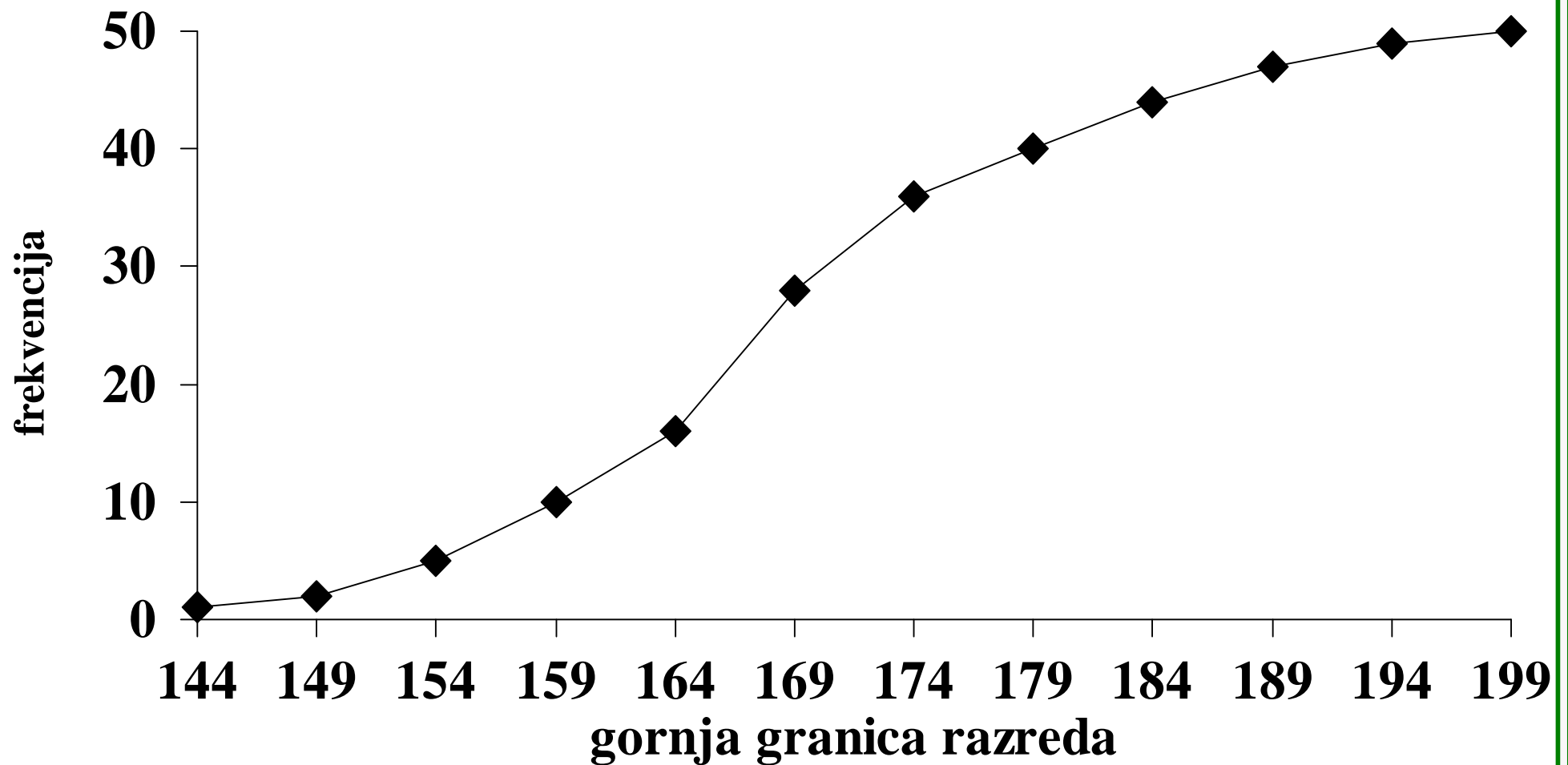
## POLIGON FREKVENCIJA





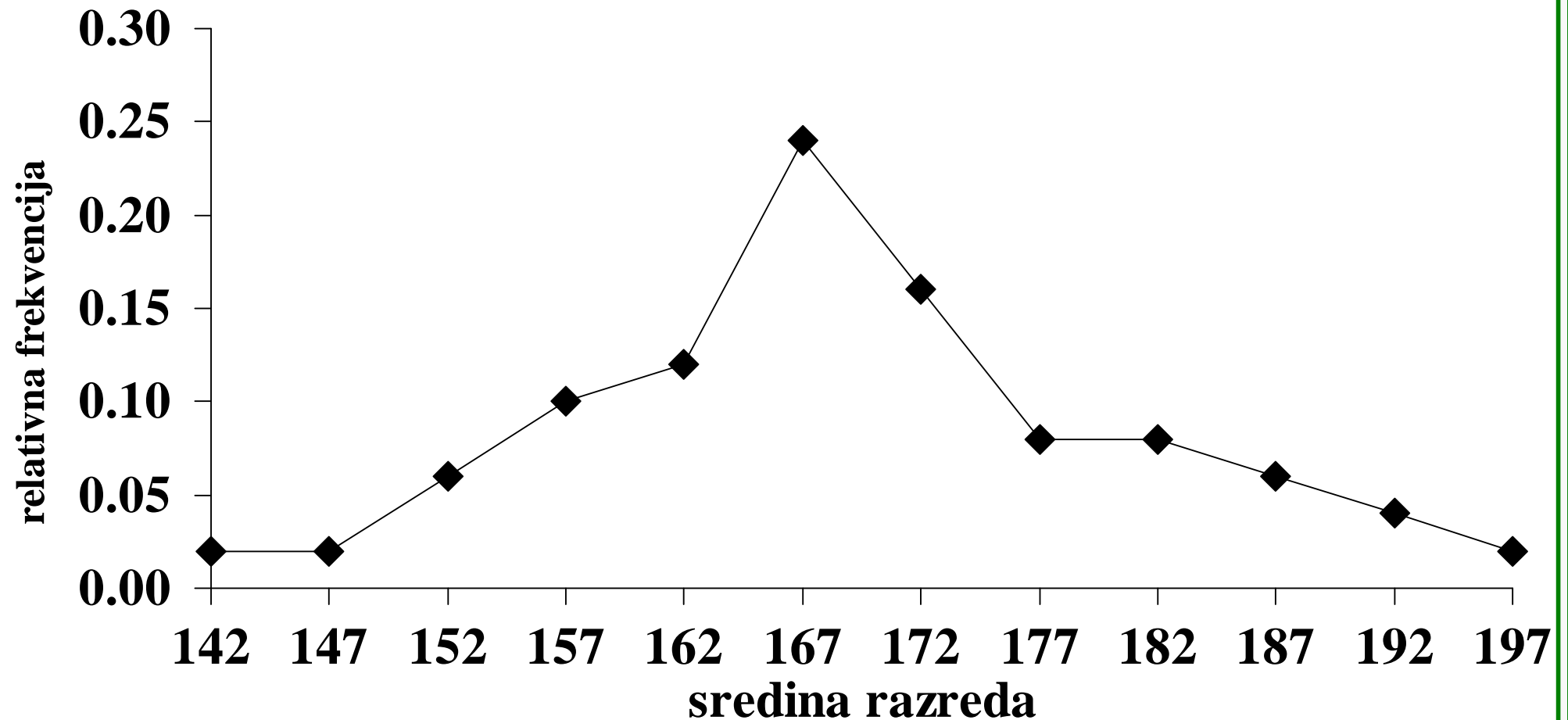
# RAZDIoba FREKVENCIJA

## POLIGON KUMULATIVNIH FREKVENCIJA

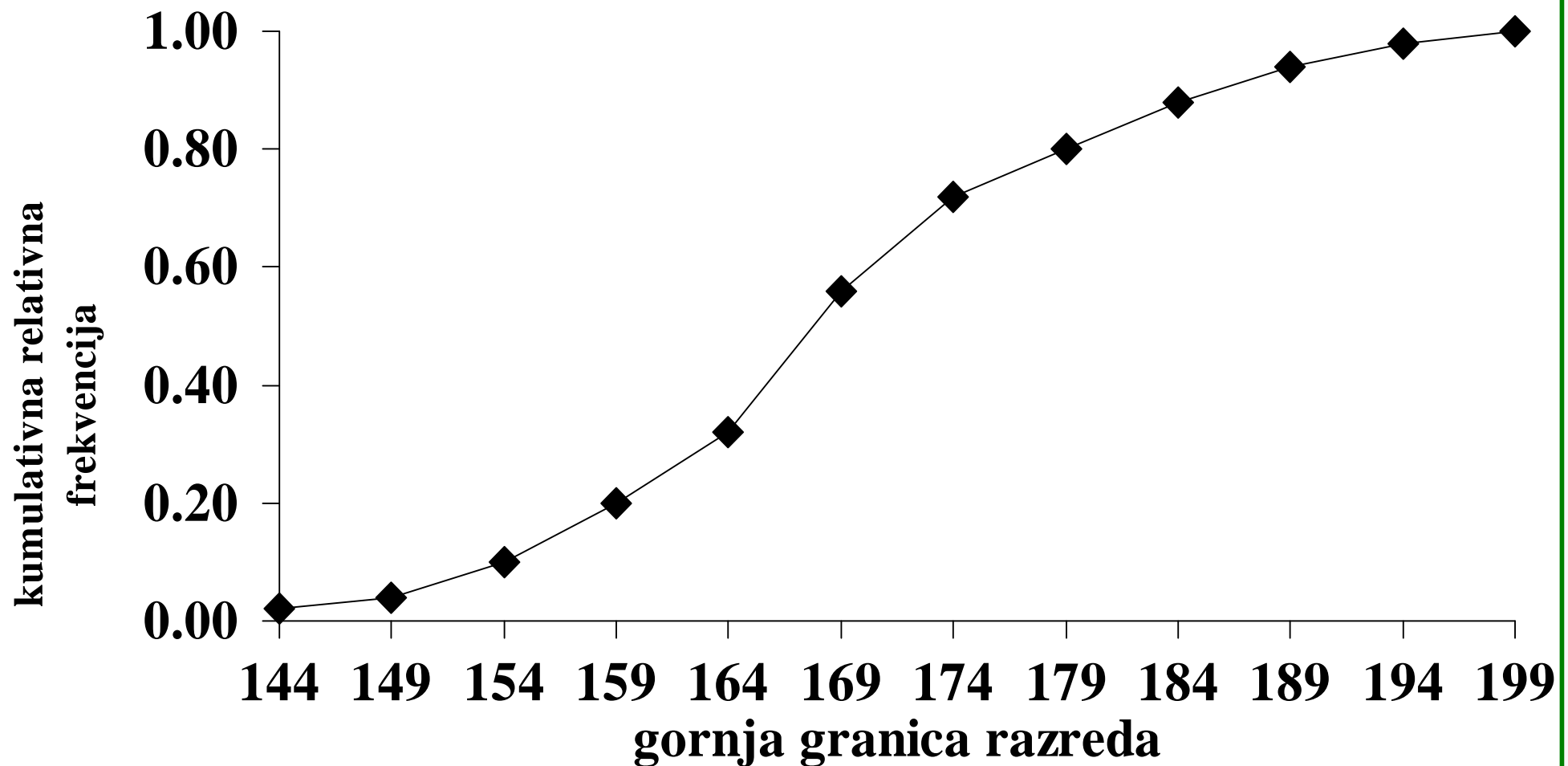


# RAZDIoba FREKVENCIJA

## POLIGON RELATIVNIH FREKVENCIJA



## POLIGON KUMULATIVNIH RELATIVNIH FREKVENCIJA



# Stablo i list (stem-and-leaf)

<b>f</b>	<b>stablo</b>	<b>list</b>
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

# 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

# PAŽNJA !

## – granice razreda

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