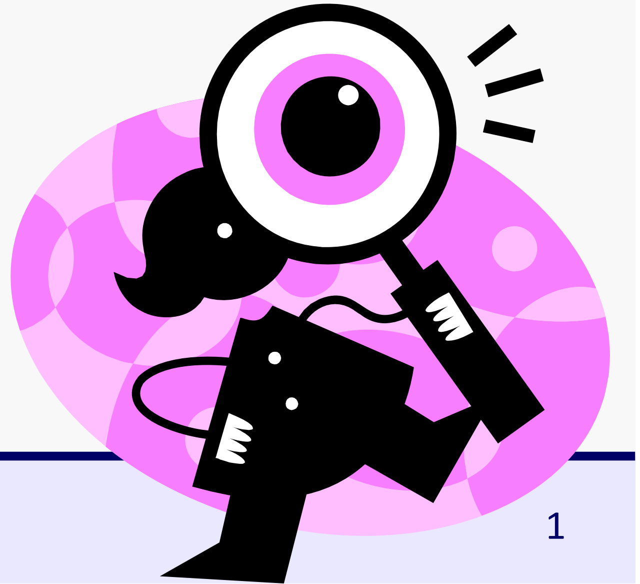
UVODNA RAZMATRANJA

izv. prof. dr. sc. Vesna Ilakovac
vesna.ilakovac@mefos.hr

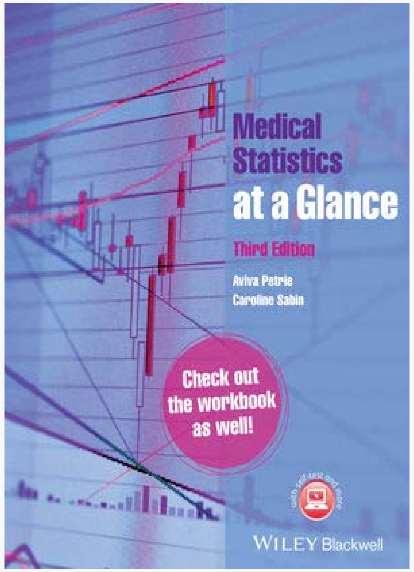


Literatura

- Dawson-Saunders B, Trapp RG. Basic & Clinical Biostatistics. Prentice-Hall Int. Inc., London
- Petrie A, Sabin C. Medical statistics at a glance. Blackwell Science, Oxford.
- Marušić M., Petrak J., Petrovečki M., Marušić A. Uvod u znanstveni rad u medicini. Medicinska naklada Zagreb.

2004.

2009.



2013.

MedCalc Manual

https://www.medcalc.org/manual/index.php.

MedCalc
easy-to-use statistical software

Home Features Download Order Contact FAQ Manual

Contents
Introduction
File menu
Edit menu
View menu
Format menu
Tools menu
Statistics menu

MedCalc manual - contents

Introduction

Program installation
Auto-update
Regional settings support
Selection of display language

Statistics menu (continued)

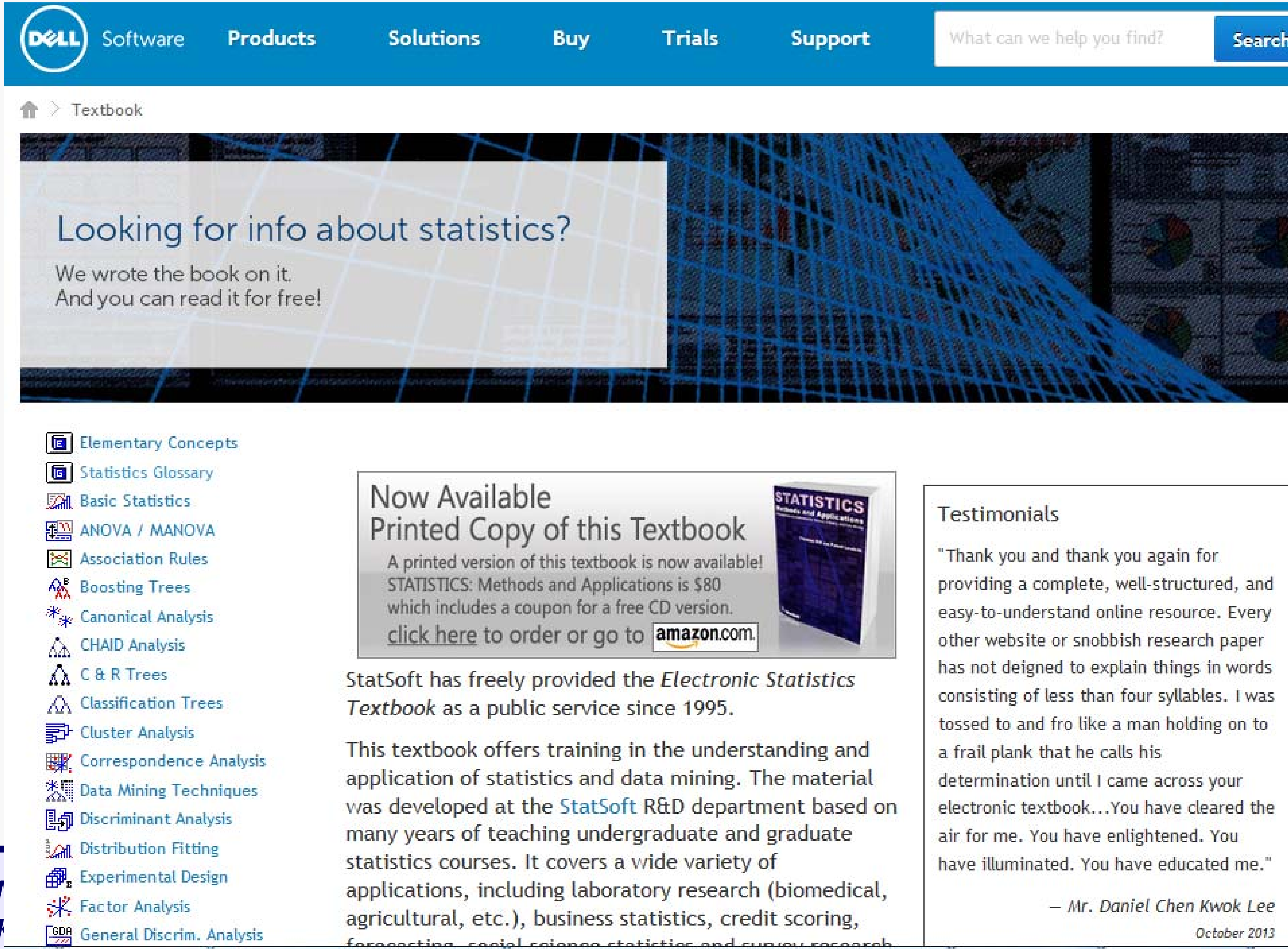
Serial measurements
Reference intervals ...
Reference interval
Age-related reference interval

PDF:

http://www.medcalc.eu/download/medcalcmanual.pdf

Literatura

StatSoft, Inc. (2013). Electronic Statistics Textbook. Tulsa, OK: StatSoft. WEB: http://www.statsoft.com/textbook/.

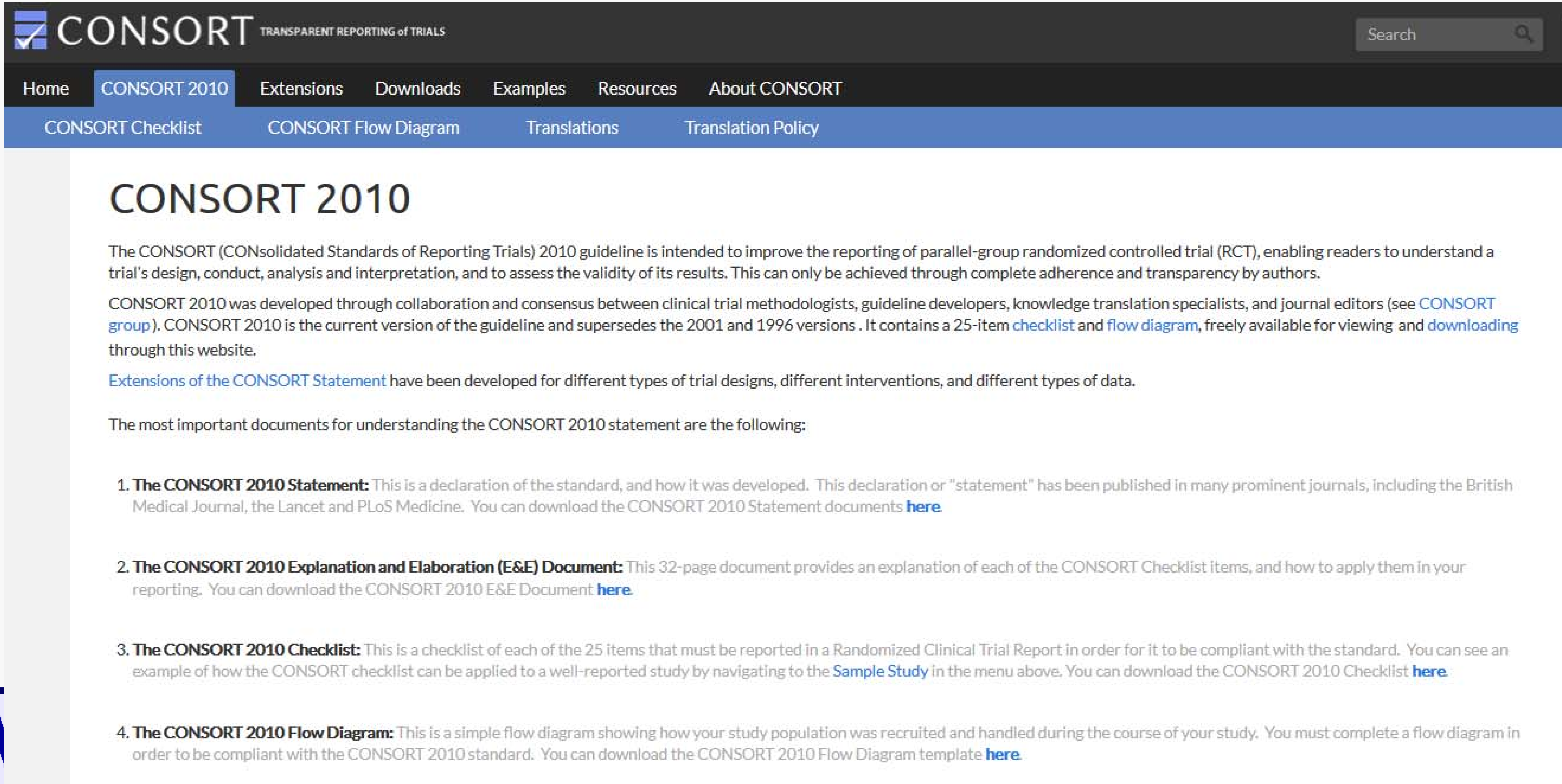


Literatura

Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Ann Int Med 2010;152.

(**http://annals.org/article.aspx?articleid=745807**)

http://www.consort-statement.org/consort-2010



CONSORT TRANSPARENT REPORTING of TRIALS

Search

Home CONSORT 2010 Extensions Downloads Examples Resources About CONSORT

CONSORT Checklist CONSORT Flow Diagram Translations Translation Policy

CONSORT 2010

The CONSORT (CONsolidated Standards of Reporting Trials) 2010 guideline is intended to improve the reporting of parallel-group randomized controlled trial (RCT), enabling readers to understand a trial's design, conduct, analysis and interpretation, and to assess the validity of its results. This can only be achieved through complete adherence and transparency by authors.

CONSORT 2010 was developed through collaboration and consensus between clinical trial methodologists, guideline developers, knowledge translation specialists, and journal editors (see CONSORT group). CONSORT 2010 is the current version of the guideline and supersedes the 2001 and 1996 versions . It contains a 25-item checklist and flow diagram, freely available for viewing and downloading through this website.

Extensions of the CONSORT Statement have been developed for different types of trial designs, different interventions, and different types of data.

The most important documents for understanding the CONSORT 2010 statement are the following:

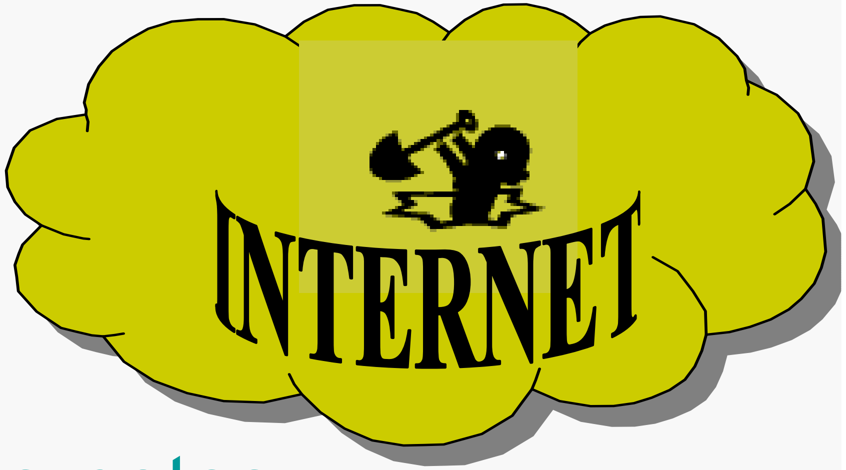
1. **The CONSORT 2010 Statement:** This is a declaration of the standard, and how it was developed. This declaration or "statement" has been published in many prominent journals, including the British Medical Journal, the Lancet and PLoS Medicine. You can download the CONSORT 2010 Statement documents **here**.

2. **The CONSORT 2010 Explanation and Elaboration (E&E) Document:** This 32-page document provides an explanation of each of the CONSORT Checklist items, and how to apply them in your reporting. You can download the CONSORT 2010 E&E Document **here**.

3. **The CONSORT 2010 Checklist:** This is a checklist of each of the 25 items that must be reported in a Randomized Clinical Trial Report in order for it to be compliant with the standard. You can see an example of how the CONSORT checklist can be applied to a well-reported study by navigating to the Sample Study in the menu above. You can download the CONSORT 2010 Checklist **here**.

4. **The CONSORT 2010 Flow Diagram:** This is a simple flow diagram showing how your study population was recruited and handled during the course of your study. You must complete a flow diagram in order to be compliant with the CONSORT 2010 standard. You can download the CONSORT 2010 Flow Diagram template **here**.

Dopunsko štivo



BMJ STATISTICS NOTES

www.csm-oxford.org.uk/?o=1292

www.bmj.com/specialties/statistics-notes

How to Upset the Statistical Referee

www-users.york.ac.uk/~mb55/talks/upset.htm

Lang T. Twenty statistical errors even YOU can find in biomedical research articles. Croat Med J 2004;45:361-70.

www.cmj.hr/2004/45/4/15311405.htm

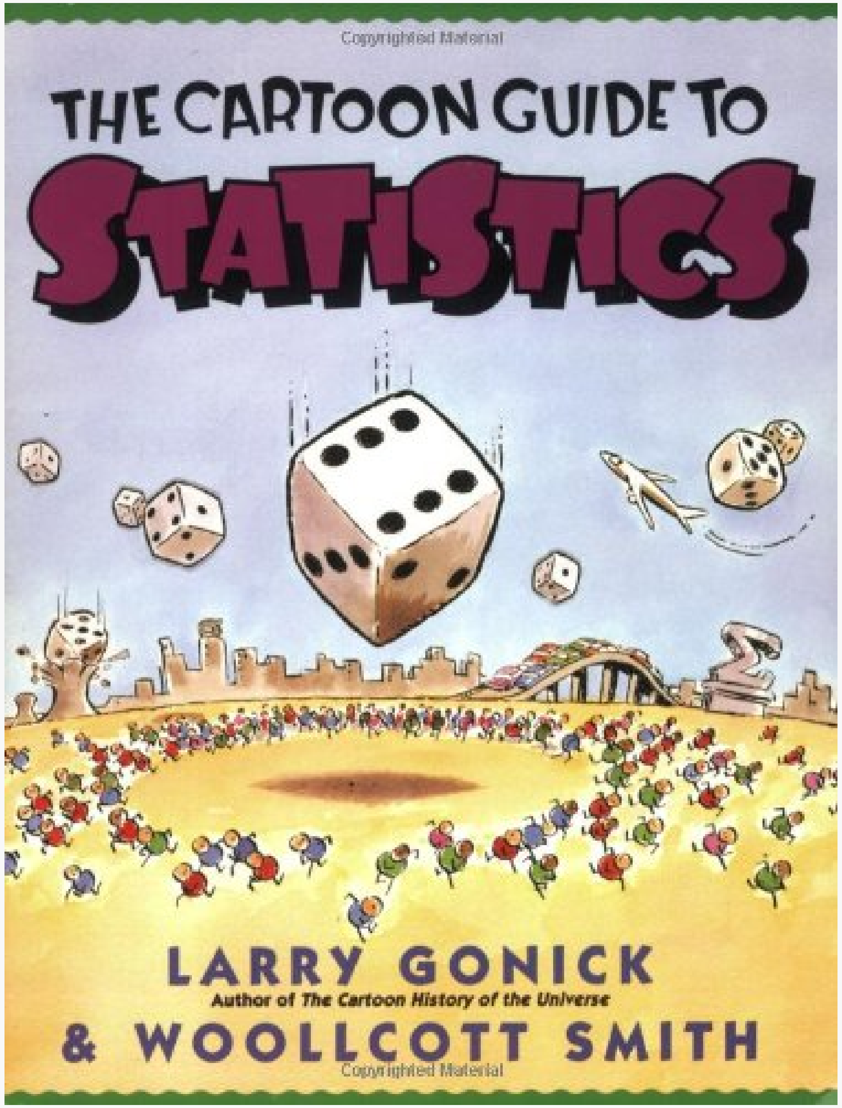
....

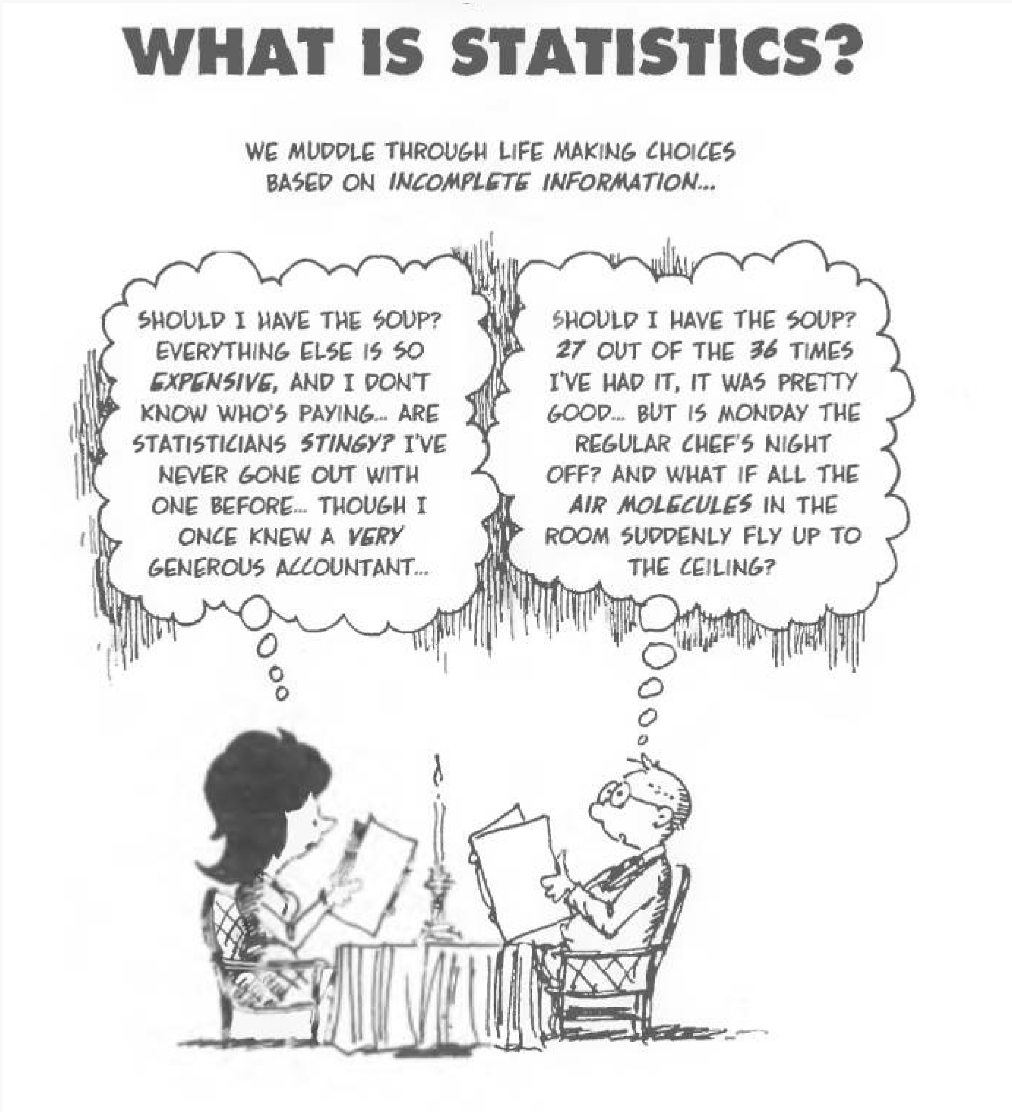
Lako čitljiva literatura......

Petz B, Kolesarić V, Ivanec D.
Petzova statistika : osnovne statističke metode za nematematičare. Naklada Slap.

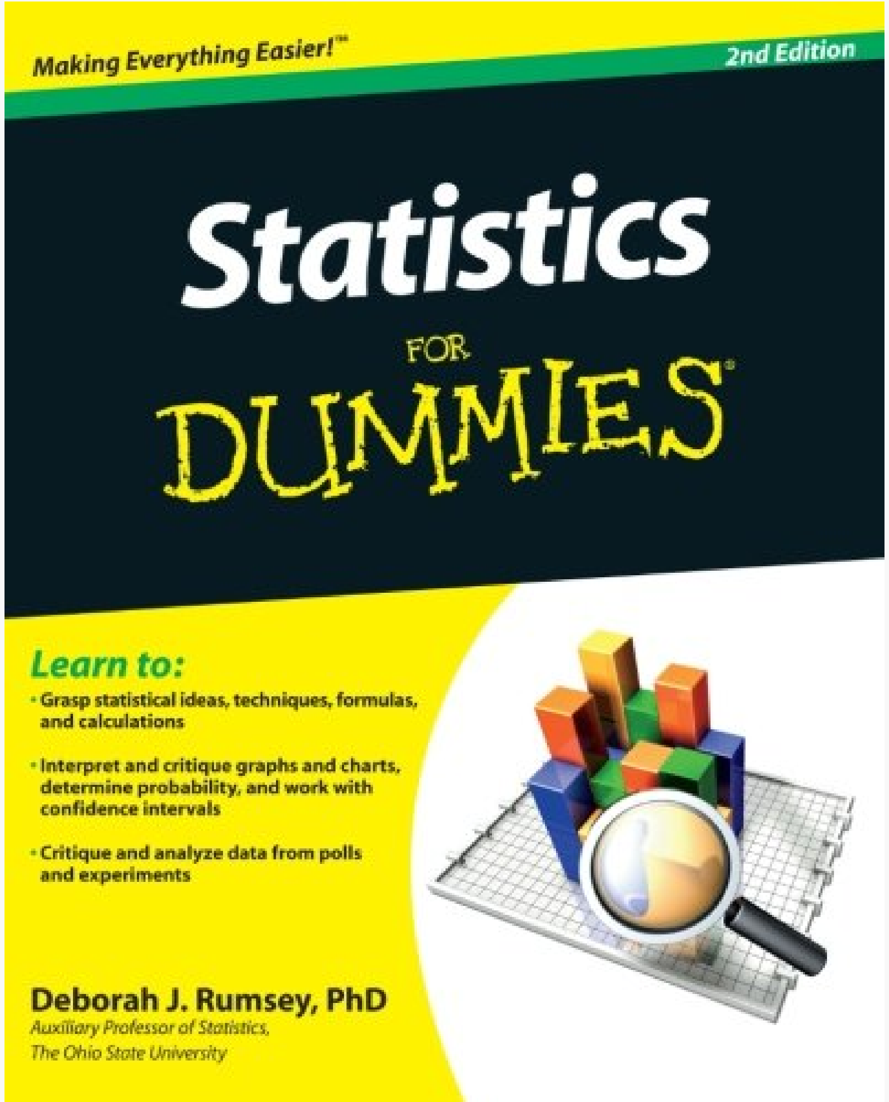
2012.

Ako ništa ne pomaže, pokušajte započeti s ......





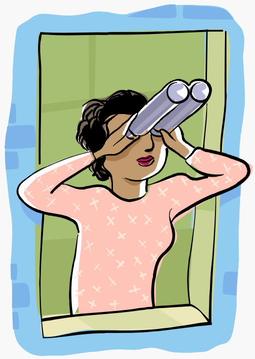
...... ili


Making Everything Easier!™
2nd Edition
Statistics
FOR
DUMMIES®
Learn to:
• Grasp statistical ideas, techniques, formulas, and calculations
• Interpret and critique graphs and charts, determine probability, and work with confidence intervals
• Critique and analyze data from polls and experiments
Deborah J. Rumsey, PhD
Auxiliary Professor of Statistics, The Ohio State University

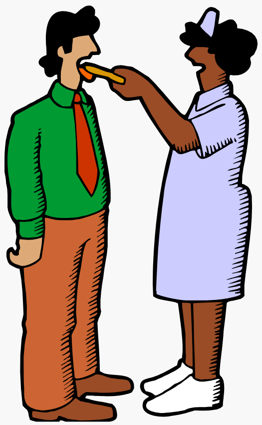
SVRHA ISTRAŽIVANJA

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

TEMELJNE VRSTE ISTRAŽIVANJA

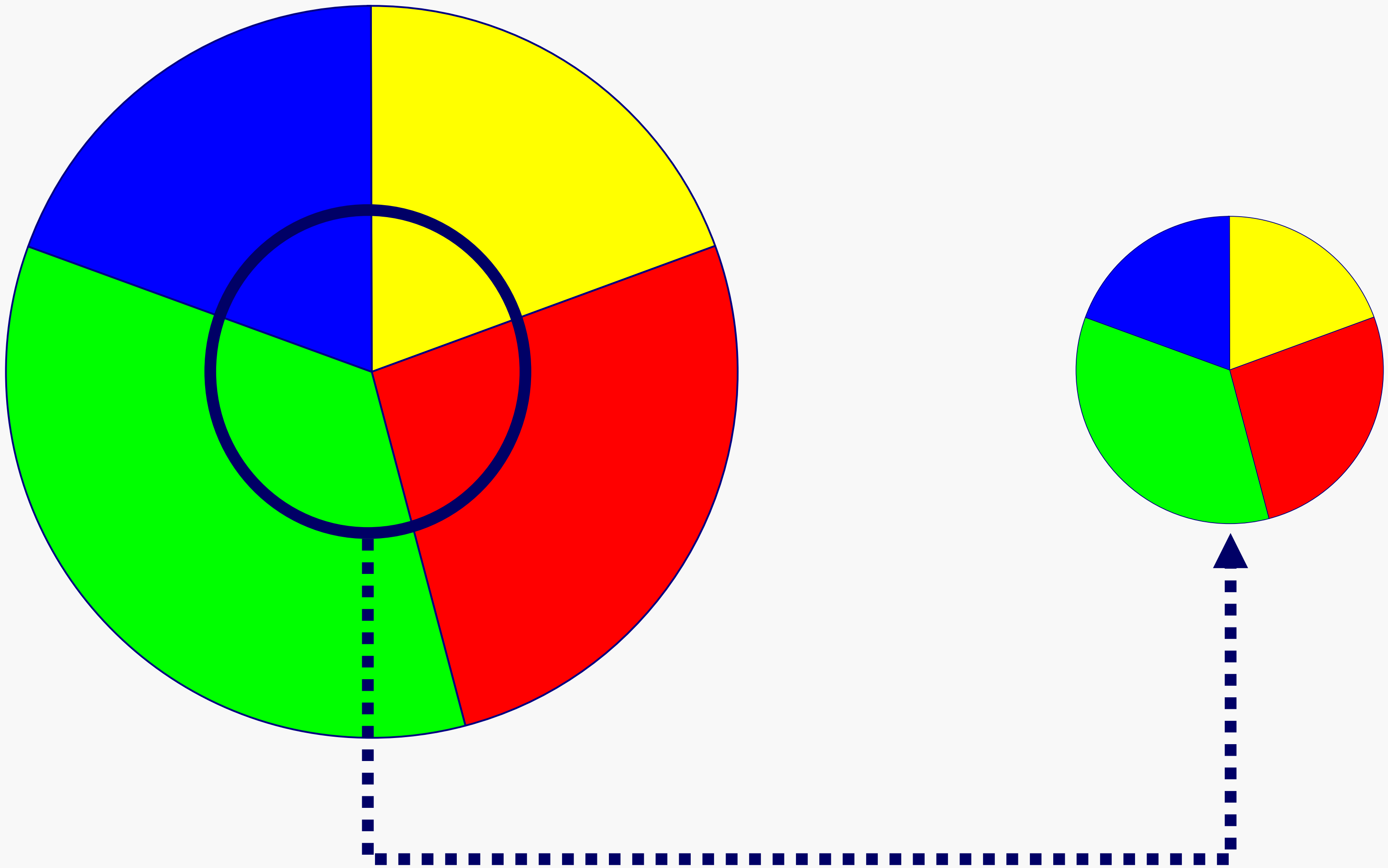


- **opažajna istraživanja (observational studies)**
  - presječno istraživanje (cross-sectional)
  - istraživanje slučajeva i kontrola (case-control)
  - kohortno istraživanje (cohort)
- **pokusna istraživanja**
  - randomizirani kontrolirani pokus (randomised controlled trial-RCT)



PRESJEČNO ISTRAŽIVANJE (cross-sectional)

"Što se dešava?"



ISTRAŽIVANJE SLUČAJEVA I KONTROLA (case-control)

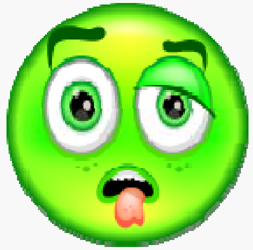
"Što se desilo?"

slučaj (case)

IZLOŽENI

NEIZLOŽENI

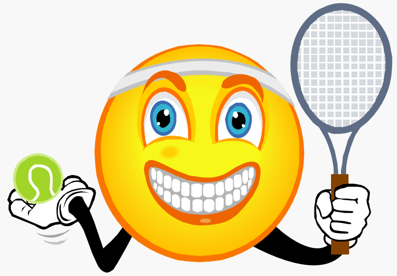
BOLESNI



IZLOŽENI

NEIZLOŽENI

ZDRAVI



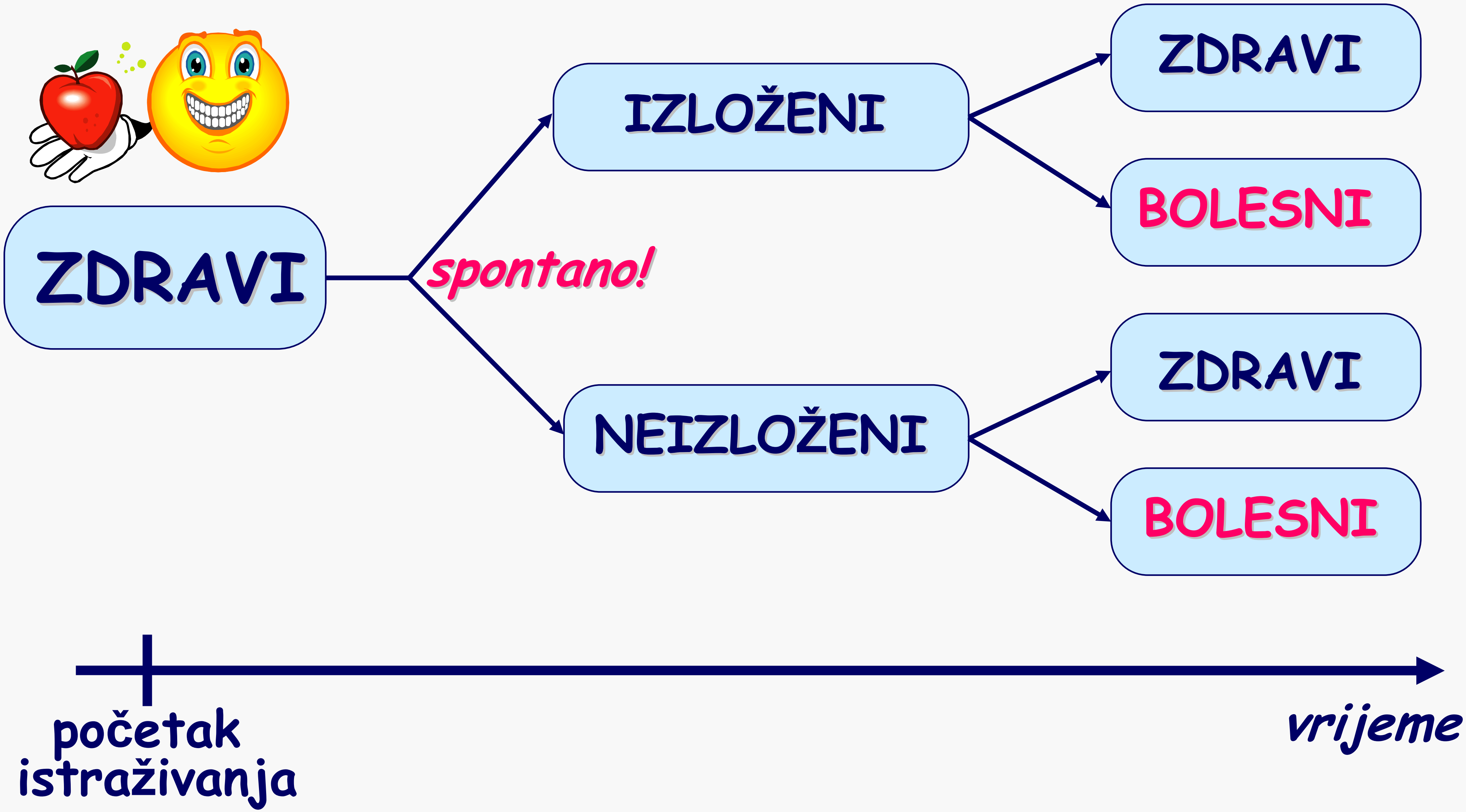
kontrola (control)

početak istraživanja

vrijeme

KOHORTNO ISTRAŽIVANJE

"Što će se desiti?"



RANDOMIZIRANI KONTROLIRANI POKUS (RCT)

UZORAK

randomizacija!

KONTROLNA SKUPINA

ZDRAVI

BOLESNI

POKUSNA SKUPINA

ZDRAVI

BOLESNI

INTERVENCIJA

početak istraživanja

vrijeme

PRIKRIVANJE (*masking, blinding*)

- **jednostruko prikriveni (slijepi) pokus**
  - ispitanik ne zna kojoj skupini pripada

- **dvostruko slijepi pokus**
  - niti ispitanik niti istraživač ne znaju kojoj skupini pripada ispitanik

- **trostruko slijepi pokus**
  - niti ispitanik, niti istraživač niti osoba koja obrađuje podatke ne zna tko pripada ispitivanoj skupini

"DOKAZI" IZ ISTRAŽIVANJA

- različite vrste istraživanja donose "dokaze" različite “snage”

- dvostruko slijepi RCT smatra se “zlatnim standardom” za donošenje najsnažnijih uzročno-posljedičnih znanstvenih zaključaka

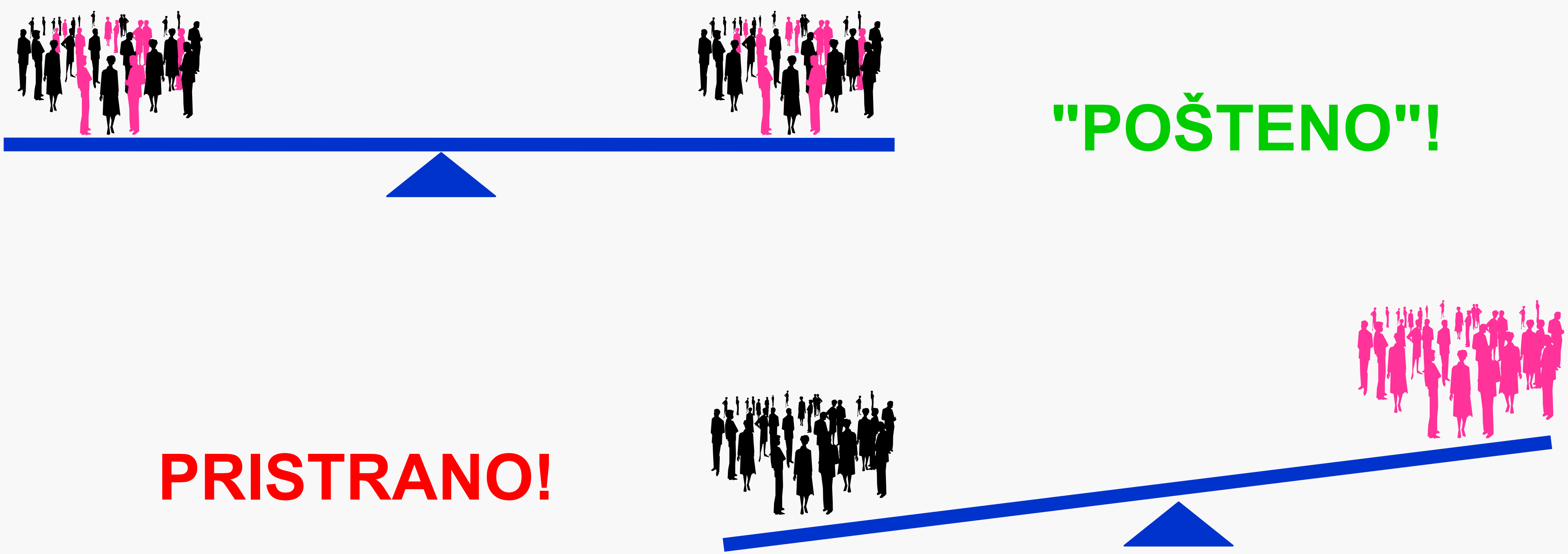
ISTRAŽIVANJE

"DOKAZNA" SNAGA ISTRAŽIVANJA

- **OPAŽAJNO**
  - bez kontrole
  - slučajeva i kontrola
  - kohortno
- **POKUSNO**
  - ne randomizirana kontrola
  - randomizirana kontrola (RCT)
    - bez prikrivanja
    - jednostruko slijepi
    - dvostruko slijepi

ZAŠTO JE POTREBNA RANDOMIZACIJA (nasumični odabir)?

- **slučajnost izbora** u postupku izbora grupa (tko dobiva koji tretman) **izjednačava grupe** s obzirom na poznate i nepoznate prognostičke faktore



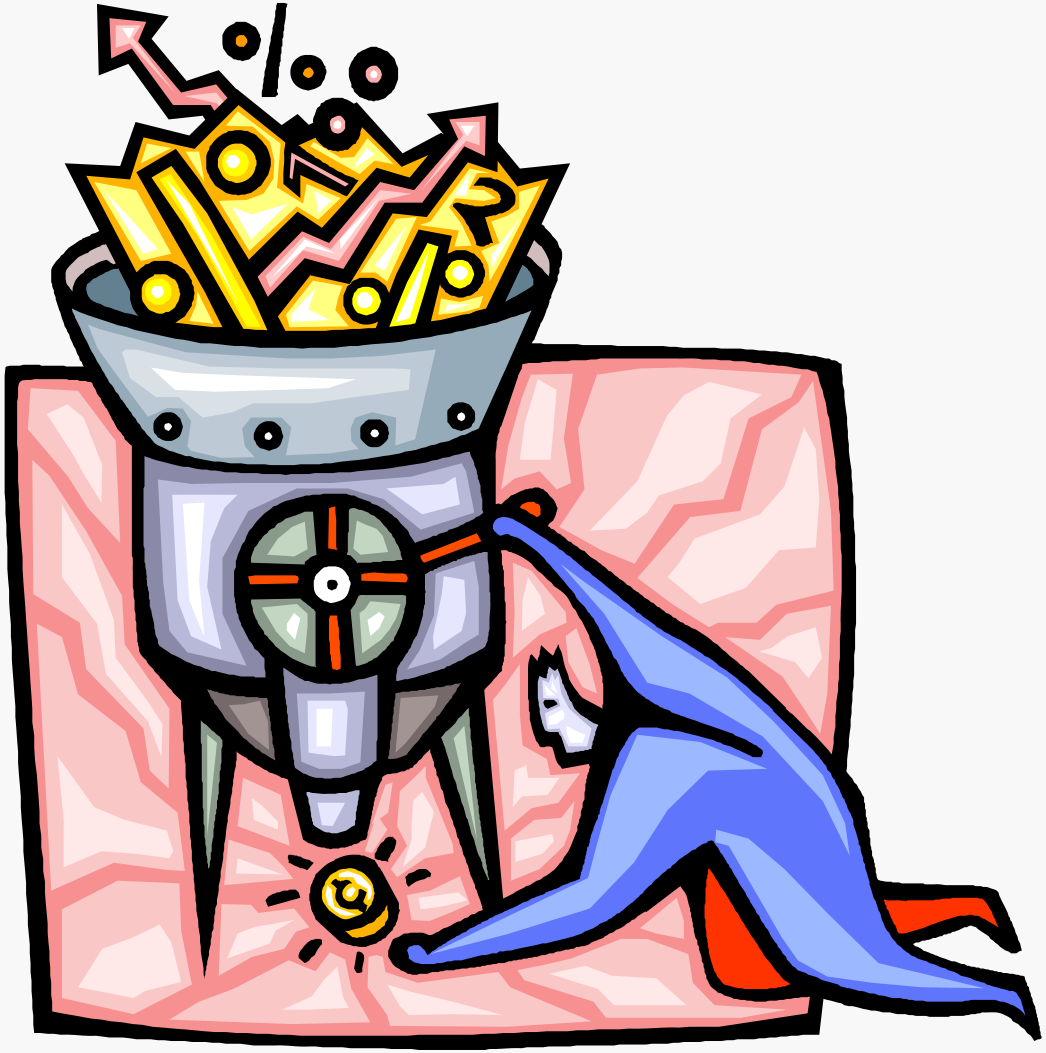
RANDOMIZACIJA (nasumični odabir)

- generatori slučajnih brojeva
  - mjerenja fizikalnih pojava (atmosferski šum, termalni šum, ...)
  - matematički algoritmi (pseudo-slučajni brojevi)
- biranje slučajnih brojeva:
  - Excel: RAND(), RANDBETWEEN()
  - https://www.random.org/
  - http://www.graphpad.com/quickcalcs/randMenu/
  - http://www.randomizer.org/

ZAŠTO JE POTREBNO PRIKRIVANJE?

- prikrivanje alokacije tretmana je dio dobrog nacrta studije
- izbjegavaju se moguće optužbe za pristranost ukoliko niti ispitanici niti ispitivači ne znaju pripadnost pojedinim grupama
- dvostruko slijepi pokus je najbolji način da se očuva znanstvena strogost RCT

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"..

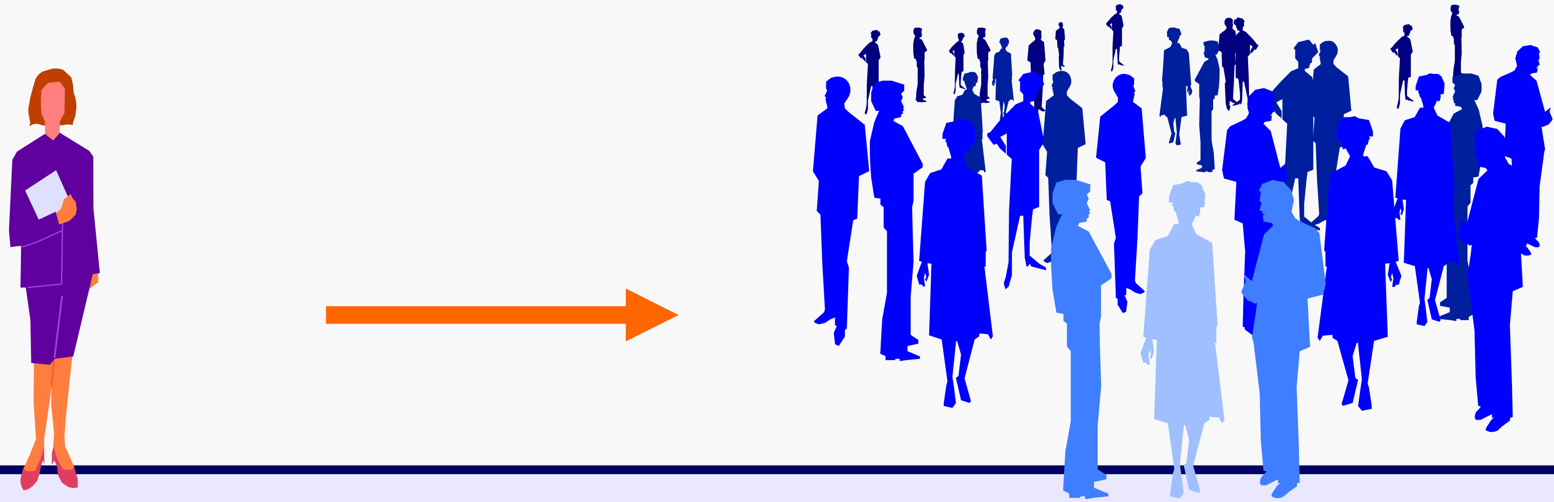
Kada i pod kojim uvjetima možemo zaključivati o populaciji iz rezultata mjerenja provedenih na nekoj skupini ispitanika?

Kako odabrati ispitanike?

Koliko ispitanika uključiti u promatranje?

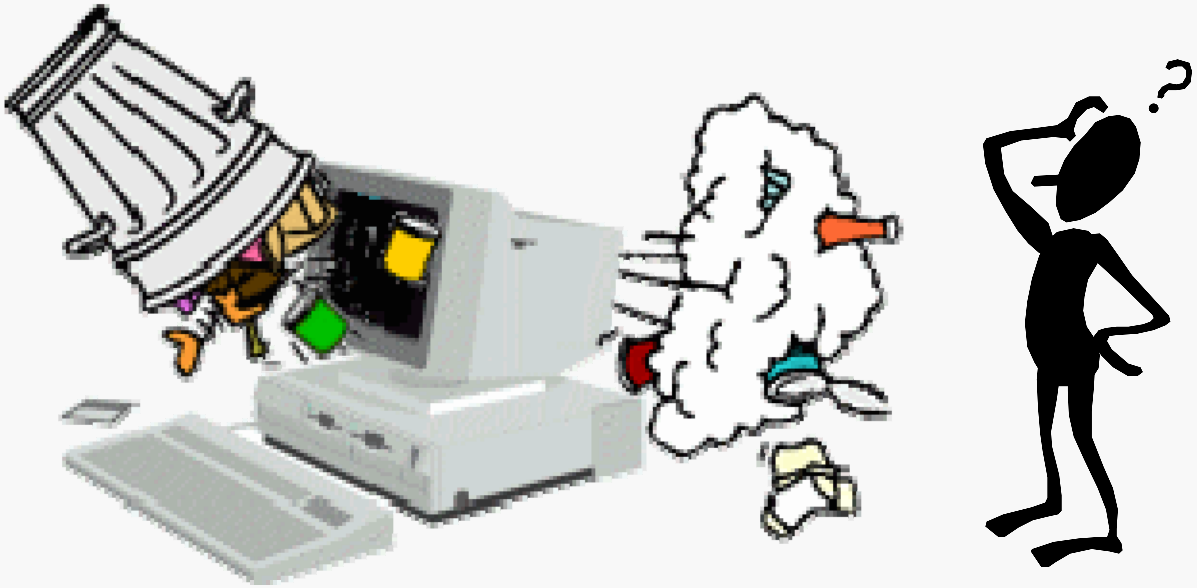
Kolika je pogreška pri zaključivanju?

...



neki razlozi - planiranje istraživanja i pokusa

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



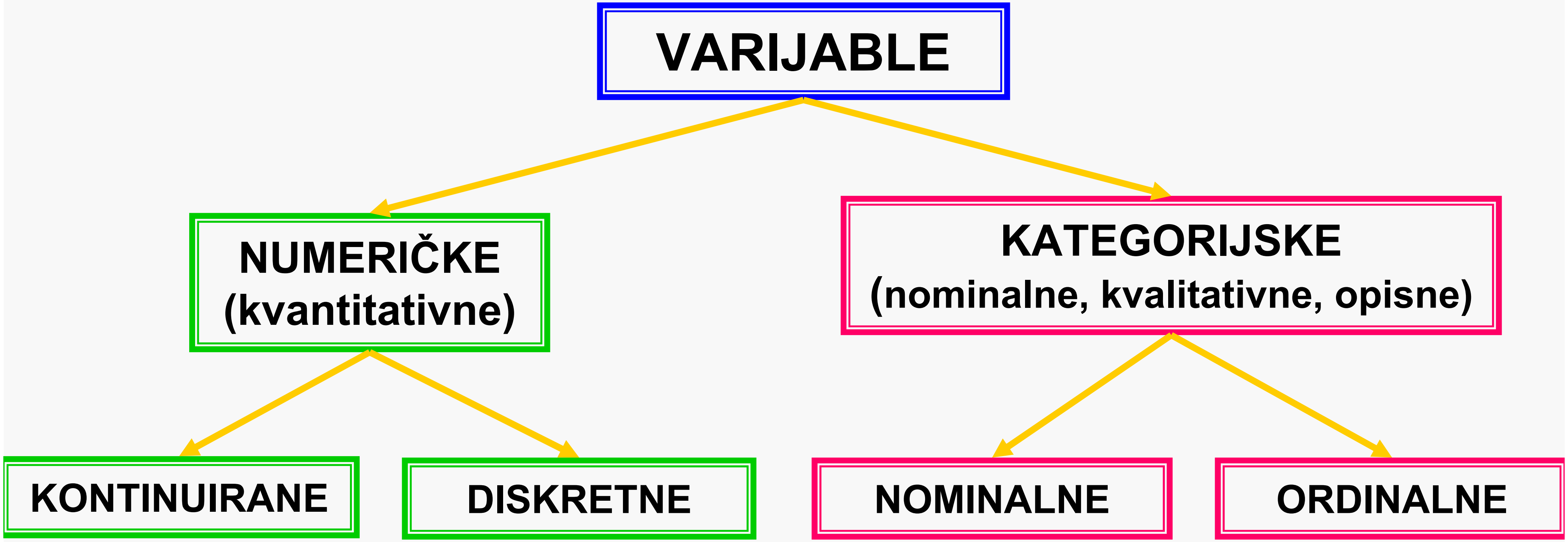
Sir ***Ronald Aylmer Fisher*** *(1890. –1962.)*

"To consult the statistician after an experiment is finished is often merely to ask him to conduct a *post mortem* examination.

He can perhaps say what the experiment died of."

R. A. Fisher, 1938.

VARIJABLE


VARIJABLE
NUMERIČKE (kvantitativne)
KATEGORIJSKE (nominalne, kvalitativne, opisne)
KONTINUIRANE
DISKRETNE
NOMINALNE
ORDINALNE

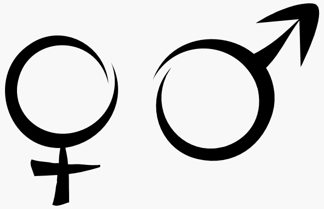
Kvantitativne (numeričke) varijable

- **kontinuirane varijable**
  - varijable koje teorijski pretpostavljaju postojanje beskonačnog broja vrijednosti
  - **u pravilu dobivene mjerenjem**
    - visina, težina, dob, tlak, ...
- **diskretne varijable**
  - varijable čija se vrijednost uvećava ili umanjuje za cijeli broj jedinica
  - **u pravilu dobivene brojanjem**
    - broj popušenih cigareta na dan, broj otkucaja srca u minuti, broj djece, broj proteinskih obroka tjedno,....

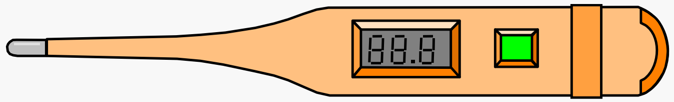
Kvalitativne (opisne) varijable

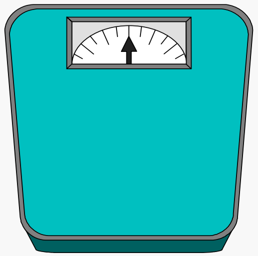
- **bez redosljeda** (*nominalne*)
  - bračno stanje, spol, krvna grupa, zanimanje, radni status, ...
- **s poznatim redosljedom** (*ordinalne*)
  - samostalnost: nije ovisan, ovisan u manjem stupnju, ovisan u višem stupnju, ovisan u visokom stupnju, potpuno ovisan
  - sluh: dobar, oštećen, gluh
  - ......
- **dvije kategorije** (*binarne, dihotomne*)
  - spol (muški/ženski), pušač (da/ne), nesanica (da/ne),....

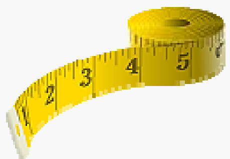
LJESTVICE MJERENJA

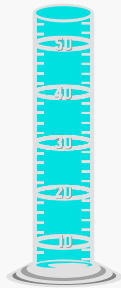


1. nominalna
2. ordinalna
3. intervalna
4. omjerna









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

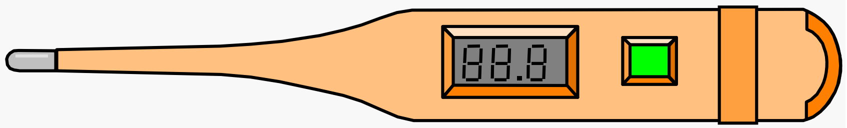
1 2 3

- ima sve karakteristike nominalne ljestvice i dodatno uključuje **redoslijed** (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**

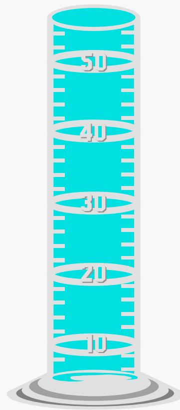
temperatura, kvocijent inteligencije

$100^0F \neq 2*50^0F$ jer je $100^0F=38^0C$

$50^0F=10^0C$

$$\left(\frac{t\,^0C}{100} = \frac{t\,^0F - 32}{180} \right)$$

4. Omjerna ljestvica



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

visina, težina, dob

90kg = 3*30kg

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

Još neke vrste podataka...

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

ejekcijska frakcija, ...

- <u>**omjeri**</u> (ratios)
  - omjer dva raznovrsna podatka (može biti i omjer dviju varijabli)

BMI

...još neke vrste podataka...

- **<u>stope</u>** (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.)

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

Apgar, GCS, ...

...još neke vrste podataka...

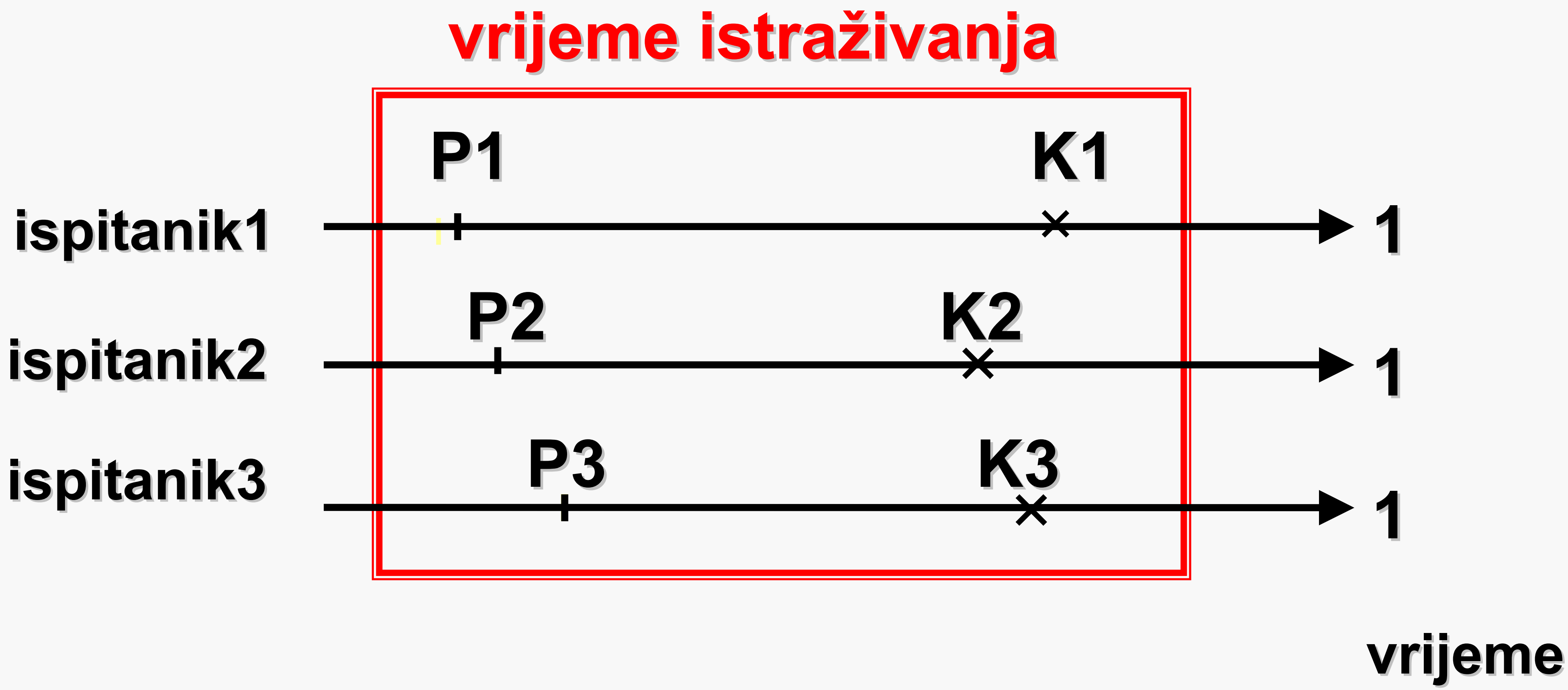
- **<u>cenzorirani</u>** (censored)
  - podatci koji se ne mogu točno izmjeriti, ali je poznato da prelaze neku granicu mjerenja
- najčešće se pojavljuju:
  - u laboratorijskim mjerenjima – vrijednosti ispod/iznad mogućnosti detekcije nekog uređaja
  - u studijama praćenja (follow-up):
    - očekivano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

Cenzuriranje

- događaj se ostvaruje = 1
- sve ostalo = 0 (cenzurirani podatci)

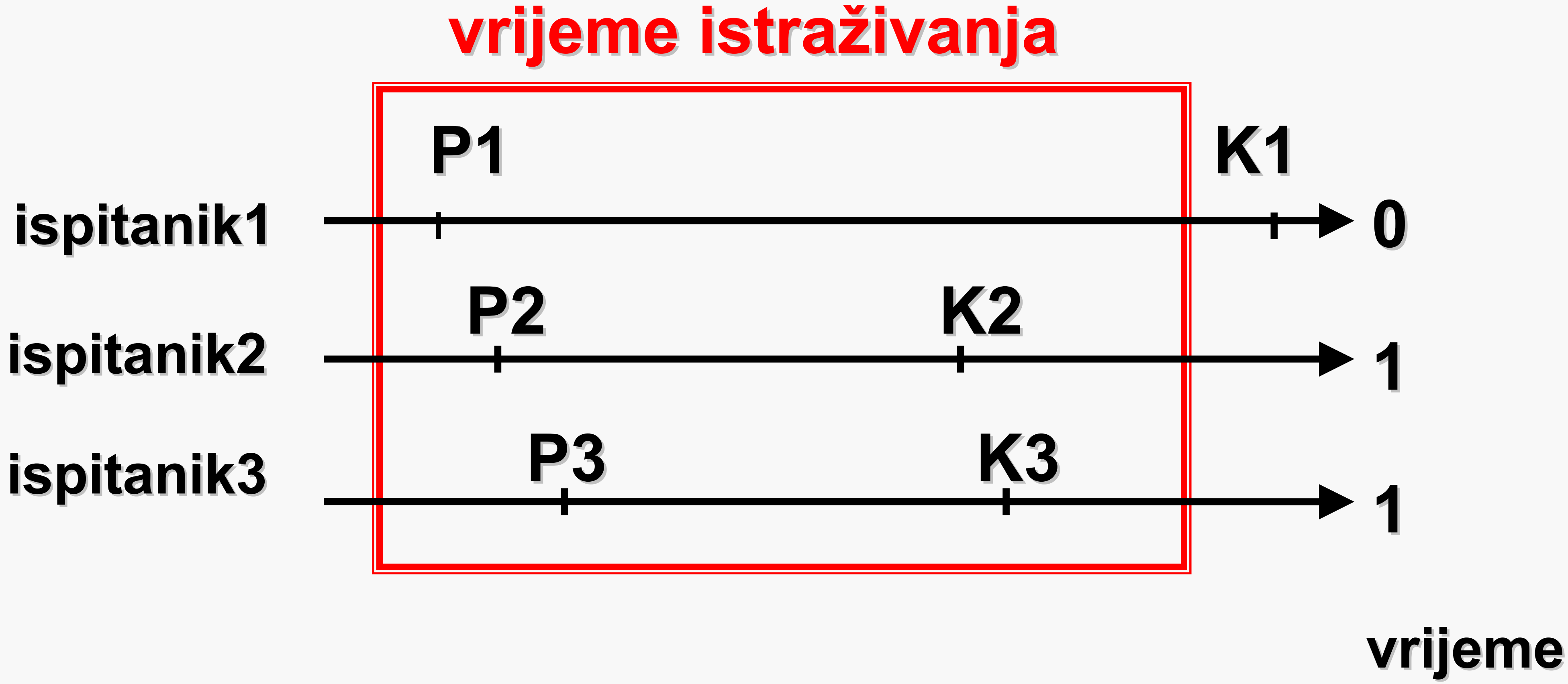
Cenzuriranje

- **potpuni podatci** (potpuno praćenje)
  - događaj se ostvario u promatranom vremenu istraživanja



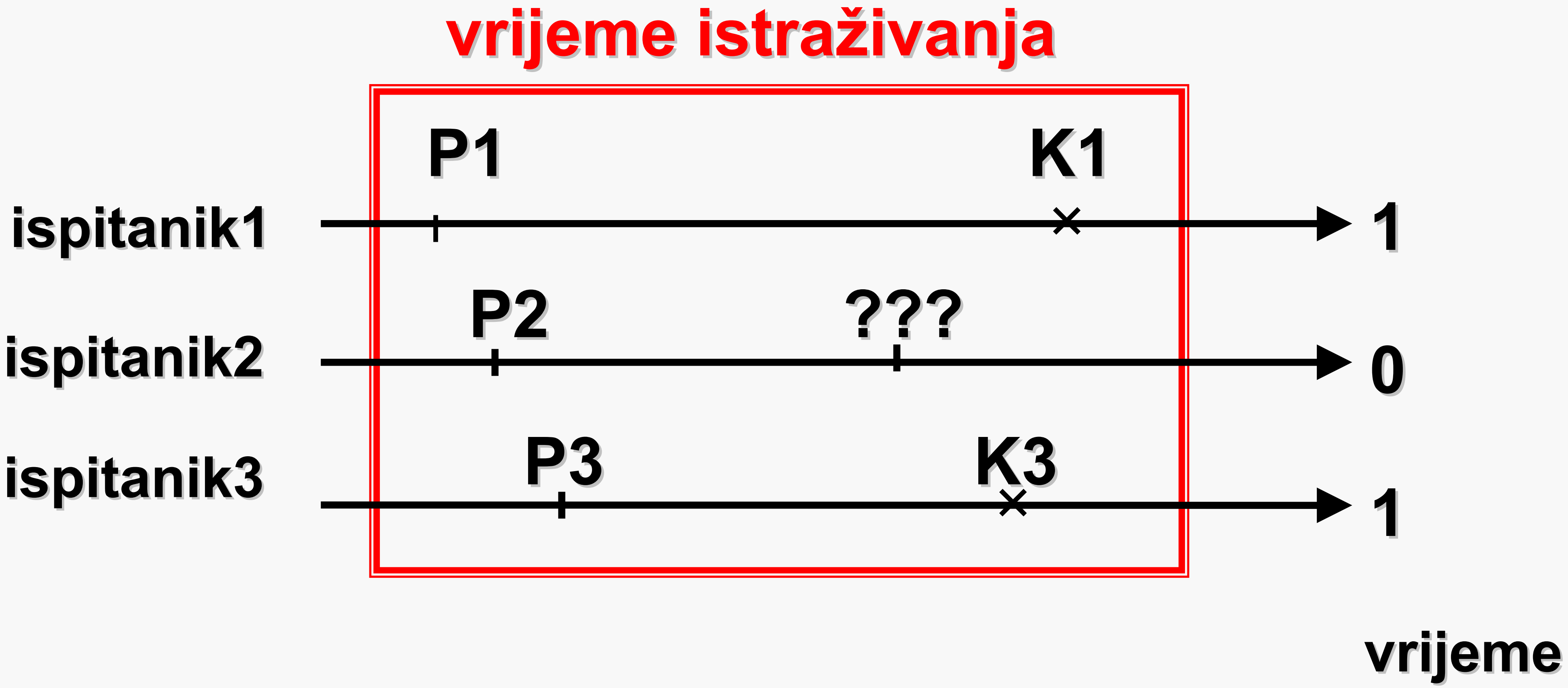
Cenzuriranje

- **cenzurirani podatci** (nepotpuni podatci)
  - za neke ispitanike događaj se ostvario IZVAN okvira promatranog vremena istraživanja



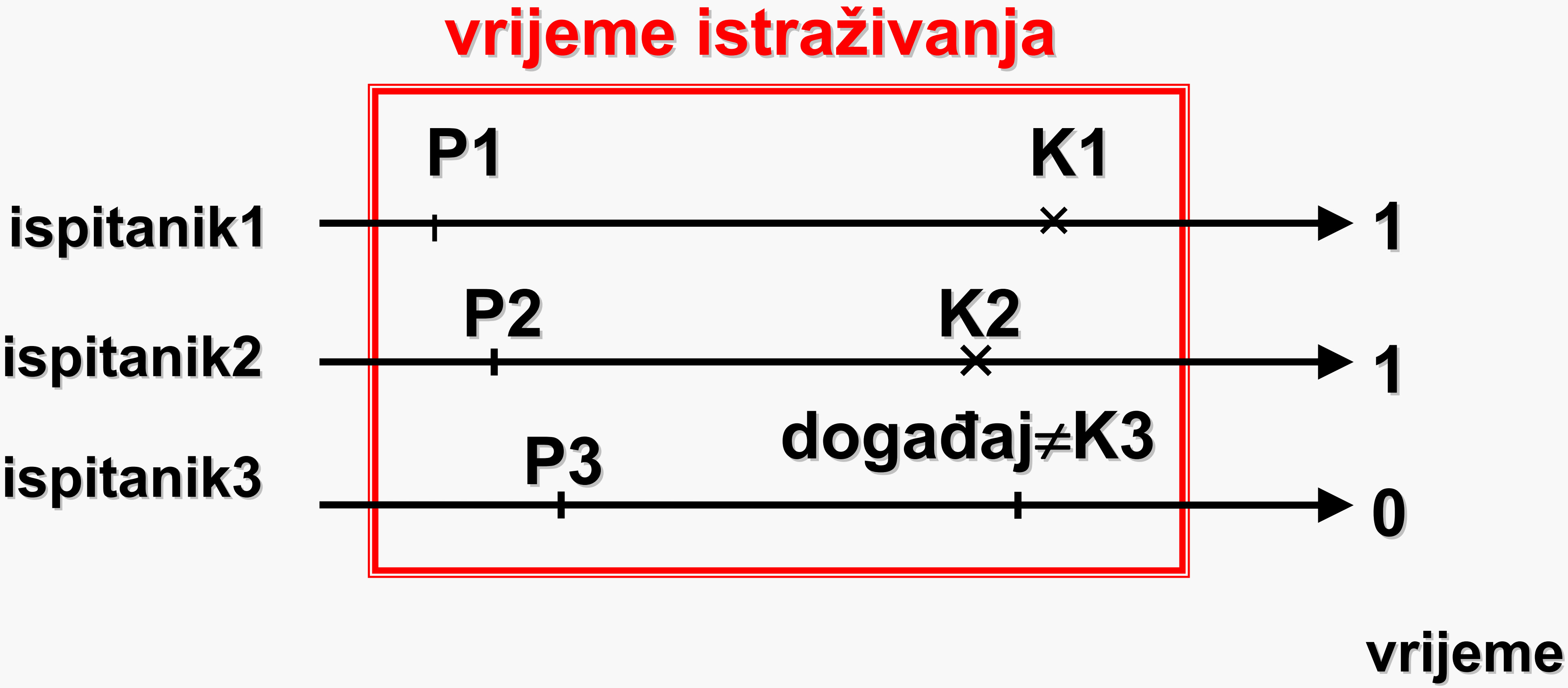
Cenzuriranje

- **cenzurirani podatci** (nepotpuni podatci)
  – izgubljeni ispitanici ("lost to follow-up")



Cenzuriranje

- **cenzurirani podatci** (nepotpuni podatci)
  - ispitanici koji iz nekih razloga ne mogu više biti praćeni (promjena mjesta boravka, smrt iz razloga nepovezanih s promatranim događajem)



Razlozi VIP tretmana cenzuriranih podataka - ocjena sredine -

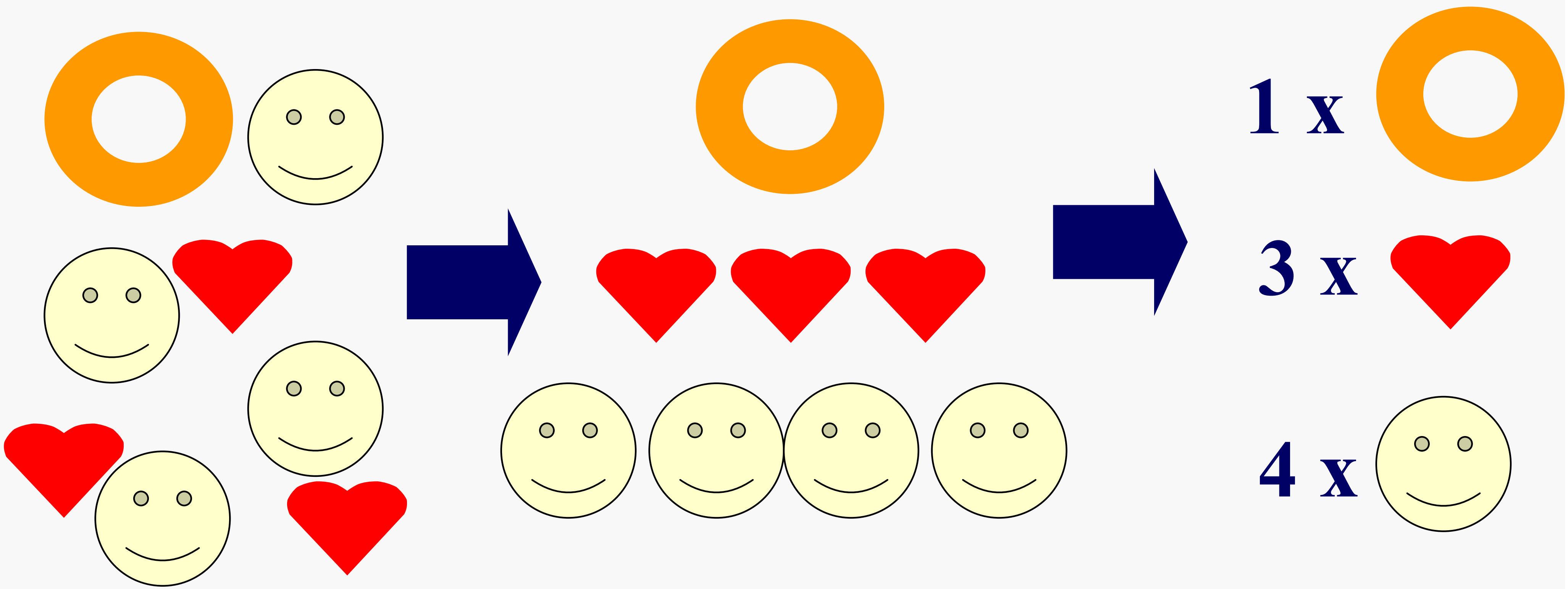
- **aritmetička sredina vremena preživljenja (od postavljanja dijagnoze) oboljelih od određene vrste karcinoma:**
  - srednje vrijeme preživljenja ovisi o tome KADA su podatci analizirani
  - nema smisla dok SVI ispitanici ne umru
- **medijan vremena preživljenja (od postavljanja dijagnoze) oboljelih od određene vrste karcinoma:**
  - može se procijeniti nakon što je POLOVINA ispitanika umrla
- **čekati????**

Razlozi VIP tretmana cenzuriranih podataka

- **čovjek-godine:**
  - svaki ispitanik doprinosi ukupnom vremenu preživljenja svojim vremenom provedenim u studiji (npr. 26 ispitanika promatrani su tijekom studije ukupno 336 mjeseci)
- **čovjek-godine**: suma vremena provedenih u studiji SVIH ispitanika svedena na godine:

  336/12 = 28 čovjek-godina
- **problem:** je li 1 čovjek praćen 28 godina ili 28 ljudi 1 godinu ili .... ???
- usporedba????

RAZDIOBA OBILJEŽJA



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:

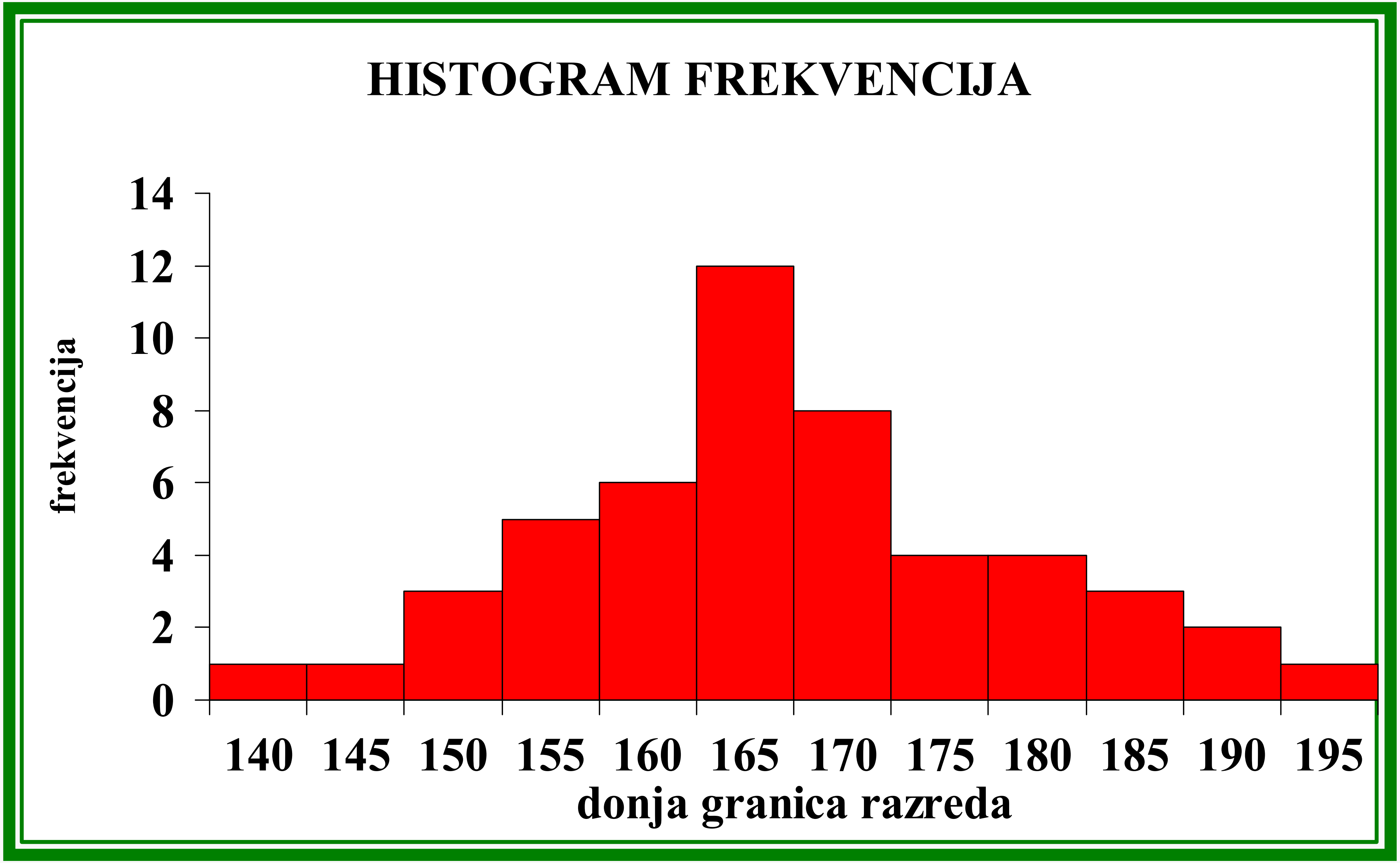
- apsolutna frekvencija:
  - koliko mjerenja ima vrijednosti iz intervala tog razreda
- apsolutna kumulativna frekvencija:
  - koliko mjerenja ima vrijednost manju ili jednaku gornjoj granici tog razreda
- relativna frekvencija:
  - koliki udio (postotak) mjerenja od ukupnog broja mjerenja ima vrijednost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

Tablica frekvencija podataka iz primjera

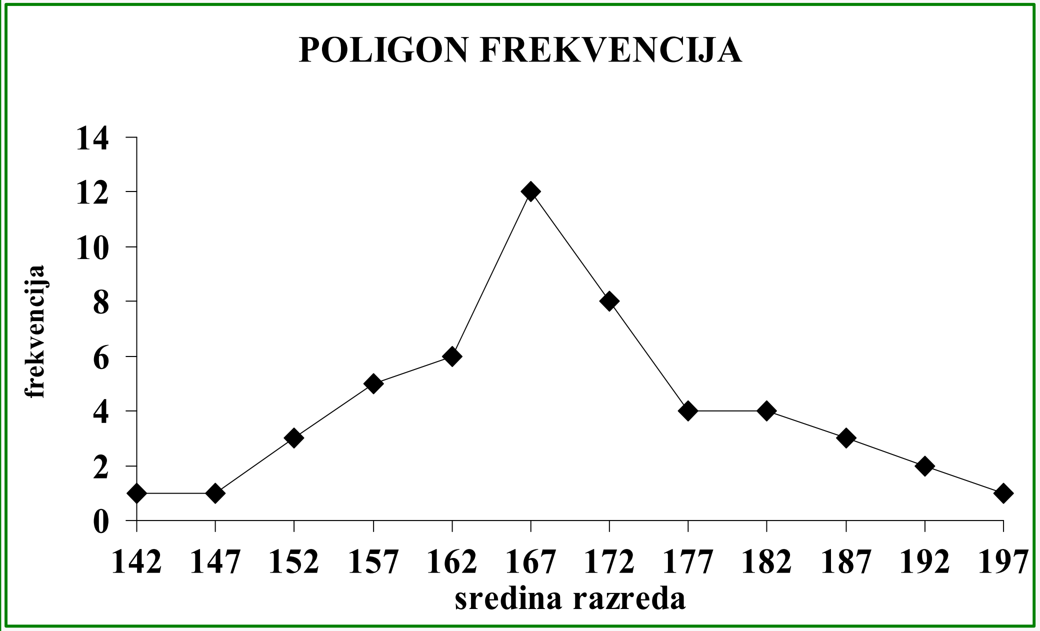
min = 140 max = 196

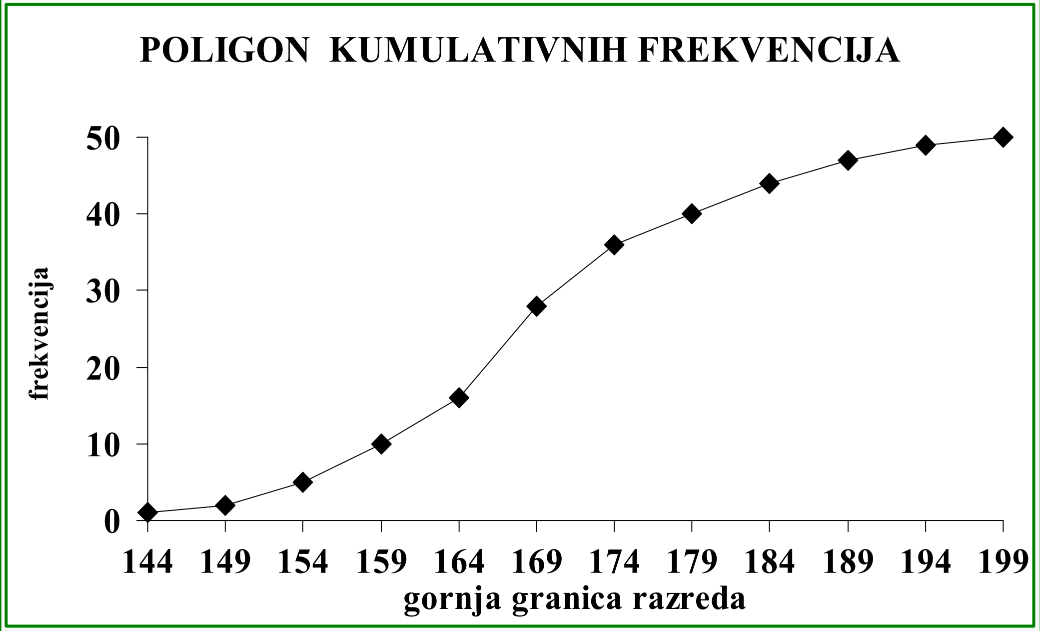
Broj r.	Granice r.	Sredina r.	f	cf	rf	crf
1	140 – 144	142	1	1	0,02	0,02
2	145 – 149	147	1	2	0,02	0,04
3	150 – 154	152	3	5	0,06	0,10
4	155 – 159	157	5	10	0,10	0,20
5	160 – 164	162	6	16	0,12	0,32
6	165 – 169	167	12	28	0,24	0,56
7	170 – 174	172	8	36	0,16	0,72
8	175 – 179	177	4	40	0,08	0,80
9	180 – 184	182	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	

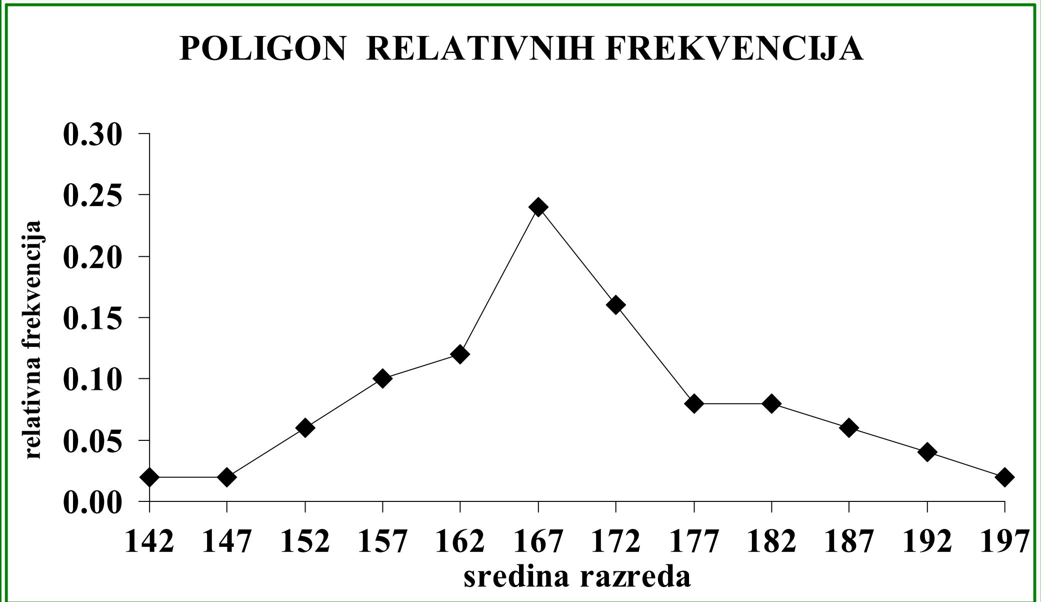
RAZDIOBA FREKVENCIJA

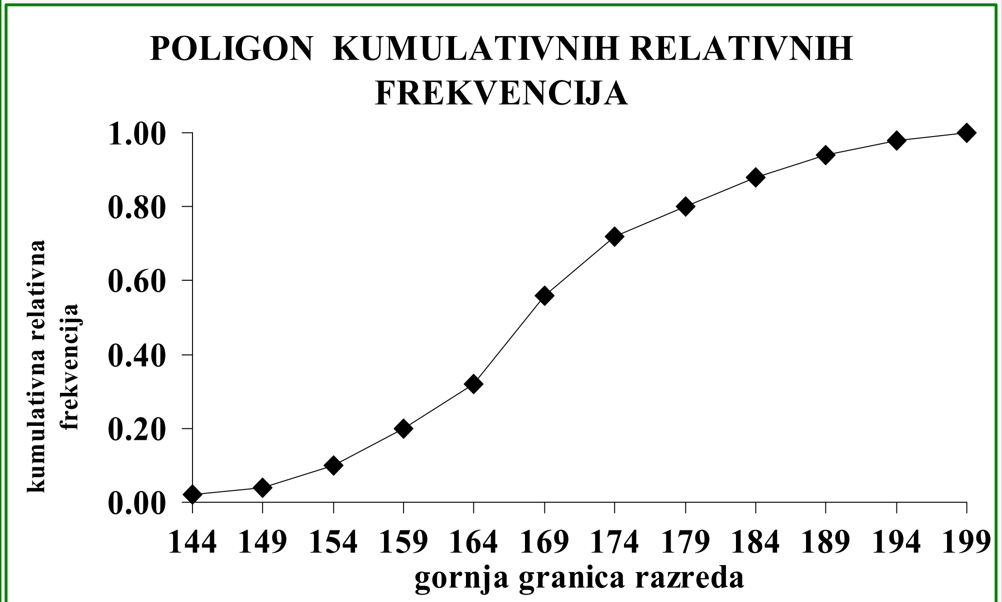


GRAFIČKI PRIKAZ RAZDIOBE









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

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!

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

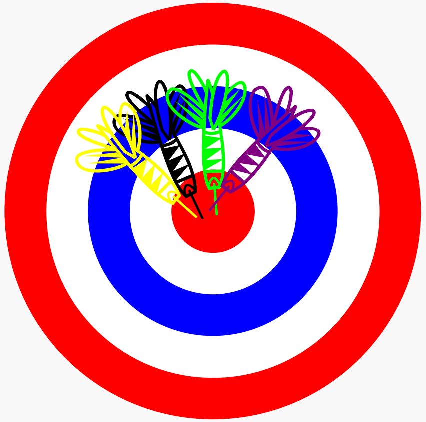
PARAMETAR I STATISTIKA

- 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

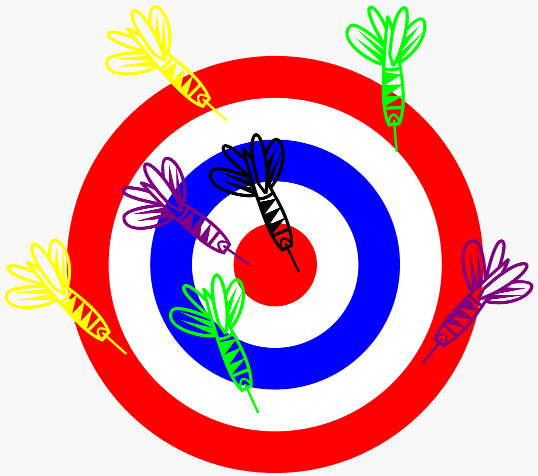
Uobičajene oznake:

	OCJENA PARAMETRA (STATISTIKA)	PARAMETAR POPULACIJE
ARITMETIČKA SREDINA	$\bar{x}$	μ
STANDARDNA DEVIJACIJA	s	σ
PROPORCIJA	p	π

OPISIVANJE RAZDIOBE PODATAKA



- sredina
- varijabilnost
- oblik



MJERE SREDINE (centralne tendencije)

(srednje vrijednosti, prosjeci, mjere lokacije)

ARITMETIČKA SREDINA

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

MEDIJAN (središnja vrijednost)

– vrijednost koja se u nizu podataka poredanih po veličini nalazi na srednjem mjestu

– dijeli podatke "na pola"

MOD (dominantna ili tipična vrijednost)

GEOMETRIJSKA SREDINA

$$\boxed{G = \sqrt[N]{x_1 \cdot x_2 \cdot \ldots \cdot x_N}}$$

HARMONIJSKA SREDINA

$$\boxed{H = \frac{N}{\sum_{i=1}^{N} \frac{1}{x_i}}}$$

ARITMETIČKA SREDINA

- oznake: $\overline{X}$ ..... uzorak μ..... populacija

ARITMETIČKA SREDINA INDIVIDUALNIH PODATAKA

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

x_i ... vrijednosti mjerenog 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?

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

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 GRUPIRANIH PODATAKA

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

f_i ... frekvencija i-tog razreda

x_{Si} ... sredina i-tog razreda

k ... broj razreda

N ... ukupan broj podataka

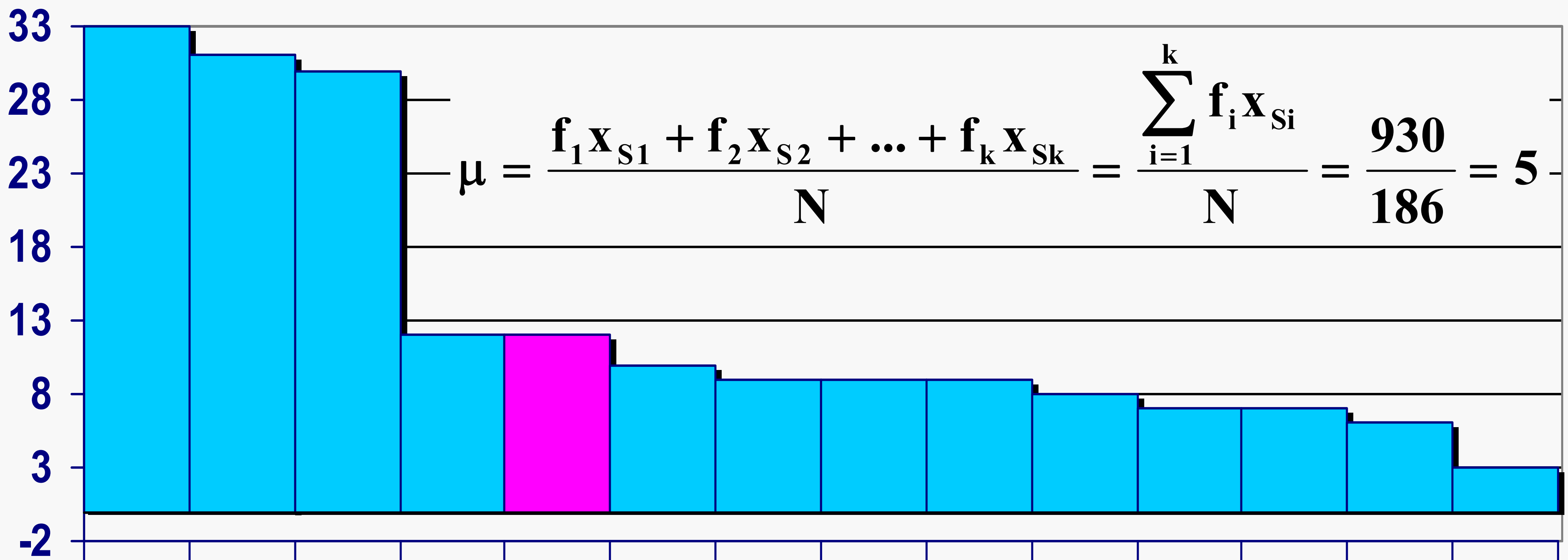
Kolika je aritmetička sredina podataka u sljedećoj tablici:

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

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

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

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



$$\mu = \frac{f_1 x_{S1} + f_2 x_{S2} + \ldots + f_k x_{Sk}}{N} = \frac{\sum_{i=1}^{k} f_i x_{Si}}{N} = \frac{930}{186} = 5$$

	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

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

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

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

x_i ... aritmetička sredina dobivena iz N_i mjerenja

n ... broj skupina 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
Σ	15	15	25

$$\boxed{\mathbf{uspjeh}_{S1} = \frac{2 \cdot 4 + 5 \cdot 9 + 5 \cdot 8 + 3 \cdot 4}{25} = \frac{105}{25} = 4.20}$$

$$\mathbf{uspjeh}_{S2} = \frac{5 \cdot 4 + 2 \cdot 9 + 3 \cdot 8 + 5 \cdot 4}{25} = \frac{82}{25} = 3.28$$

U dva navrata vršeno je mjerenje neke dužine i dobiveni su slijedeći rezultati:

$\overline{x}_1 = 20\text{cm} \quad ; N_1 = 15$

$\overline{x}_2 = 23\text{cm} \quad ; N_2 = 60$

a) Kolika je zajednička aritmetička sredina?

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

$$\text{a)} \quad \overline{\mathbf{x}}_{\text{zaj}} = \frac{20 \cdot 15 + 23 \cdot 60}{15 + 60} = \frac{300 + 1380}{75} = 22.4\,\text{cm}$$

$$\text{b)} \quad \overline{\mathbf{x}}_{\text{zaj}} = \frac{20 \cdot 60 + 23 \cdot 15}{15 + 60} = \frac{1200 + 354}{75} = 20.6\,\text{cm}$$

aritmetička sredina osjetljiva je **ne samo na <u>vrijednost</u> nego i na <u>broj</u> podataka**

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

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

Imamo niz podataka:

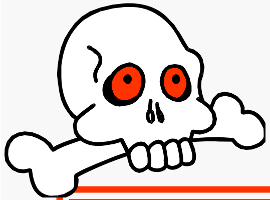
A: 2 2.5 3.5 3 4

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

SVOJSTVA ARITMETIČKE SREDINE



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

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

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

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. $\overline{X} = 3$

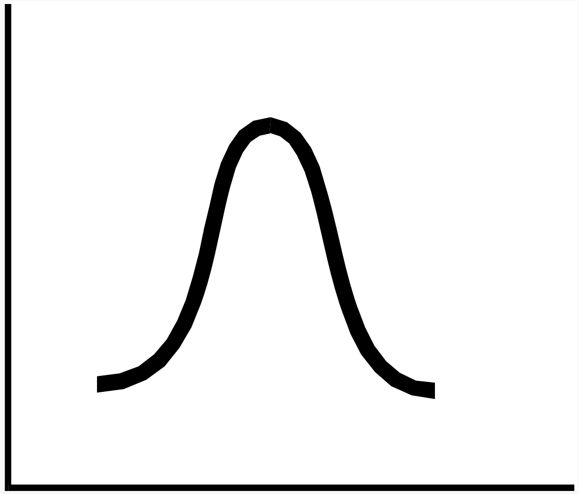
niz B: medijan.... **Me=1** arit. sred. $\overline{X} = 3$

MOD (dominantna vrijednost)

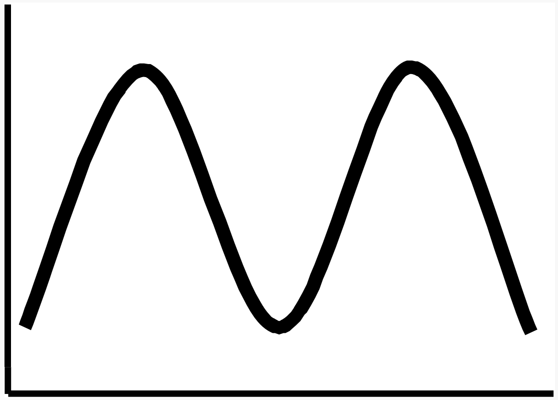
- vrijednost koja se u nizu mjerenja **najčešće** javlja (dominira svojom frekvencijom)

- na mod ne utječu **ni broj ni veličina** podataka, već **samo frekvencija**

  oznaka: Mo



UNIMODALNA



BIMODALNA

UNIFORMNA

PAŽNJA !

PARAMETRIJSKA MJERA SREDINE

NEPARAMETRIJSKE MJERE SREDINE

	Aritmetička sredina	Medijan	Mod
Ljestvica	Najmanje intervalna	Najmanje ordinalna	Sve ljestvice
Prisustvo ekstremnih vrijednosti	NIJE POGODNA!	Ne smetaju	Ne smetaju
Razdioba	Normalna ili najmanje simetrična	Nije važno	Unimodalna

MJERE VARIJABILNOSTI

(mjere disperzije, raspršenja)

RASPON - max-min

KVANTILE - mjere varijabilnosti po položaju

- najčešće: kvartile (Q_1 - 25 %, Q_3 - 75 %)

VARIJANCA

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

STANDARDNA DEVIJACIJA

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

KOEFICIJENT VARIJABILNOSTI

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

RASPON

$$\boxed{R = \max - \min}$$

- 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
prva	1%
druga	2%
treća	3%
....	

decila	obuhvat jedinica promatranja
prva	10%
druga	20%
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 svojstava neke grupe ispitanika**

  b) **razlike u varijabilnosti istog svojstva u različitim grupama ispitanika**

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)

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

NULTI I PRVI MOMENT

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

prvo svojstvo aritmetičke sredine

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

DRUGI MOMENT

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

VARIJANCA

TREĆI MOMENT

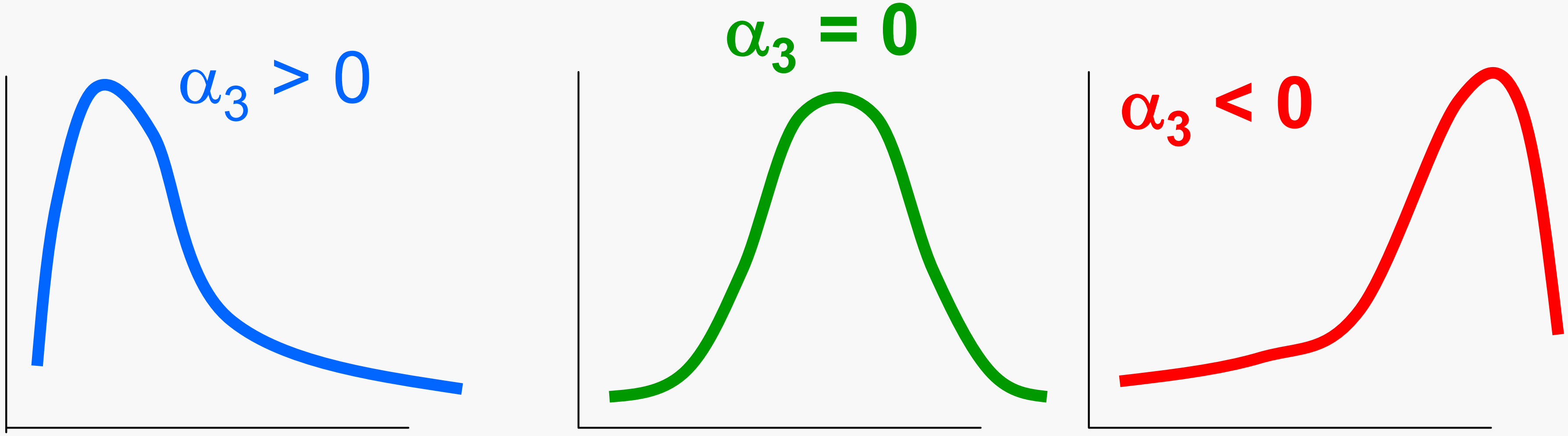
$$\mu_3 = \frac{\sum_{i=1}^{N} (x_i - \mu)^3}{N}$$

- za simetrične raspodjele $\mu_3 = 0$

KOEFICIJENT ASIMETRIJE (*coefficient of skewness*)

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

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



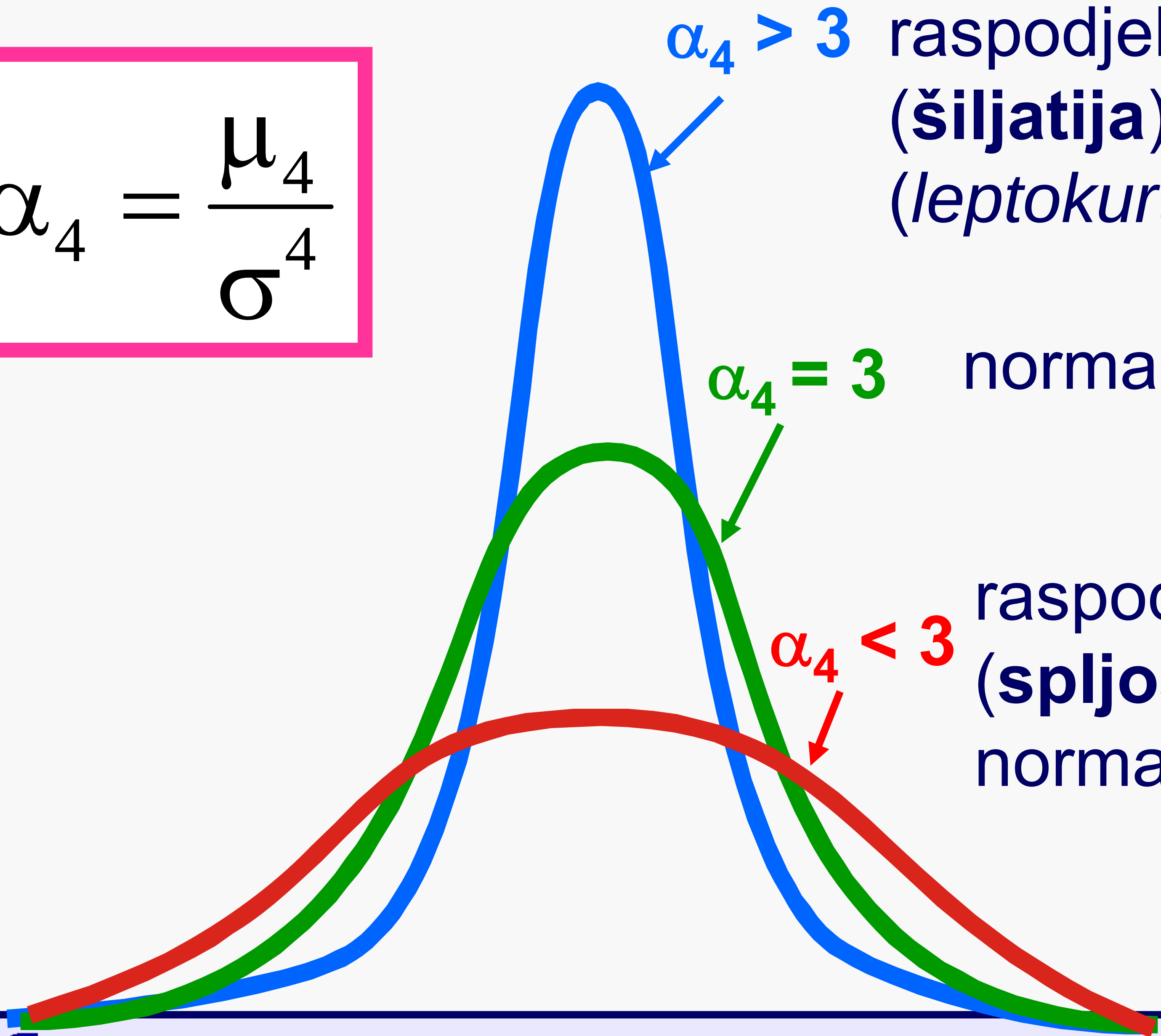
ČETVRTI MOMENT

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

- koristi se za mjeru spljoštenosti

KOEFICIJENT SPLJOŠTENOSTI (*coefficient of kurtosis*)

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



$\alpha_4 > 3$ raspodjela je **uža** (**šiljatija**) od normalne (*leptokurtic*)

$\alpha_4 = 3$ normalna spljoštenost

$\alpha_4 < 3$ raspodjela je **šira** (**spljoštenija**) od normalne (*platykurtic*)

KOEFICIJENT SPLJOŠTENOSTI (*coefficient of kurtosis*)

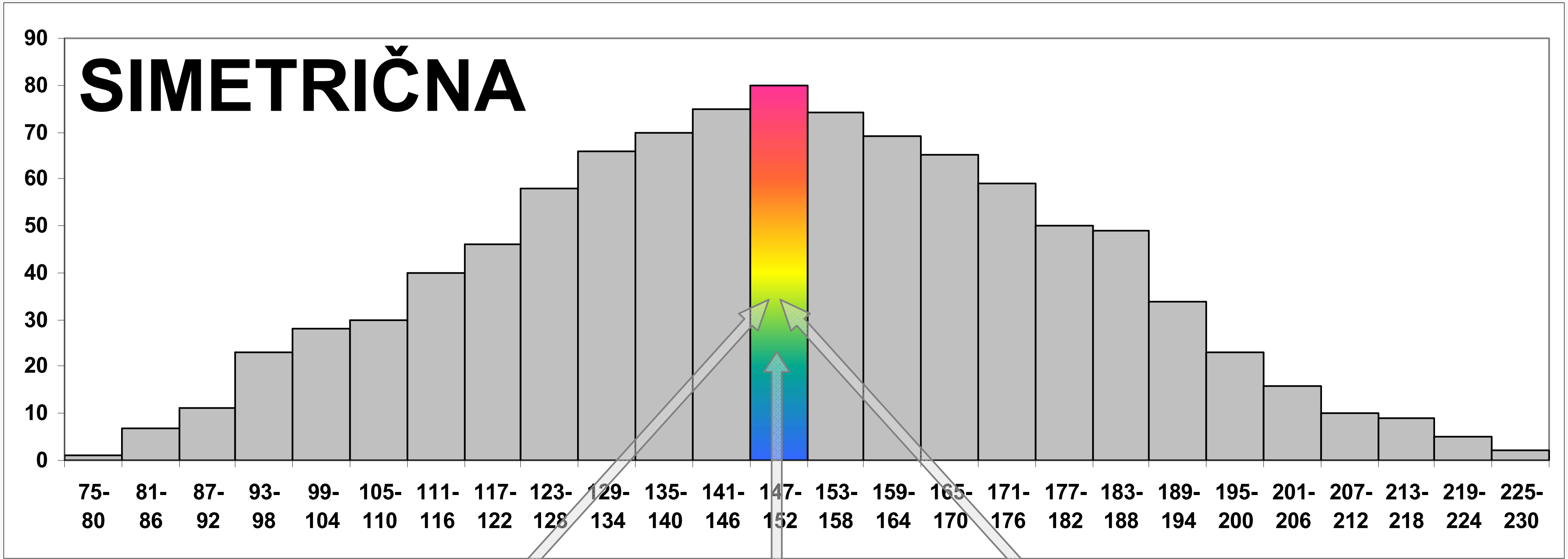
- statistički programi koeficijent spljoštenosti prikazuju kao **eksces spljoštenosti** (kurtosis excess)

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

- **za normalnu raspodjelu**

  eksces spljoštenosti = 0

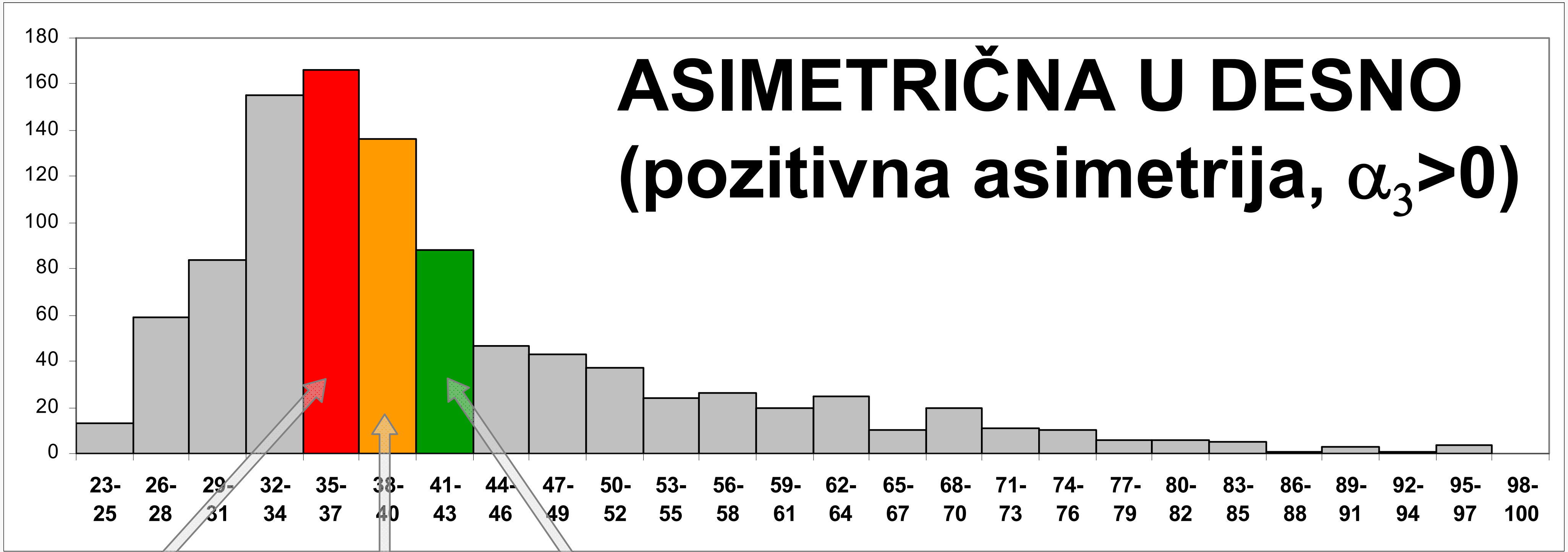
ODNOS MJERA SREDINE I ASIMETRIJE



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

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

ODNOS MJERA SREDINE I ASIMETRIJE



Mo = 37

$\overline{X} = 42$

Me = 38

$$\boxed{Mo < Me < \overline{x}}$$

ODNOS MJERA SREDINE I ASIMETRIJE

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

MJERE SREDINE PREMA LJESTVICI MJERENJA

Ljestvica mjerenja	MOD	MEDIJAN	ARITMETIČKA SREDINA
NOMINALNA	✓	✗	✗
ORDINALNA	✓	✓	✗
INTERVALNA/OMJERNA (asimetrična raspodjela)	✓	✓	✗
INTERVALNA/OMJERNA (simetrična raspodjela)	✓	✓	✓

MJERE VARIJABILNOSTI PREMA LJESTVICI MJERENJA

Ljestvica mjerenja	RASPON	KVARTILE	VARIJANCA STANDARDNA DEVIJACIJA KOEFICIJENT VARIJABILNOSTI
NOMINALNA	✗	✗	✗
ORDINALNA	✓	✓	✗
INTERVALNA/OMJERNA (asimetrična raspodjela)	✓	✓	✗
INTERVALNA/OMJERNA (simetrična raspodjela)	✓	✓	✓

MJERE VARIJABILNOSTI

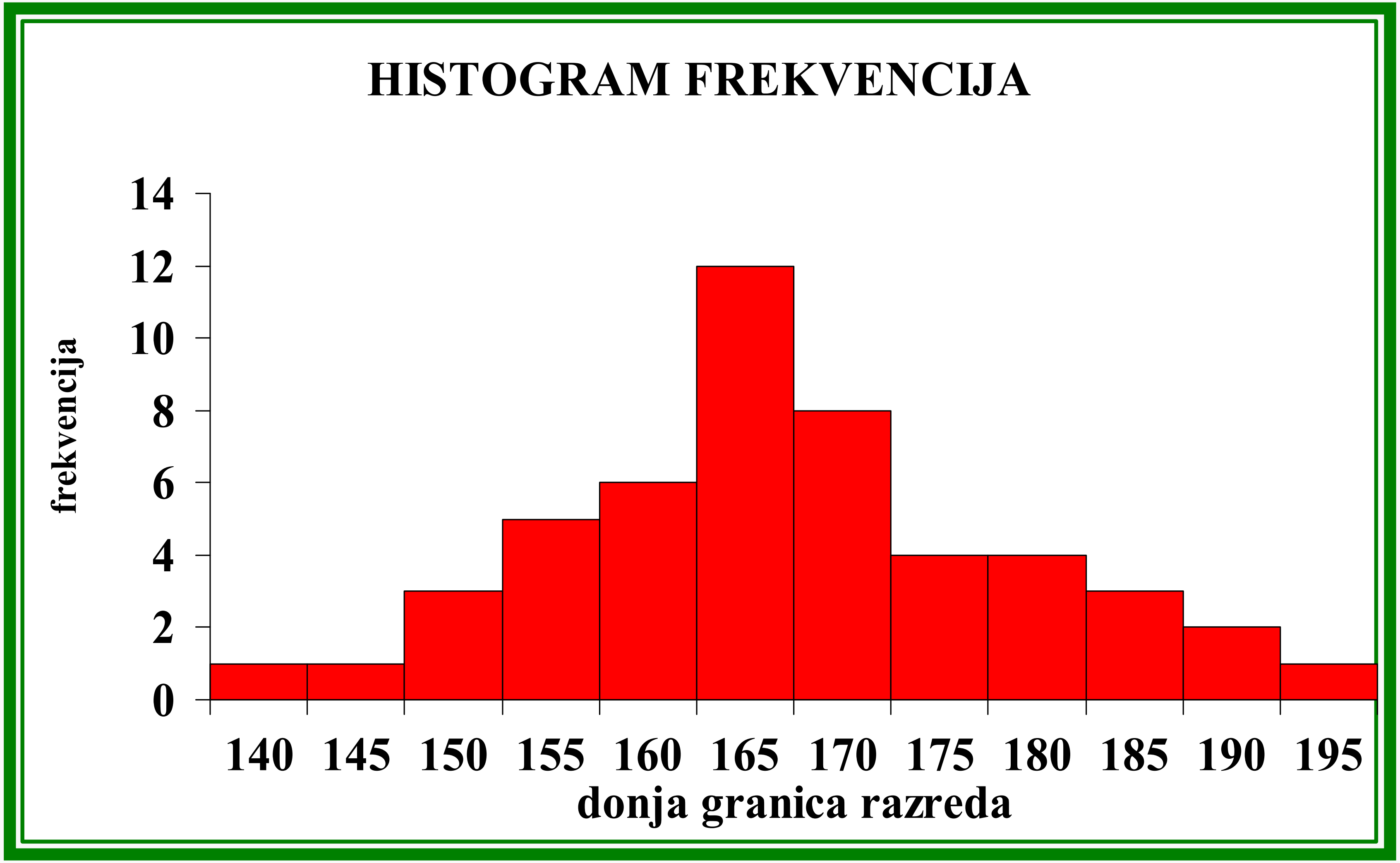
- iskazuju se uz odgovarajuću mjeru sredine

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

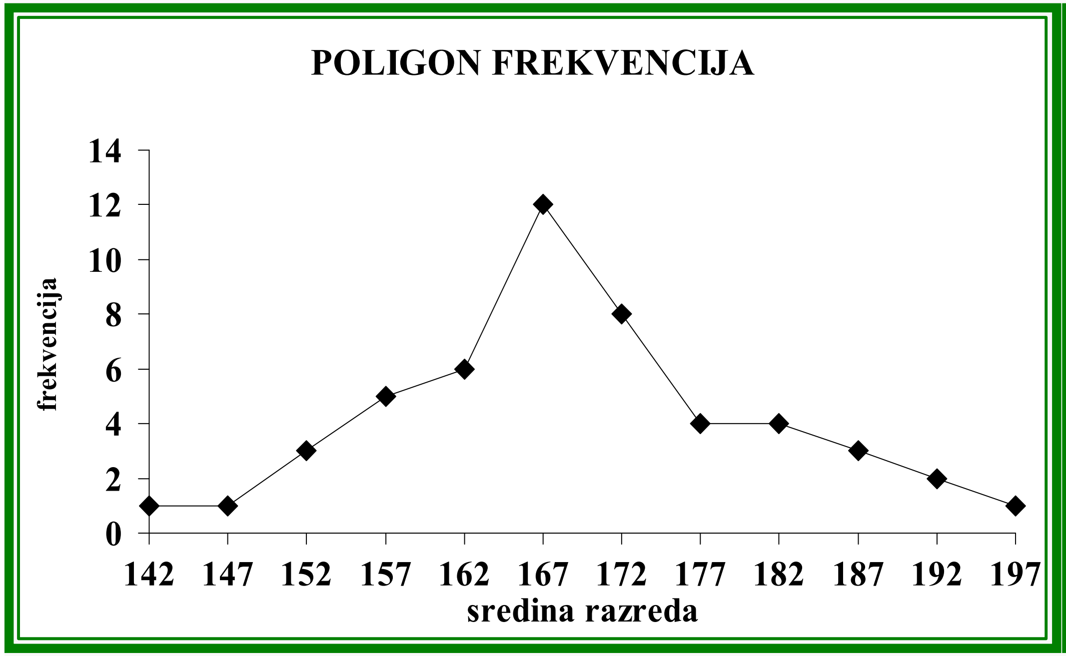
GRAFIČKI PRIKAZ

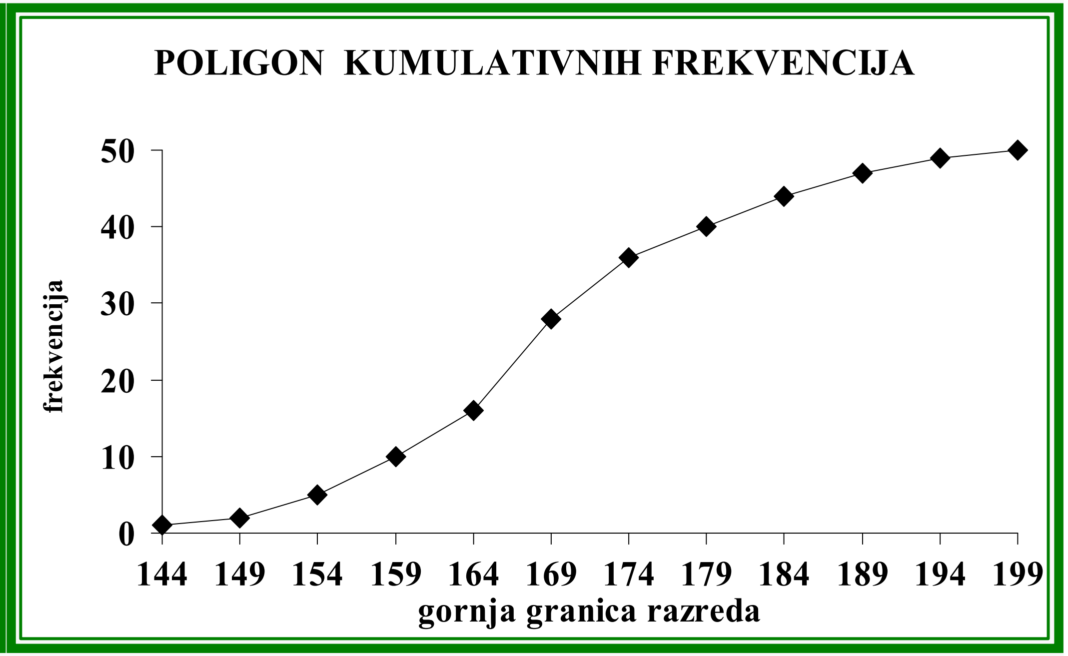
RAZDIOBA OBILJEŽJA

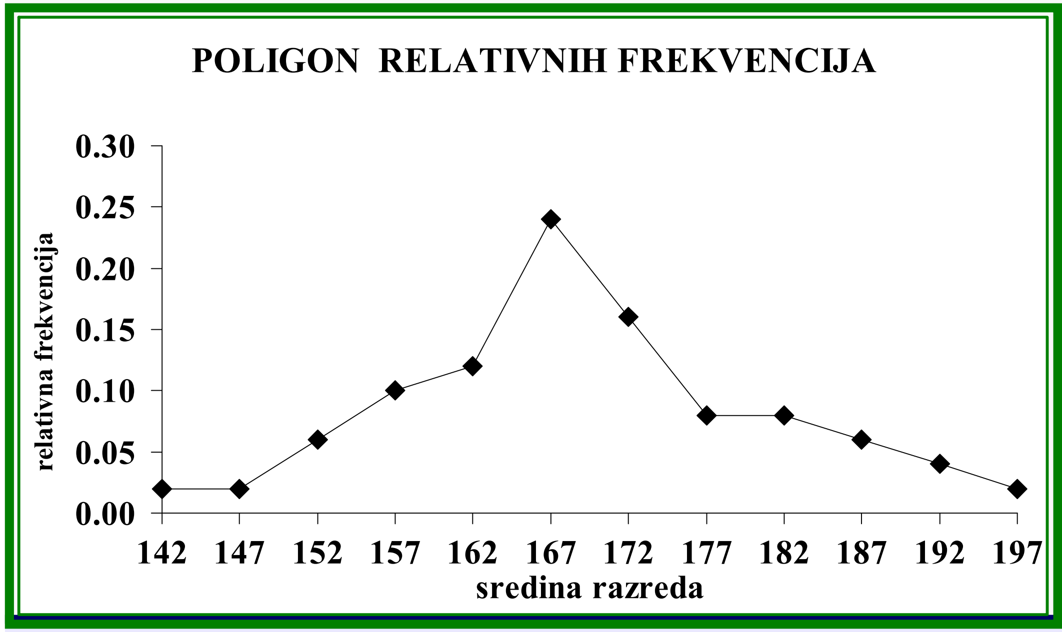
- prikazuje se histogramom (stupičasti grafikon, "*column chart*") ili poligonom frekvencija (linijski grafikon, "*line chart*")

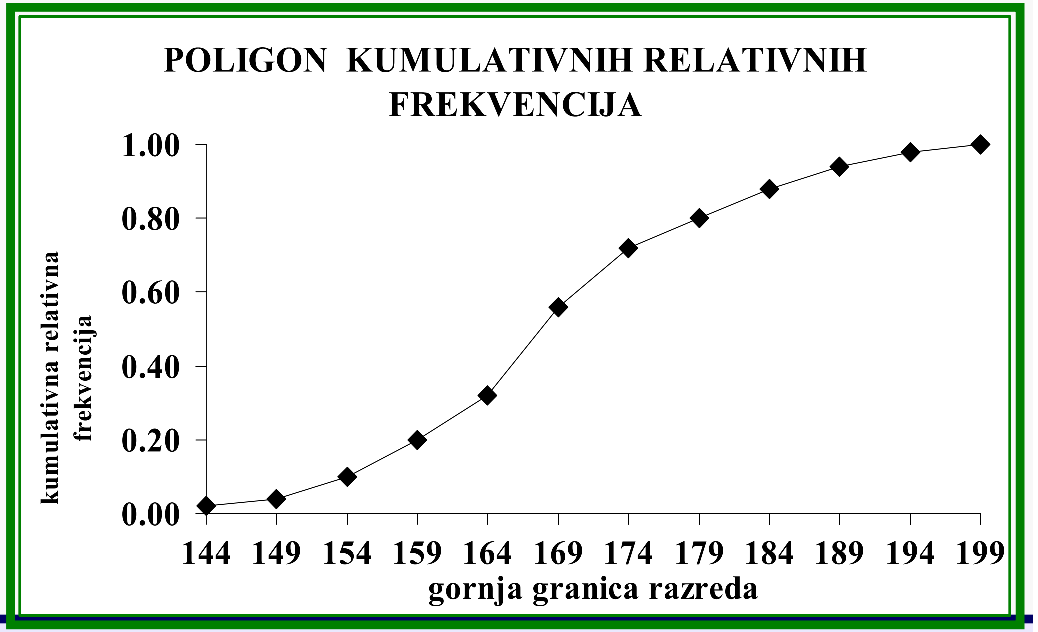


RAZDIOBA OBILJEŽJA



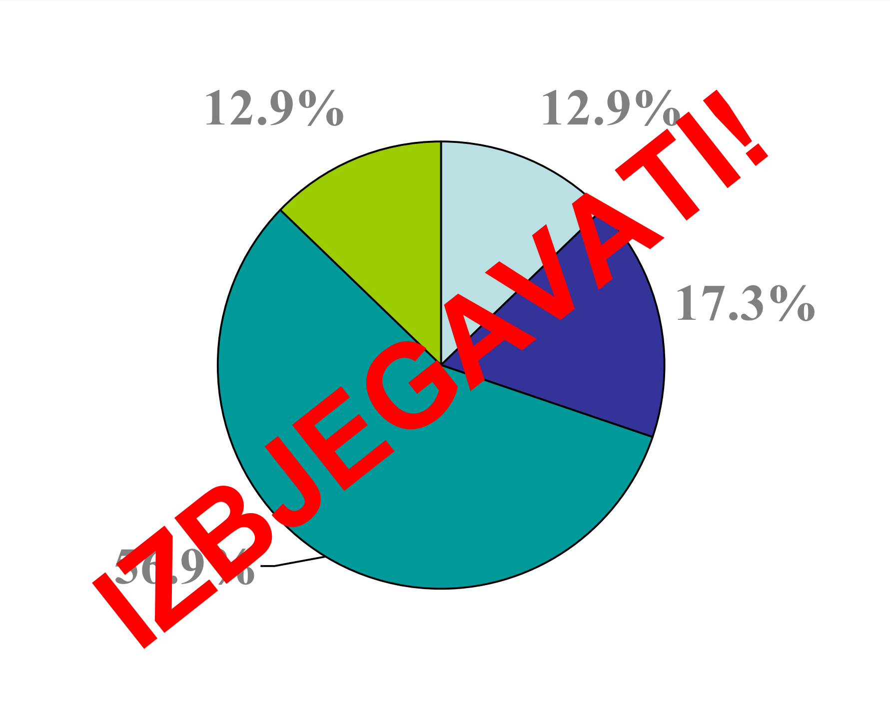


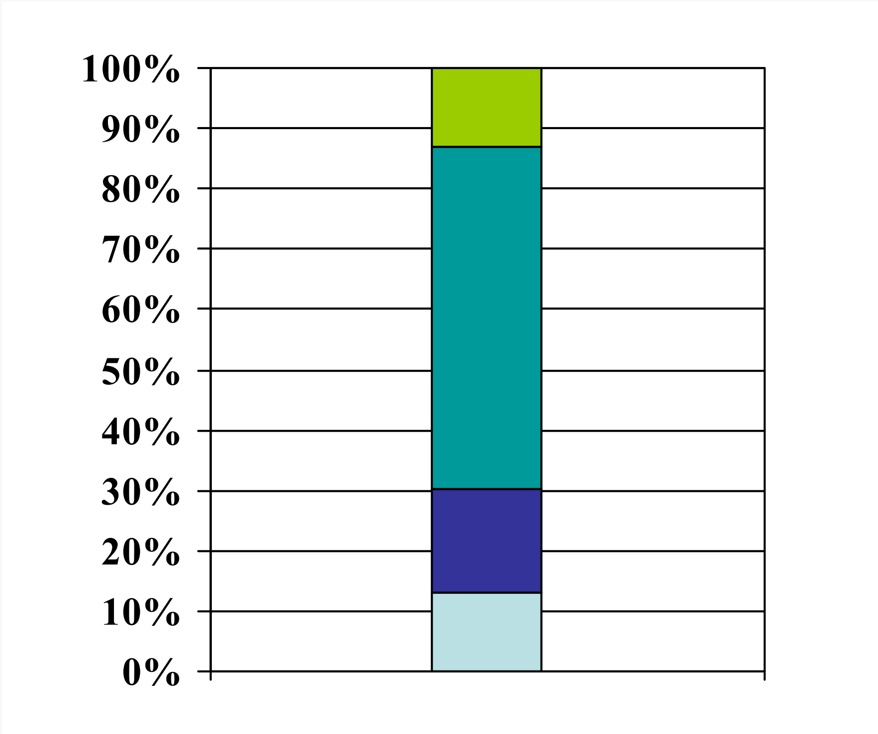




STRUKTURA OBILJEŽJA

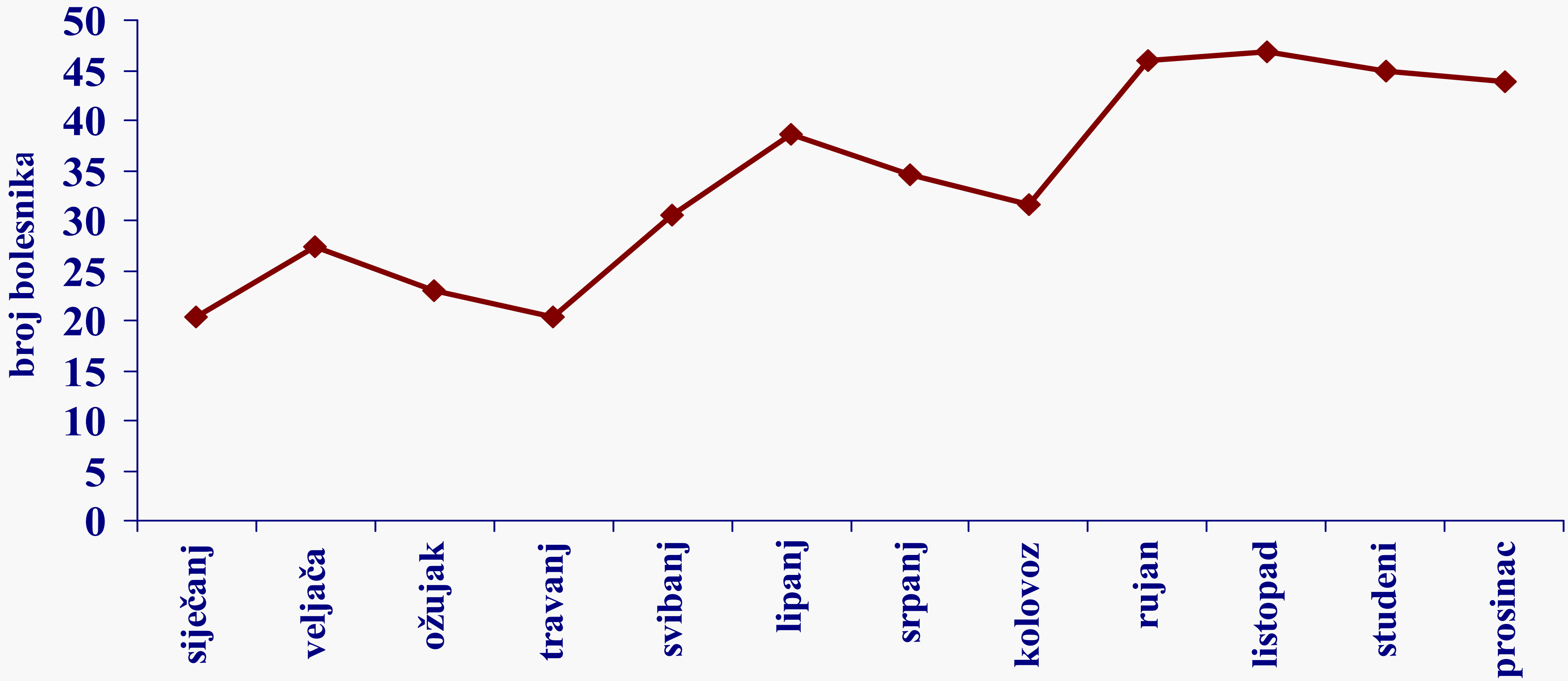
- **pokazuje udio pojedinih kategorija u ukupnom broju promatranja/mjerenja**
- **prikazuje se kružnim grafikonom ("torta", "*pie chart*") ili složenim stupičastim grafikonom ("*100% stacked column chart*")**





PROMJENE U VREMENU

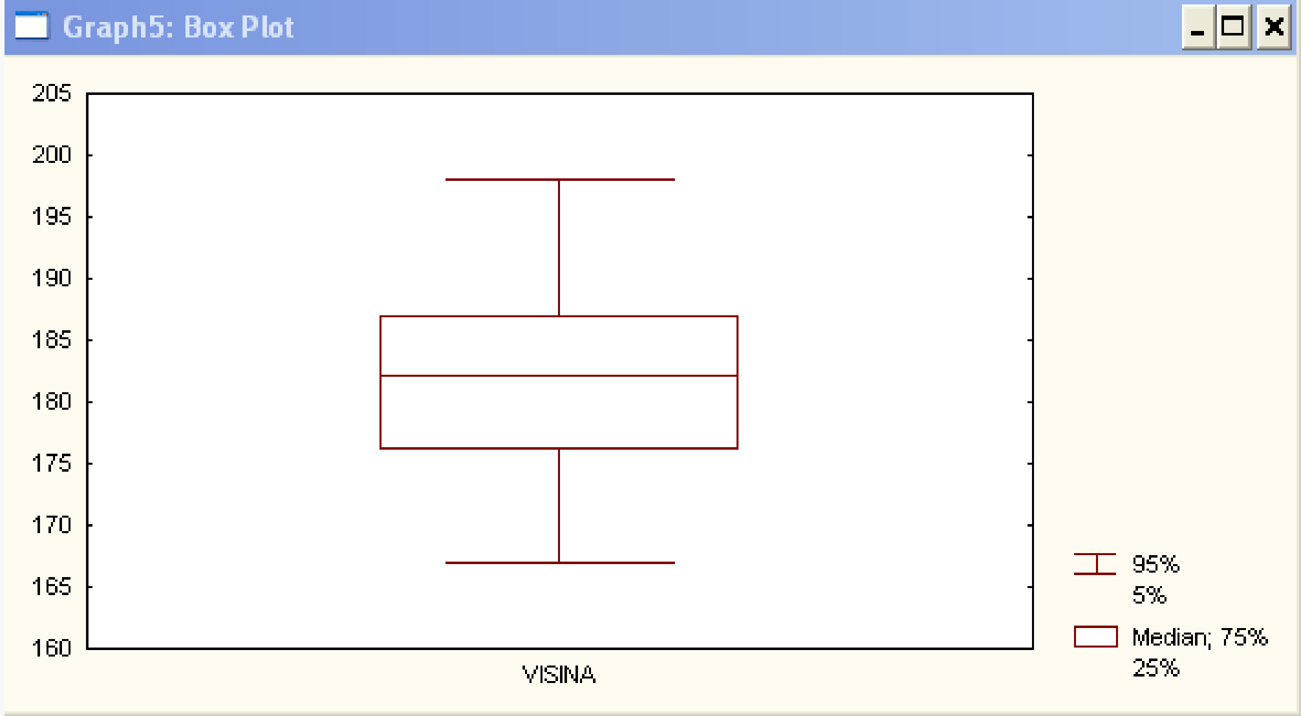
- **prikazuju se linijskim grafikonom (*“line chart”*)**
- **na apscisu se nanose vremenski intervali, a na ordinatu vrijednost promatrane varijable**

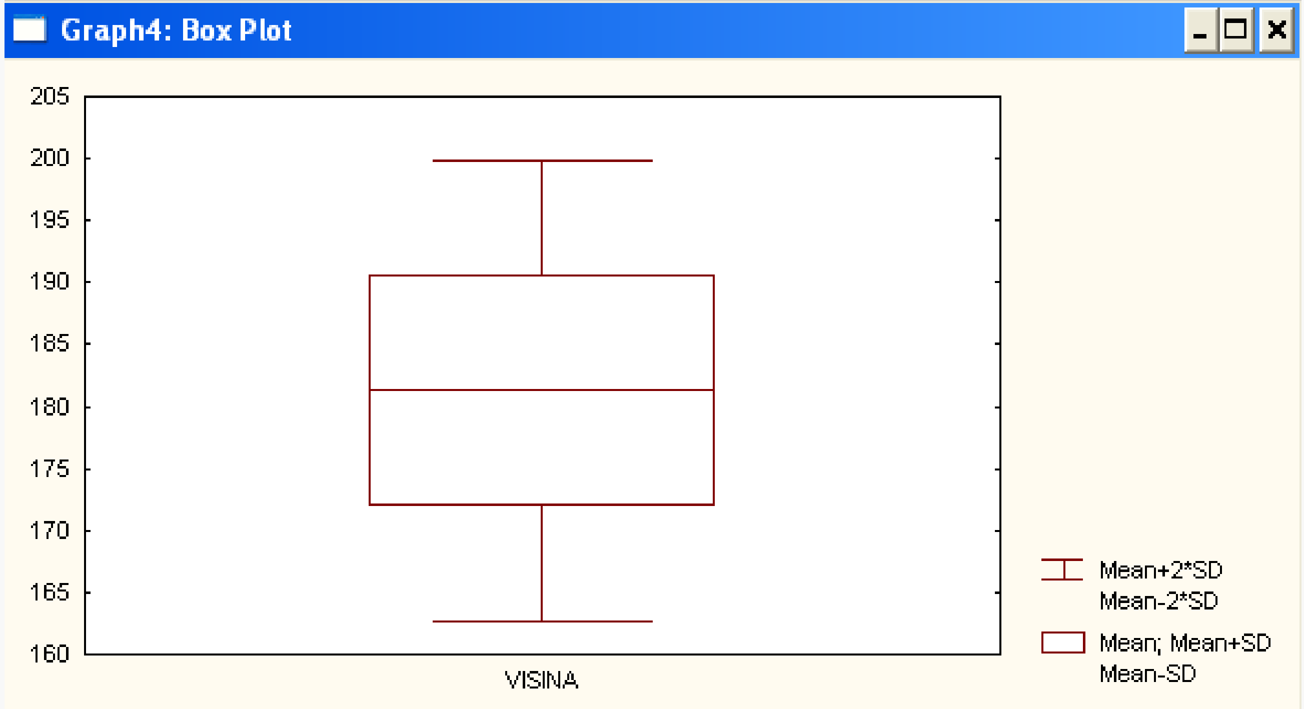


OSNOVNE MJERE SREDINE I RASPRŠENJA

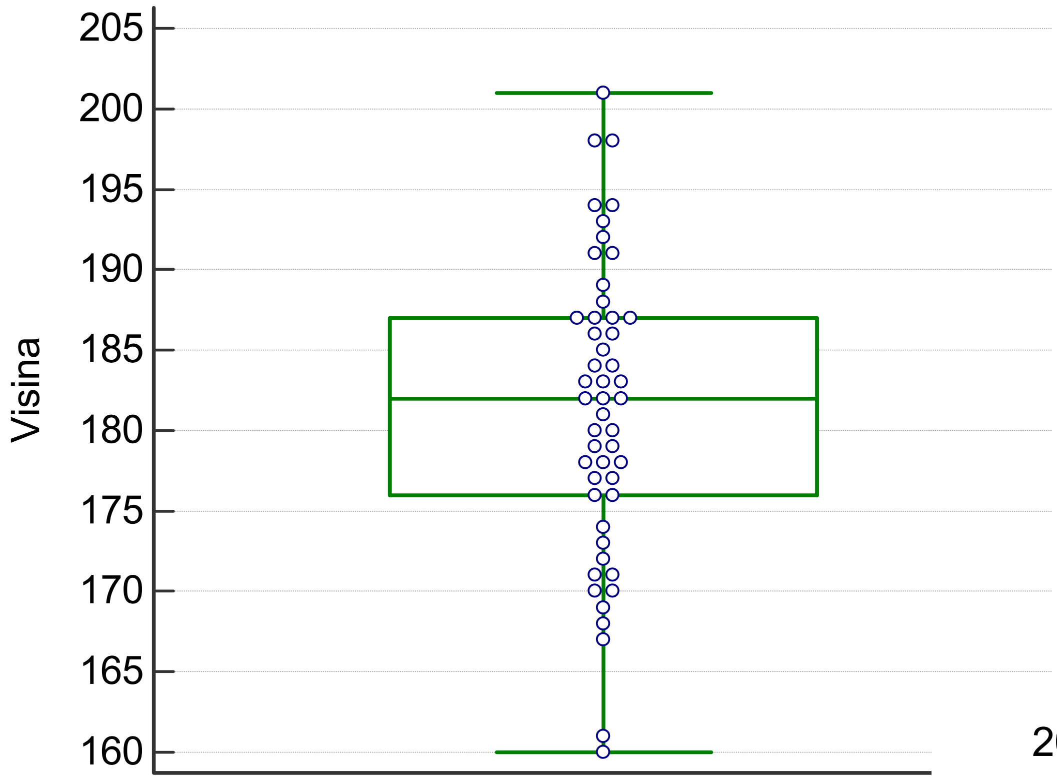
- “kutija i brkovi” grafikon (“Box-and-Whisker” plot)
- najčešće prikazuje kombinacije:

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

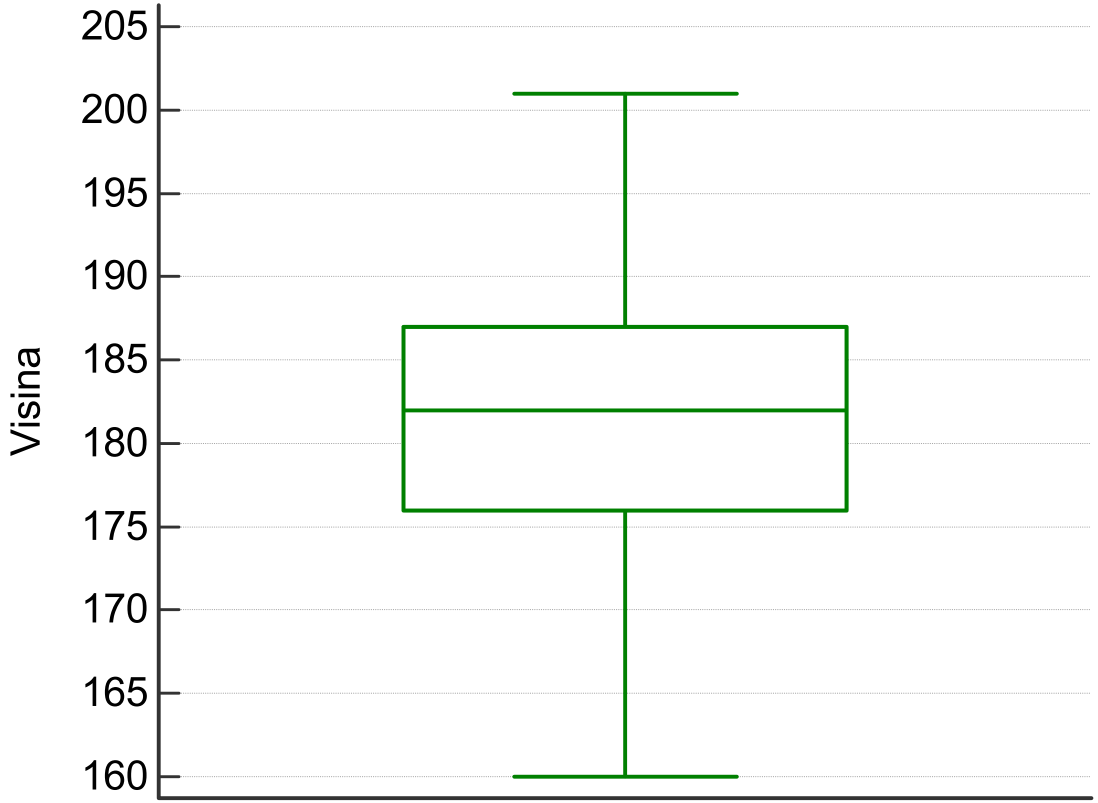

Graph5: Box Plot
205
200
195
190
185
180
175
170
165
160
VISINA
95%
5%
Median; 75%
25%


Graph4: Box Plot
205
200
195
190
185
180
175
170
165
160
VISINA
Mean+2*SD
Mean-2*SD
Mean; Mean+SD
Mean-SD

Statistica

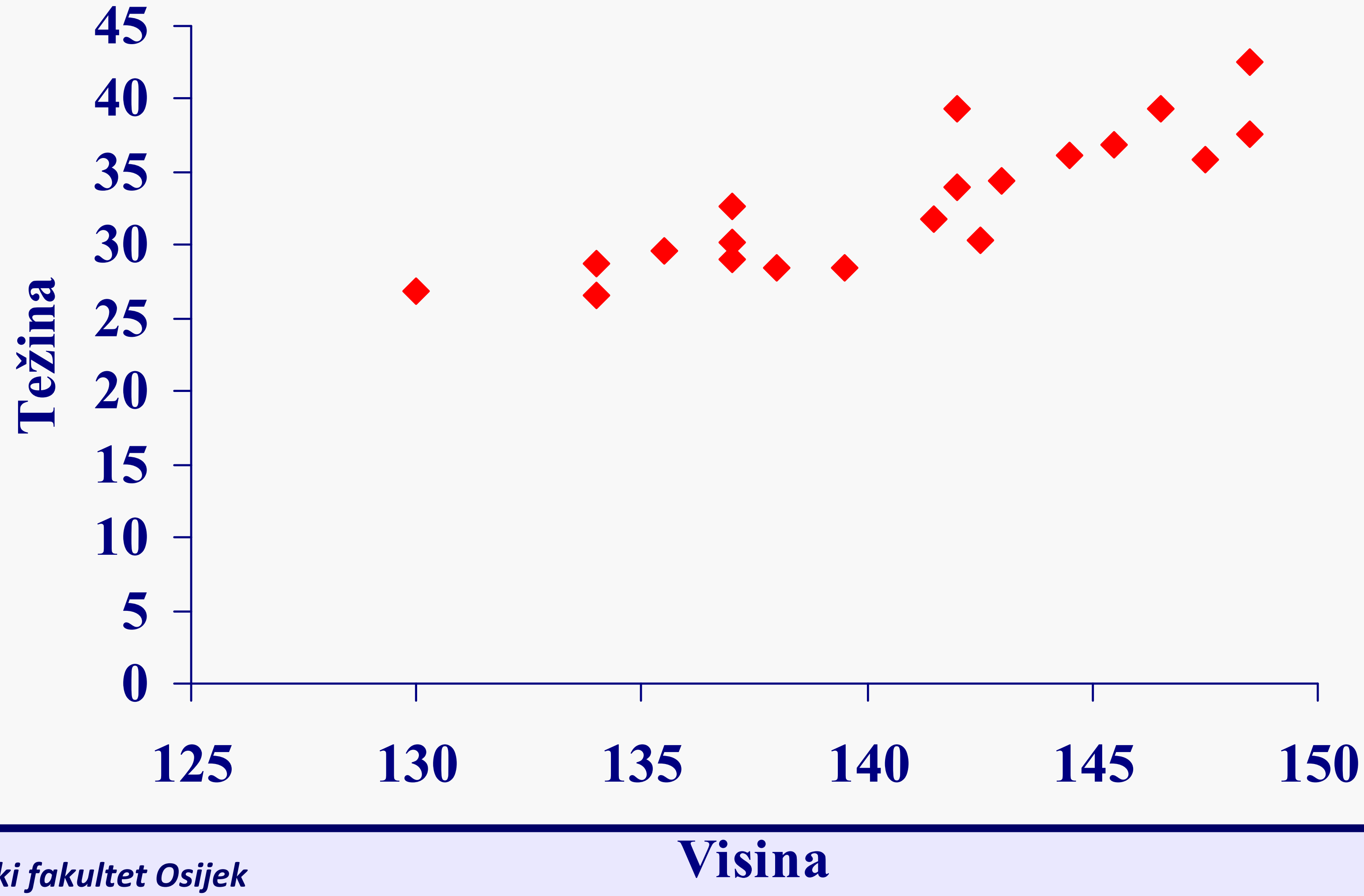

205
200
195
190
185
180
175
170
165
160
Visina

MedCalc


205
200
195
190
185
180
175
170
165
160
Visina

POVEZANOST DVIJU VARIJABLI

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



VJEROJATNOST

OSNOVNI POJMOVI TEORIJE VJEROJATNOSTI

- **teorija vjerojatnosti**
  - **matematička teorija slučajnih događaja**

- **SLUČAJNI DOGAĐAJ**
  - **događaj koji ne mora bezuvjetno nastupiti** (pod određenim okolnostima se može, ali i ne mora dogoditi)

POKUS

- **proces dobivanja rezultata opažanja**

- bacanje dva novčića
- određivanje krvne grupe 100 bolesnika i opažanje rezultata
- bacanje dvije igraće kocke
- prebrojavanje bolesnika koji uzimaju terapiju snižavanja lipida u krvi

PROSTOR ISHODA

– skup svih mogućih ishoda nekog pokusa

Prostor ishoda za pokus bacanja dva novčića

(P - pismo, G - glava)

prostor ishoda = { GG, GP, PG, PP}

Prostor ishoda za izvlačenje dvije igraće karte (s obzirom na boju)

prostor ishoda = {▤ ▤, ▤ ▤, ▤ ▤, ▤ ▤}

PROSTOR ISHODA

Prostor ishoda za pokus bacanja dvije igraće kocke

prostor ishoda= {(1,1) (1,2) (1,3) (1,4) (1,5) (1,6)
(2,1) (2,2) (2,3) (2,4) (2,5) (2,6)
(3,1) (3,2) (3,3) (3,4) (3,5) (3,6)
(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)
(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)
(6,1) (6,2) (6,3) (6,4) (6,5) (6,6) }

PROSTOR ISHODA

Prostor ishoda za pokus određivanja krvne grupe u 100 bolesnika

Krvna grupa	Broj bolesnika
O	42
A	43
B	11
AB	4
Ukupno	100

$$\text{prostor ishoda} = \{\underbrace{O, O, .. O}_{42 \times O}, \underbrace{A, A, .. A}_{43 \times A}, \underbrace{B, B, .., B}_{11 \times B}, \underbrace{AB, AB, AB, AB}_{4 \times AB}\}$$

DOGAĐAJ

podskup prostora ishoda, tj. skup ishoda

Pokus: bacanje dva novčića

Događaj: pojava dvije "glave"

(P - pismo, G - glava)

prostor ishoda = { **GG**, GP, PG, PP}

ishod koji realizira događaj

DOGAĐAJ

Pokus: bacanje dvije igraće kocke

Događaj: pojava barem jedne šestice

prostor ishoda= {							
	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	**(1,6)**	
	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	**(2,6)**	
	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	**(3,6)**	
	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	**(4,6)**	
	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	**(5,6)**	
	(6,1)	**(6,2)**	**(6,3)**	**(6,4)**	**(6,5)**	**(6,6)**	}

ishodi koji realiziraju događaj

DOGAĐAJ

Pokus: određivanje krvne grupe u 100 bolesnika

Događaj: krvna grupa je AB

Krvna grupa	Broj bolesnika
O	42
A	43
B	11
AB	<u>4</u>
Ukupno	100

ishodi koji realiziraju događaj

$$\text{prostor ishoda} = \{\underbrace{O, O, ..O}_{42 \times O}, \underbrace{A, A, ..A}_{43 \times A}, \underbrace{B, B, .., B}_{11 \times B}, \underbrace{\underline{AB, AB, AB, AB}}_{4 \times AB}\}$$

KLASIČNA DEFINICIJA VJEROJATNOSTI (vjerojatnost A PRIORI)

ELEMENTARNI DOGAĐAJ

svaki od n jednako mogućih **ishoda nekog događaja**

prostor elementarnih događaja

skup svih elementarnih događaja

$$\boxed{P(D) = \frac{m(D)}{n}}$$

vjerojatnost događaja D

m(D).... broj povoljnih ishoda (događaja koji realiziraju događaj D)

n.......... broj svih ishoda (kardinalni broj prostora elementarnih događaja)

PRIMJER

pokus: bacanje dva novčića

D ....oba novčića su pala na glavu

prostor ishoda = { **GG**, GP, PG, PP}

- taj događaj realizira 1 od 4 jednako moguća događaja

$$\mathbf{P(D)=\frac{1}{4}}$$

pokus: bacanje dvije igraće kocke
D....pojava barem jedne šestice

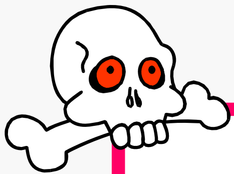
prostor ishoda= {(1,1) (1,2) (1,3) (1,4) (1,5) (**1,6**)
(2,1) (2,2) (2,3) (2,4) (2,5) (**2,6**)
(3,1) (3,2) (3,3) (3,4) (3,5) (**3,6**)
(4,1) (4,2) (4,3) (4,4) (4,5) (**4,6**)
(5,1) (5,2) (5,3) (5,4) (5,5) (**5,6**)
(**6,1**) (**6,2**) (**6,3**) (**6,4**) (**6,5**) (**6,6**) }

- taj događaj realizira 11 od 36 jednako mogućih događaja

$$\mathbf{P(D)=\frac{11}{36}}$$

RELATIVNA FREKVENCIJA

– omjer broja povoljnih ishoda i ukupnog broja pokusa ili ispitanika na kojima promatramo događaj



$$\boxed{f(D) = \frac{m(D)}{n}}$$

RELATIVNA FREKVENCIJA - primjer

Od 674 stanovnika otoka Suska, 312 stanovnika ima krvnu grupu O, 340 grupu A, 17 grupu B a 5 grupu AB. Kolika je vjerojatnost da jedan slučajno odabrani stanovnik ima krvnu grupu O?

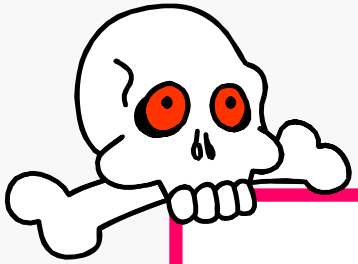
D....krvna grupa je O

broj povoljnih ishoda

$$P(D) = \frac{312}{674} \approx 0.46$$

broj ispitanika

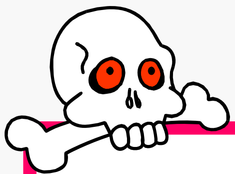
ZAKON VELIKIH BROJEVA



KADA BROJ POKUSA RASTE, APSOLUTNA RAZLIKA IZMEĐU RELATIVNE FREKVENCIJE I VJEROJATNOSTI SE SMANJUJE

VJEROJATNOST A POSTERIORI (statistička vjerojatnost)

- granična vrijednost relativne frekvencije kada broj pokusa raste u beskonačnost



$$\boxed{P(D) = \lim_{n \to \infty} \frac{m(D)}{n}}$$

vjerojatnost događaja D

$$\mathbf{m(D) \leq n \Rightarrow 0 \leq P(D) \leq 1}$$

SKALA VJEROJATNOSTI

SUBJEKTIVNA VJEROJATNOST

- vjerojatnost događaju D se dodjeljuje prema subjektivnoj procjeni pojedinca

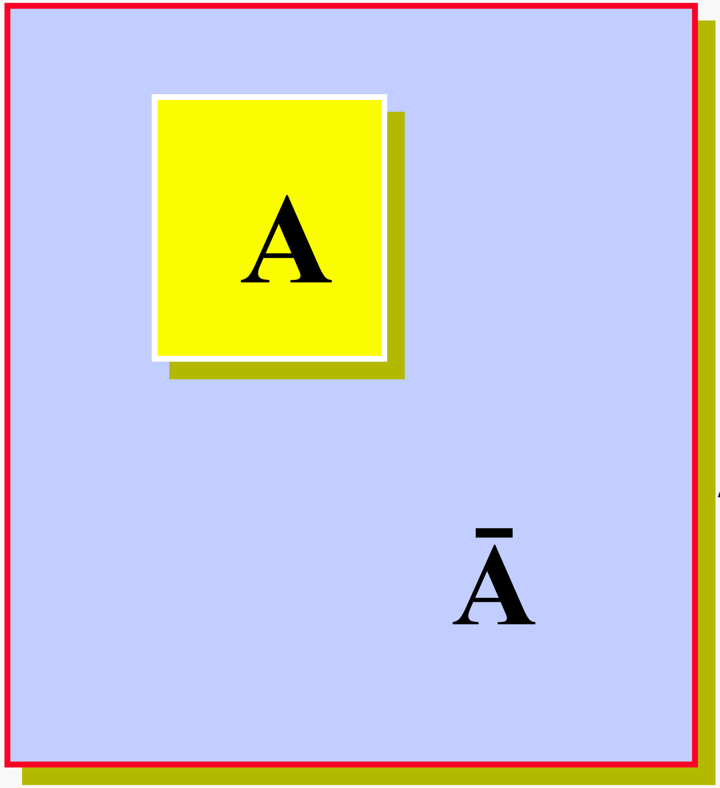
Prijatelj: Odgovor je: D

Ja: Koliko si siguran?

Prijatelj: Više od 77%

OSNOVNA PRAVILA RAČUNA VJEROJATNOSTI

PRAVILO KOMPLEMENTIRANJA
(suprotna vjerojatnost)



P(A)... vjerojatnost nastupanja događaja A

Ā ... "non A" (događaj koji označava **ne nastupanje** događaja A)

$$P(\overline{A}) = 1 - P(A)$$

PRAVILO KOMPLEMENTIRANJA - primjer

Ako je vjerojatnost rođenja muškog djeteta 0.52, kolika je vjerojatnost rođenja ženskog djeteta?

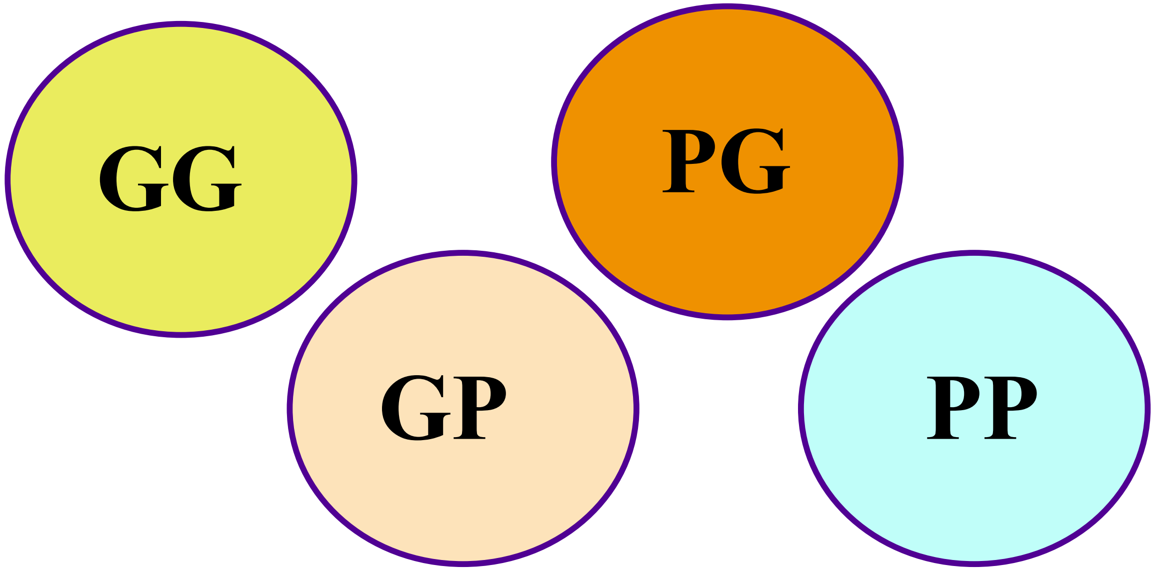
P(muško) = 0.52

P(žensko) = 1 - 0.52 = 0.48

DOGAĐAJI KOJI SE MEĐUSOBNO ISKLJUČUJU

- ne mogu nastupiti istovremeno
- disjunktni događaji

U pokusu bacanja dva novčića sva četiri moguća ishoda se međusobno isključuju



PRAVILO ADICIJE
(za događaje koji se međusobno isključuju)

$A_1, A_2, A_3, \ldots, A_k$ događaji koji se međusobno isključuju

A_1 **U** A_2 **U** ....**U** A_k složeni događaj, nastaje kada nastupi **ili** A_1 **ili** A_2 **ili..ili** A_k

$$\boxed{P(A_1 U A_2 U \ldots U A_k) = P(A_1) + P(A_2) + \ldots + P(A_k)}$$

PRAVILO ADICIJE - primjer

U nekoj populaciji vjerojatnosti pojedinih krvnih grupa su:

$P(O)=0.42$, $P(A)=0.43$, $P(B)=0.11$ i $P(AB)=0.04$.

Kolika je vjerojatnost da slučajno odabrani pripadnik te populacije ima krvnu grupu A ili B?

$$\mathbf{P(A \text{ ili } B) = 0.43 + 0.11 = 0.54}$$

PRAVILO ADICIJE - primjer

Ako je vjerojatnost svijetle kose 0.20, vjerojatnost smeđe kose 0.40 a vjerojatnost crne kose 020, kolika je vjerojatnost da kosa bude:

a) smeđa ili crna

b) svijetla ili smeđa ili crna?

a) P(smeđa ili crna) = 0.40 + 0.20 = 0.60

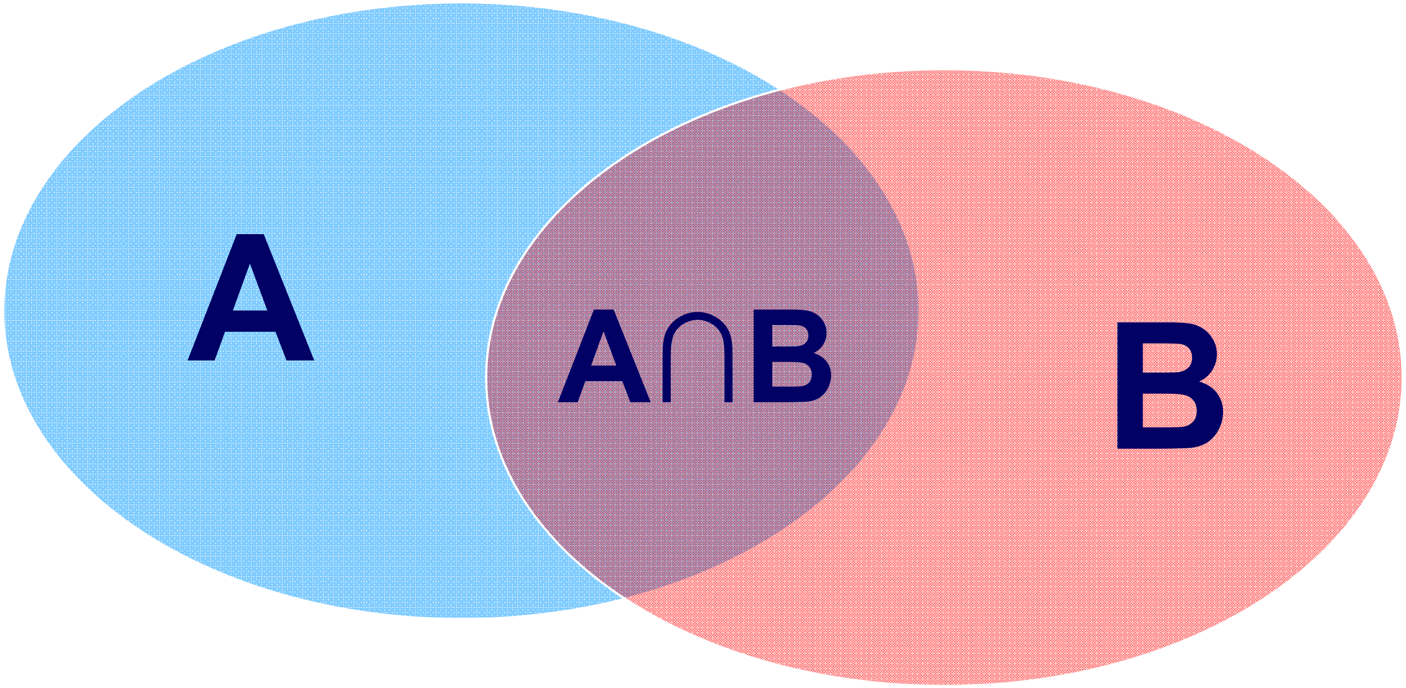
b) P(svijetla ili smeđa ili crna) = 0.20 + 0.40 + 0.20 = 0.80

NEZAVISNI DOGAĐAJI

- događaji koji mogu nastupiti istovremeno, pri čemu vjerojatnost nasupanja nekog od njih ne ovisi o realizaciji drugih

PRAVILO MULTIPLIKACIJE
(za nezavisne događaje)

A , B ...... nezavisni događaji



A ∩ B novi događaj koji nastupi kada se istovremeno realiziraju događaji A i B

PRAVILO MULTIPLIKACIJE
(za nezavisne događaje)

$A_1, A_2, A_3, \ldots, A_k$ ... nezavisni događaji

$A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_k$ događaj koji nastupi kada se istovremeno realiziraju događaji $A_1, A_2, A_3, \ldots, A_k$

$$P(A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_k) = P(A_1) \cdot P(A_2) \cdot P(A_3) \cdot \ldots \cdot P(A_k)$$

PRAVILO MULTIPLIKACIJE - primjer

Ako je vjerojatnost svijetle kose 0.30, a vjerojatnost crnih očiju 0.20, kolika je vjerojatnost istovremenog pojavljivanja svijetle kose i crnih očiju?

P(svijetla kosa i crne oči) = 0.30 * 0.20 = 0.06

primjer

Rezultati studije pokazali su da 60% majki djece do 10 godina radi puno radno vrijeme. Ako slučajno odaberemo tri majke, kolika je vjerojatnost da barem jedna od njih radi puno radno vrijeme?

$$
\begin{aligned}
P(\text{barem jedna radi PRV}) &= 1 - P(\text{niti jedna ne radi PRV}) = \\
&= 1 - [(0.4)(0.4)(0.4)] = \\
&= 1 - (0.4)^3 = 1 - 0.064 = \\
&= 0.936
\end{aligned}
$$

...općenito....

vjerojatnost da se u nizu od **m** pokusa događaj **A** pojavi BAREM jedan puta dana je sa:

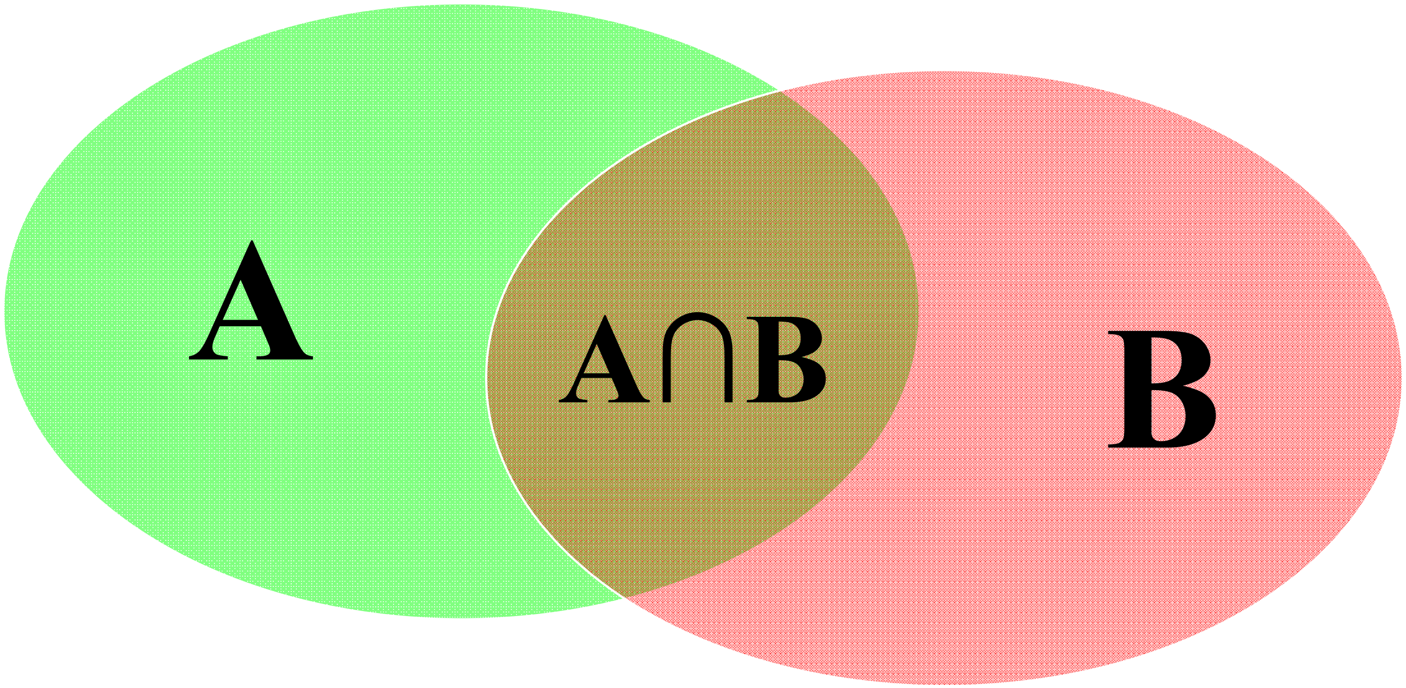
$$\boxed{\mathbf{P}(\text{barem jedan A}) = 1 - q^{\mathbf{m}}}$$

q.... vjerojatnost da se **A** NE DOGODI

$q = 1 - P(A)$

OPĆE PRAVILO ADICIJE

A , B ...... događaji koji mogu istovremeno nastupiti



$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

OPĆE PRAVILO ADICIJE - primjer

Od 150 studenata u domu 40 je imalo CD, 80 TV a 30 i CD i TV. Kolika je vjerojatnost da slučajno odabrani student ima ili CD ili TV?

A ....student ima CD $P(A) = 40/150 = 0.2667$

B ....student ima TV $P(B) = 80/150 = 0.5333$

AiB ...student ima i CD i TV $P(A \cap B) = 30/150 = 0.2$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.2667 + 0.5333 - 0.2 = 0.6$$

UVJETNA VJEROJATNOST

- osnovna vjerojatnost u prirodnim i humanističkim istraživanjima
- ishodu nekog događaja prethodi neki drugi događaj kao uvjet za slijedeći potencijalni događaj

letalitet (stopa umrlih od neke bolesti) je tipična uvjetna vjerojatnost - mora biti zadovoljen uvjet da je osoba oboljela od te bolesti

UVJETNA VJEROJATNOST

A, B...događaji

vjerojatnost događaja B uz uvjet da se događaj A odigrao dana je sa:

$$\boxed{P(B \mid A) = \frac{P(A \cap B)}{P(A)}}$$

UVJETNA VJEROJATNOST - primjer

Učestalost sljepoće za boje u ljudskoj populaciji različita je prema spolu (X-kromosomska nasljedna anomalija).

Učestalost (%)

	muškarci	žene	ukupno
slijepi za boje	4,23	0,65	4,88
normalni	48,48	46,64	95,12
ukupno	52,21	47,29	100,00

UVJETNA VJEROJATNOST - primjer

Incidencija sljepoće za boje u muškoj subpopulaciji vodi na uvjetnu vjerojatnost

$$P(S\,|\,M) = \frac{P(S \cap M)}{P(M)}$$

$P(S \cap M)$

	muškarci	žene	ukupno
slijepi za boje	0.0423	0.0065	0.0488
normalni	0.4848	0.4664	0.9512
ukupno	0.5221	0.4729	1.0000

$P(M)$

OPĆE PRAVILO MULTIPLIKACIJE

neka su A, B događaji koji NISU nezavisni

$$\boxed{\begin{aligned} P(A \cap B) &= P(A)P(B|A) \\ P(A \cap B) &= P(B)P(A|B) \end{aligned}}$$

A...preživljavanje kemoterapije ; P(A)=0.9

B...izlječenje leukemije

B|A ....izlječenje leukemije pod uvjetom preživljavanja kemoterapije; P(B|A)= 0.8

Kolika je vjerojatnost da bolesnik preživi kemoterapiju i bude izliječen od leukemije?

$$P(A \cap B)=P(A)P(B|A)=0.9*0.8=0.72$$

OPĆE PRAVILO MULTIPLIKACIJE
primjer

Morbiditet od neke bolesti u populaciji je 0.10, a letalitet od te iste bolesti u istoj populaciji 0.08. Kolika je vjerojatnost da slučajno odabran član te populacije oboli i umre?

P(A)= 0.10

$$\mathbf{P(B|A) = 0.08}$$

$$\mathbf{P(A \cap B) = P(A)P(B|A) = 0.10 * 0.08 = 0.008}$$

NEZAVISNI DOGAĐAJI

događaji A i B su nezavisni ako vrijedi:

$$\boxed{P(A|B) = P(A)}$$

ili

$$\boxed{P(A \cap B)=P(A)P(B)}$$

NEZAVISNI DOGAĐAJI

	gluhoća		
	postoji	ne postoji	ukupno
slijepi za boje	0.0004	0.0796	0.0800
normalni	0.0046	0.9154	0.9200
ukupno	0.0050	0.9950	1.0000

$P(G)$

$P(\overline{G})$

$$P(G|S) = 0.0004/0.0800 = 0.0050$$

$$P(\overline{G}|S) = 0.0796/0.0800 = 0.9950$$

$$\mathbf{P(S \cap G) = P(S) \bullet P(G|S)}$$

$$\mathbf{P(G|S) = P(G)}$$

$$\mathbf{P(S \cap G) = P(S) \bullet P(G)}$$ ....događaji su nezavisni

RASPODJELE VJEROJATNOSTI

SLUČAJNA VARIJABLA

- pravilo po kojem se pojedinim slučajnim događajima dodjeljuju brojevi (atributi) tako da je uz svaki od njih pridružena određena vjerojatnost

Npr. varijabla **broj infekcija** može poprimiti vrijednosti 0, 1, 2, ...

Zovemo ju **slučajnom** ako je svakoj od vrijednosti pridružena određena vjerojatnost (npr. vjerojatnost da netko ima dvije infekcije u nekoj skupini ljudi je 0,07)

DISKRETNA SLUČAJNA VARIJABLA (definicija)

Diskretna slučajna varijabla je varijabla X koja poprima niz vrijednosti

$$x_1, x_2, \ldots, x_k$$

ali svaku od njih s određenom vjerojatnošću

$$p(x_1), p(x_2), \ldots, p(x_k), \quad p(x_i) \geq 0, \forall i$$

pri čemu za vjerojatnosti $p(x_i)$ vrijedi

$$\sum_{i=1}^{k} p(x_i) = 1$$

RASPODJELA slučajne varijable X

skup svih parova $\{ x_i, p(x_i) \}$, $i=1,2,...$

FUNKCIJA VJEROJATNOSTI slučajne varijable X

zakon $p(x)$ po kojem svakoj vrijednosti x_i pripada vjerojatnost $p(x_i)$

FUNKCIJA RASPODJELE slučajne varijable X

$$\boxed{F(x_0) = \sum_{x_i \leq x_0} p(x_i)}$$

- pokazuje kolika je vjerojatnost da slučajna varijabla X poprimi bilo koju vrijednost manju ili jednaku x_0 tj.

$$F(x_0) = P\{x \leq x_0\}$$

PRIMJER:

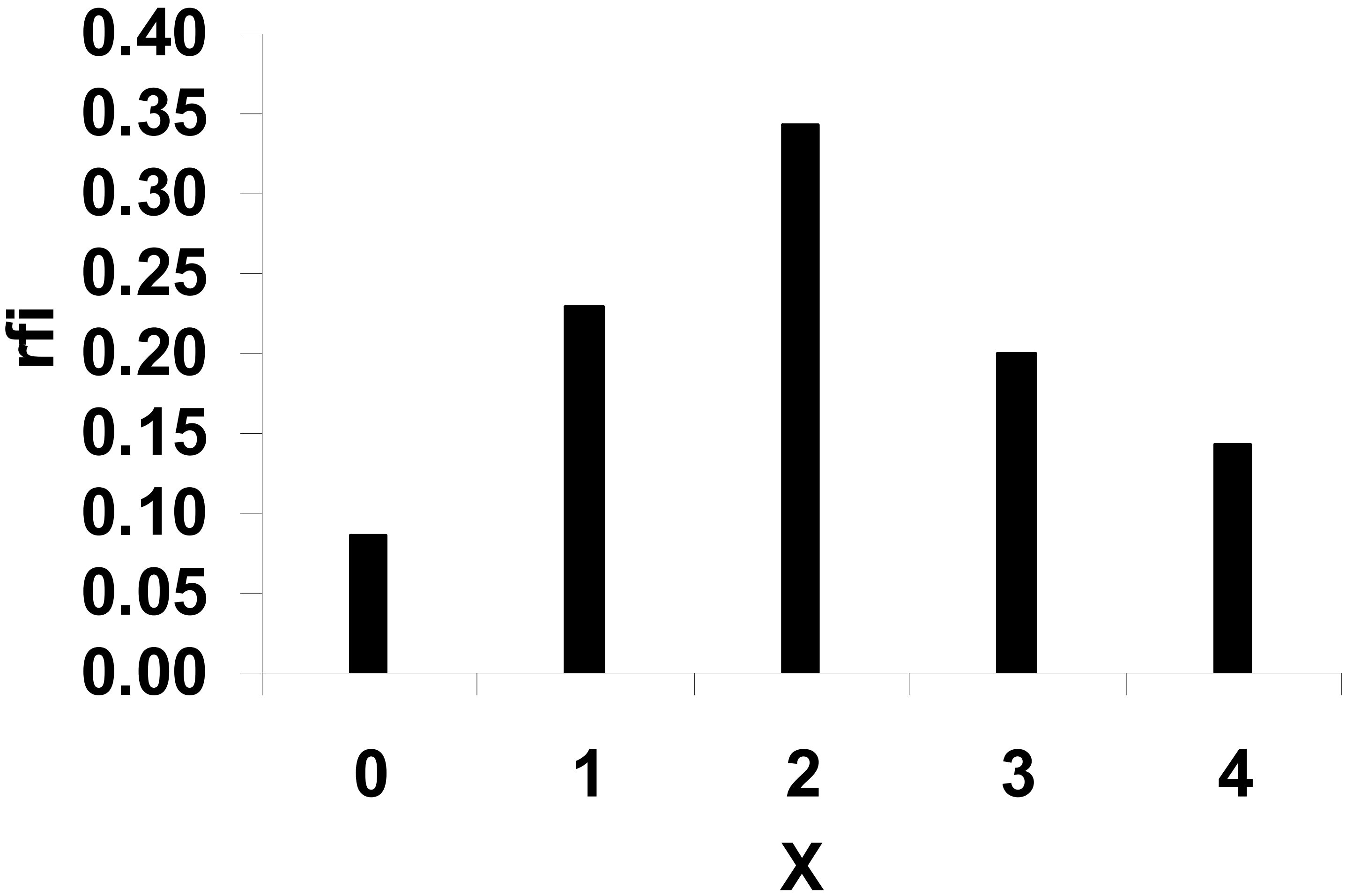
Na 35 preparata nekog biološkog materijala mjeren je broj nezrelih stanica. Dobiveni su slijedeći rezultati:

X_i	f_i	rf_i	crf_i
0	3	0.09 $=p(x_0)$	0.09 $= F(x_0)=p(x_0)$
1	8	0.23 $=p(x_1)$	0.32 $= F(x_1)=p(x_0)+p(x_1)$
2	12	0.34 $=p(x_2)$	0.66 $= F(x_2)=p(x_0)+p(x_1)+p(x_2)$
3	7	0.20 $=p(x_3)$	0.86 :
4	5	0.14 $=p(x_4)$	1.00 :
	35	1.00	

rf_i .... vjerojatnost da je broj nezrelih stanica jednak X_i - $\mathbf{p(x_i)}$

crf_i ... vjerojatnost da je broj nezrelih stanica manji ili jednak X_i - $\mathbf{F(x_i)}$

raspodjela vjerojatnosti



KONTINUIRANA SLUČAJNA VARIJABLA

- **područje vrijednosti** kontinuirane slučajne varijable je interval na brojevnom pravcu (moguće i čitav brojevni pravac)
- vjerojatnost pridružujemo intervalima brojevnog pravca
- pojedinačnim vrijednostima x_i pripada vjerojatnost 0
- nekom intervalu (x_1, x_2) pridružujemo vjerojatnost $P\{x_1<x<x_2\}$ po **funkciji vjerojatnosti f(x)**

(definicija)

Funkcija vjerojatnosti kontinuirane slučajne varijable X je funkcija f(x) koja ima slijedeća svojstva:

1. $f(x) \geq 0, \forall x$

2. $\int_{-\infty}^{+\infty} f(x)dx = 1$

3. $\int_{x_1}^{x_2} f(x)dx = P\{x_1 < x < x_2\}$

pri čemu su x_1, x_2 bilo koje dvije vrijednosti varijable x takve da je $x_1 < x_2$

značenje svojstava funkcije vjerojatnosti:

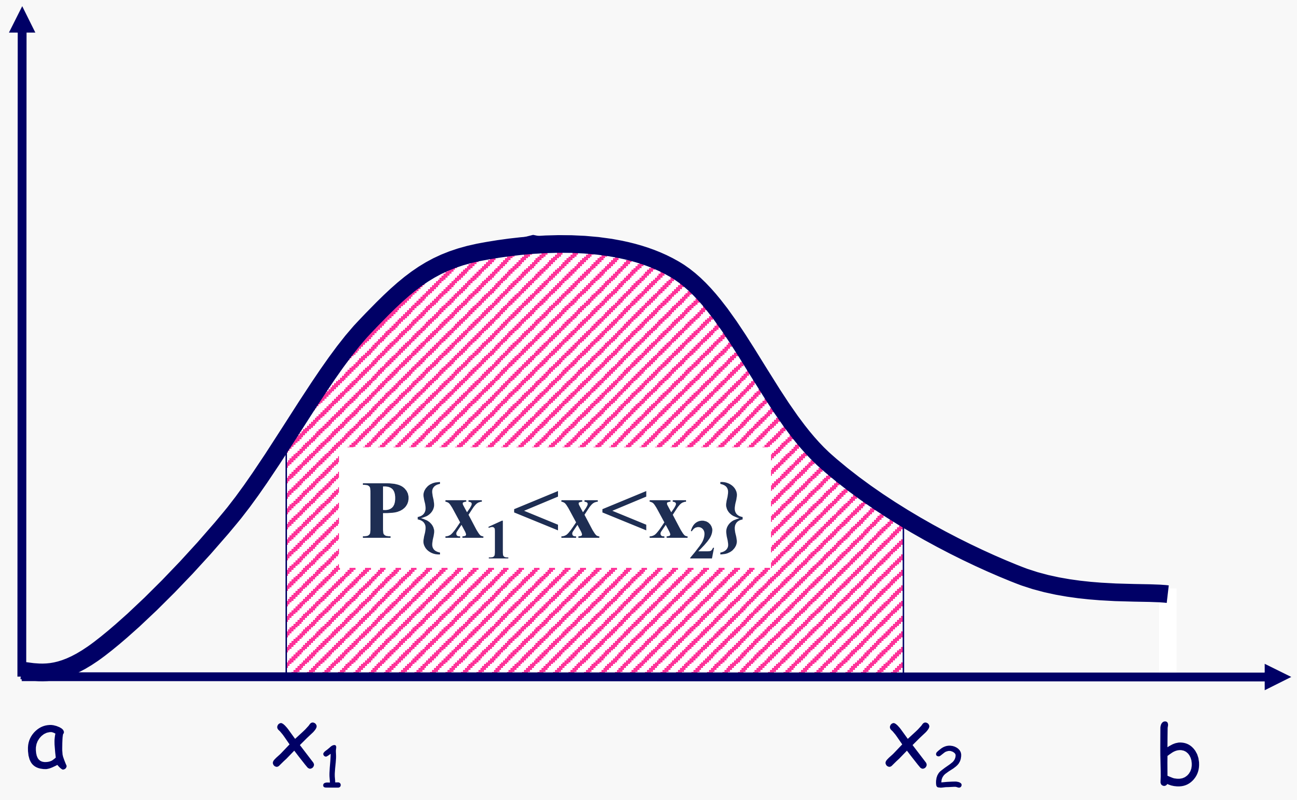
- **1. svojstvo** – NENEGATIVNOST (vjerojatnost ne može biti negativan broj)

- **2. svojstvo** – VJEROJATNOST SIGURNOG DOGAĐAJA je 1
  - ako je područje vrijednosti slučajne varijable X interval (a,b), onda 2. svojstvo poprima oblik

$$\boxed{\int_a^b f(x)dx = 1}$$

i uzimamo da je **f(x)=0** za sve vrijednosti x izvan područja vrijednosti, tj. intervala (a,b)

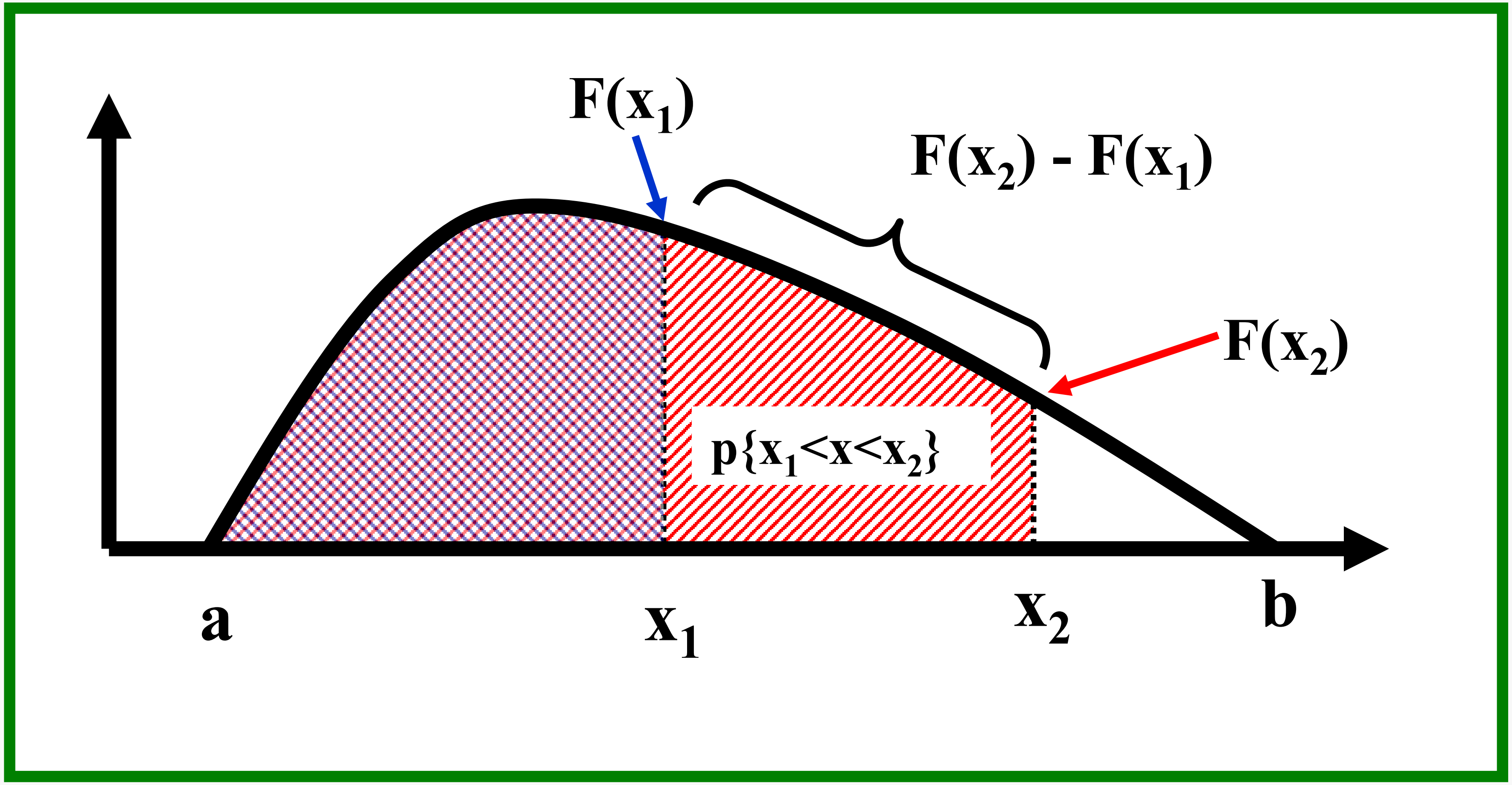
- 3. svojstvo – $P\{x_1<x<x_2\}$ je numerički jednaka površini ispod krivulje vjerojatnosti nad intervalom (x_1, x_2)



FUNKCIJA RASPODJELE kontinuirane slučajne varijable

$$\mathbf{F(x_0)} = \int_{-\infty}^{\mathbf{x_0}} f(x)dx = P\{x \leq x_0\}$$

- pokazuje kolika je vjerojatnost da slučajna varijabla X poprimi bilo koju vrijednost manju ili jednaku x_0



F(x1)
F(x2) - F(x1)
F(x2)
p{x1<x<x2}
a
x1
x2
b

KARAKTERISTIČNE VRIJEDNOSTI SLUČAJNIH VARIJABLI

OČEKIVANJE (sredina) SLUČAJNE VARIJABLE

$$E(x) = \mu = \sum_i x_i p(x_i)$$

očekivanje diskretne slučajne varijable

$$E(x) = \mu = \int_{-\infty}^{+\infty} x f(x) dx$$

očekivanje kontinuirane slučajne varijable

- kod empirijske raspodjele frekvencija to je aritmetička sredina:

$$\mu = \frac{\sum_i x_{i0}}{N} = \frac{\sum_i f_i x_i}{N} = \sum_i f_{ri} x_i$$

VARIJANCA (disperzija) SLUČAJNE VARIJABLE

$$V(x) = \sum_i (x_i - \mu)^2 p(x_i)$$

varijanca diskretne slučajne varijable

$$V(x) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x)dx$$

varijanca kontinuirane slučajne varijable

- kod empirijske raspodjele frekvencija:

$$\sigma^2 = \frac{\sum_i (x_{i0} - \mu)^2}{N} = \frac{\sum_i f_i (x_i - \mu)^2}{N} = \sum_i f_{ri} (x_i - \mu)^2$$

TEORIJSKE RASPODJELE

BINOMNA RASPODJELA X~B(n,p)

- $(p+q)^n$

BERNOULLIJEV DOGAĐAJ

- događaj *A* čija se vjerojatnost nastupanja

$$P(A) = p$$

ne mijenja tijekom pokusa, a **vjerojatnost NE nastupanja** događaja *A* je

$$q = 1 - p$$

npr. u pokusu bacanja novčića pri svakom bacanju novčića vjerojatnost pojave "grba" je p = 0,5

- vjerojatnost da Bernoullijev događaj A u nizu od n pokusa nastupi x puta:

$$\boxed{P(x)=\binom{n}{x}p^{x}q^{n-x}}$$

BERNOULLIJEVA FORMULA

gdje je $\boxed{\binom{n}{x}=\frac{n!}{x!(n-x)!}}$ BINOMNI KOEFICIJENT

BINOMNA RASPODJELA

- skup svih parova $\{x, P(x)\}$, $x = 0, 1, 2, ..., n$

Primjer rekombinacije gena:

Vjerojatnost pojave gena A je p(A)=p, a vjerojatnost pojave gena a je p(a)=q. Koje su vjerojatnosti mogućih genotipova?

Mogu nastupiti kombinacije: AA, Aa, aA, aa.
Aa i aA se ne mogu biološki razlikovati =>
prostor ishoda je {AA,Aa,aa}

$$P(AA) = \binom{2}{2} p^2 q^{2-2} = p^2$$

$$P(Aa) = \binom{2}{1} p^1 q^{2-1} = 2pq$$

$$P(aa) = \binom{2}{0} p^0 q^{2-0} = q^2$$

KARAKTERISTIKE BINOMNE RASPODJELE

- jednoznačno je određena parametrima **n**, **p**

$$\boxed{E(x) = np}$$

OČEKIVANJE

$$\boxed{\sigma^2 = V(x) = npq}$$

VARIJANCA

KARAKTERISTIKE BINOMNE RASPODJELE

KOEFICIJENT ASIMETRIJE

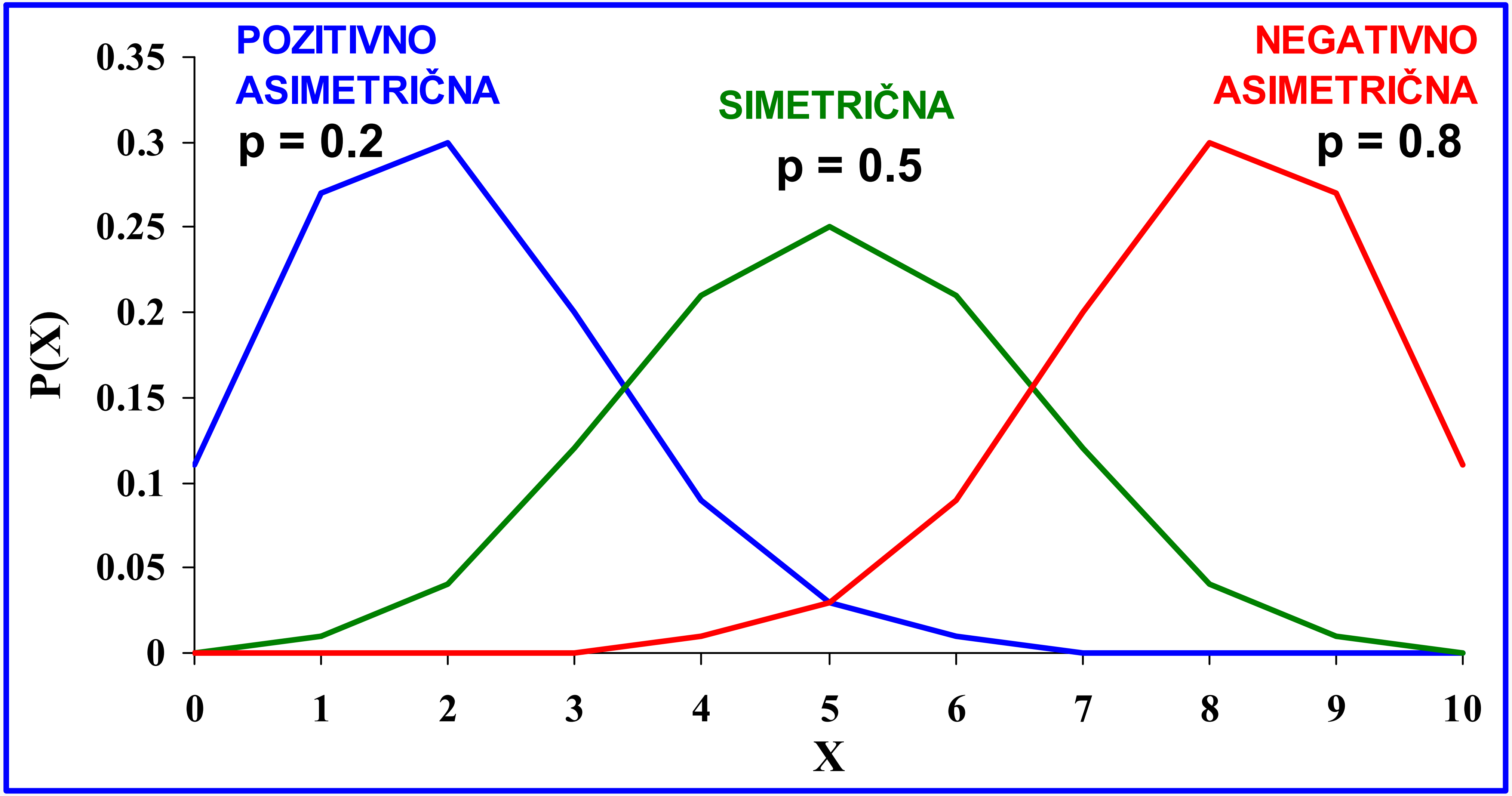
$$\boxed{\alpha_3 = \frac{q - p}{\sqrt{npq}}}$$

KOEFICIJENT SPLJOŠTENOSTI

$$\boxed{\alpha_4 = 3 + \frac{1 - 6pq}{npq}}$$

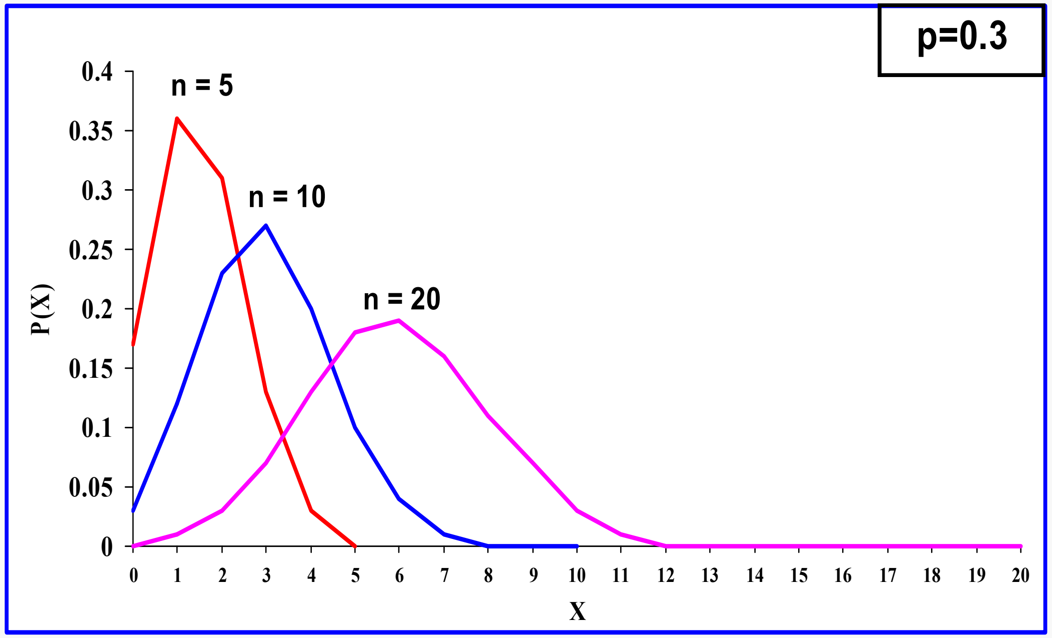
- SIMETRIČNA za $p = q = 1/2$
- POZITIVNO ASIMETRIČNA ako je $q>p$, tj. $q > 1/2$
- NEGATIVNO ASIMETRIČNA ako je $q<p$, tj. $q < 1/2$

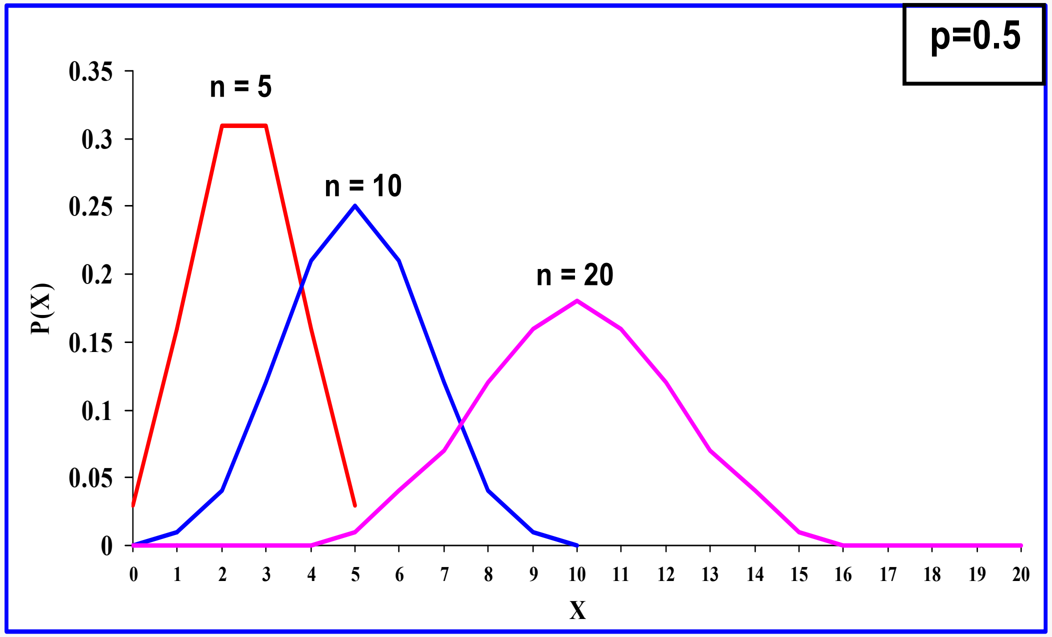
BINOMNA RASPODJELA ZA RAZLIČITE PARAMETRE p

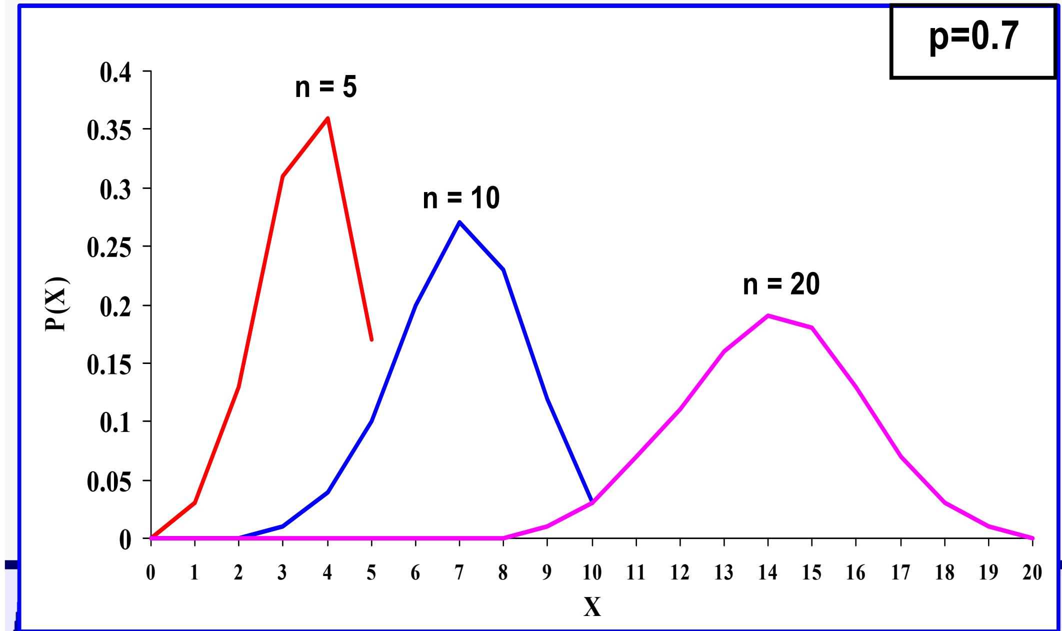


- bez obzira na međusobni odnos parametara p i q, vrijedi:

$$\alpha_3 \rightarrow 0 \quad \text{kada} \quad n \rightarrow \infty \; ; \qquad \alpha_4 \rightarrow 3 \quad \text{kada} \quad n \rightarrow \infty$$







PRIMJENA BINOMNE RASPODJELE ZA PROCJENU VJEROJATNOSTI SMRTNOG ISHODA

Ako je letalitet od neke bolesti p=0,30 a vjerojatnost preživljavanja q=0,70 pitanje vjerojatnosti smrtnog ishoda i preživljavanja za 5 bolesnika možemo prikazati kao binomnu raspodjelu (n=5; p=0,30)

PRIMJENA BINOMNE RASPODJELE ZA PROCJENU VJEROJATNOSTI SMRTNOG ISHODA

BROJ UMRLIH	VJEROJATNOST	
0	$P(0)=\binom{5}{0}p^0q^5=q^5$	$=0.16807$
1	$P(1)=\binom{5}{1}p^1q^4=5pq^4$	$=0.36015$
2	$P(2)=\binom{5}{2}p^2q^3=10p^2q^3$	$=0.30870$
3	$P(3)=\binom{5}{3}p^3q^2=10p^3q^2$	$=0.13230$
4	$P(4)=\binom{5}{4}p^4q^1=5p^4q$	$=0.02835$
5	$P(5)=\binom{5}{5}p^5q^0=p^5$	$= 0.00243$

PRILAGOĐAVANJE BINOMNE RASPODJELE EMPIRIJSKIM PODATCIMA

Pokus: 1000 bacanja 7 novčića. Kolike su očekivane frekvencije pojavljivanja grba uz pretpostavku da su novčići ispravni?

Znamo:

- novčići su ispravni => $p = 0.5,\ q = 0.5$
- u svakom je bacanju 7 novčića => $n = 7$
- vjerojatnost pojave X grbova dana je sa

$$P(x) = \binom{7}{x} p^x q^{7-x} = \binom{7}{x} 0.5^x 0.5^{7-x} = \binom{7}{x} 0.5^7$$

x	P(x)	P(x)*N	f_{ex}
0	$\binom{7}{0}0.5^7 = 0.5^7$	7.81	8
1	$\binom{7}{1}0.5^7 = 7 \cdot 0.5^7$	54.69	55
2	$\binom{7}{2}0.5^7 = 21 \cdot 0.5^7$	164.06	164
3	$\binom{7}{3}0.5^7 = 35 \cdot 0.5^7$	273.44	273
4	$\binom{7}{4}0.5^7 = 35 \cdot 0.5^7$	273.44	273
5	$\binom{7}{5}0.5^7 = 21 \cdot 0.5^7$	164.06	164
6	$\binom{7}{6}0.5^7 = 7 \cdot 0.5^7$	54.69	55
7	$\binom{7}{7}0.5^7 = 0.5^7$	7.81	8

POISSONOVA RASPODJELA

- granični prijelaz binomne raspodjele kada

  n $\to\infty$ uz uvjet **n·p = const.**

$$\boxed{P(x) = \frac{\mu^{x} e^{-\mu}}{x!}, \quad P(0) = e^{-\mu}}$$

funkcija vjerojatnosti Poissonove raspodjele

gdje je μ= n·p ;   e .... baza prirodnog logaritma

($e \approx 2.72$)

KARAKTERISTIKE POISSONOVE RASPODJELE

- jednoznačno je određena parametrom μ

$$\boxed{E(x)=\mu}$$

OČEKIVANJE

$$\boxed{\sigma^2=V(x)=\mu}$$

VARIJANCA

KARAKTERISTIKE POISSONOVE RASPODJELE

KOEFICIJENT ASIMETRIJE

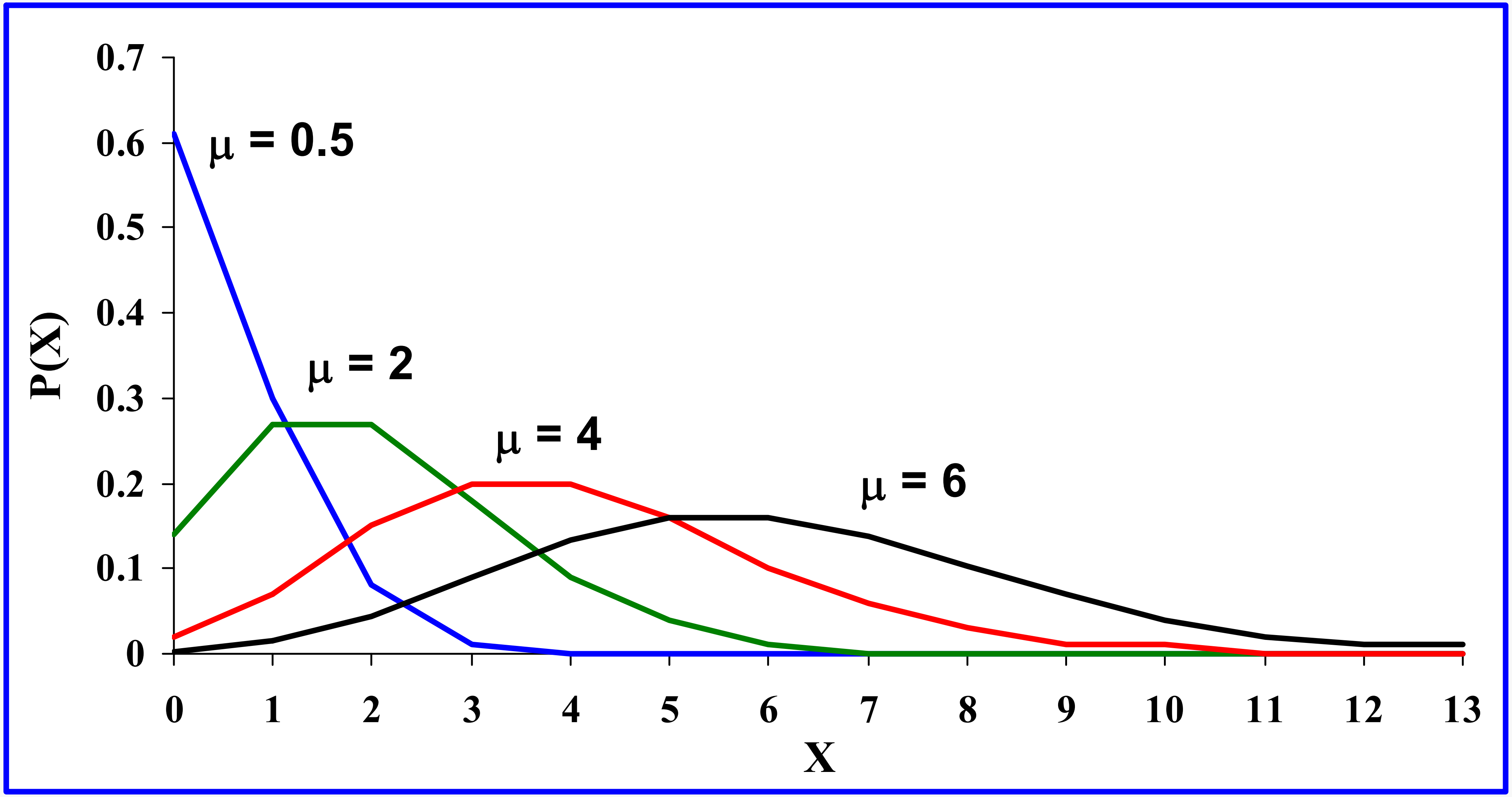
$$\boxed{\alpha_3 = \frac{1}{\sqrt{\mu}}}$$

KOEFICIJENT SPLJOŠTENOSTI

$$\boxed{\alpha_4 = 3 + \frac{1}{\mu}}$$

- $\alpha_3 > 0, \forall \mu$ (pozitivna asimetrija)
- povećanjem parametra μ asimetrija se smanjuje

POISSONOVA RASPODJELA ZA RAZLIČITE PARAMETRE μ



POISSONOVA RASPODJELA

- opisuje slučajnu raspodjelu događaja u vremenu ili sitnih čestica u prostoru
- raspodjela "rijetkih događaja"

PRIMJER: U promatranjima Rutherforda i Geigera ustanovljeno je da jedan radioaktivan izvor emitira u prosjeku 3.87 α čestica u vremenskom intervalu od 7.5 sekundi. Kolika je vjerojatnost da je u jednoj sekundi emitirana:

a) najviše 1 čestica?

b) najmanje 1 čestica?

$$P(x) = \frac{\mu^{x} e^{-\mu}}{x!}, \quad P(0) = e^{-\mu} \qquad \mu = \frac{3.87}{7.5} = 0.516$$

a) $P(x \leq 1) = P(0) + P(1)$

$$P(0) = e^{-\mu} = e^{-0.516} = 0.5969$$

$$P(1) = \frac{0.516^{1} e^{-0.516}}{1!} = 0.516 \cdot 0.5969 = 0.3080$$

$$P(x \leq 1) = P(0) + P(1) = 0.5969 + 0.3080 = 0.9049$$

b) $P(x \geq 1) = 1 - P(x < 1) = 1 - P(0) = 1 - 0.5969 = 0.4031$

Aproksimacija binomne razdiobe Poissonovom

- **u slučajevima kad je** $n \cdot p \leq 10$ uz $n > 50$

PRIMJER: U seriji automatski očitanih nalaza KKS prosječno je 1% pogrešnih.

a) Kolika je vjerojatnost da od 200 nalaza ne bude niti jedan pogrešan?
b) Kolika je vjerojatnost da će u 300 nalaza biti najviše 1 pogrešan?

a) $p=0.01$
$n=200$
$\mu=n \cdot p=2$

$$P(0) = e^{-\mu} = e^{-2} = 0.1353$$

b) $p=0.01$
$n=300$
$\mu=n \cdot p=3$

$$P(x) = \frac{\mu^{x} e^{-\mu}}{x!}$$

$$P=P(x \leq 1)=P(0)+P(1)$$

$$P(0) = e^{-\mu} = e^{-3} = 0.0498$$

$$P(1) = \frac{3^{1} e^{-3}}{1!} = 3 \cdot e^{-3} = 3 \cdot 0.0498 = 0.1494$$

$$P=P(x \leq 1)=P(0)+P(1) = 0.0498 + 0.1494 = 0.1992$$

PRILAGOĐAVANJE POISSONOVE RASPODJELE EMPIRIJSKIM PODATCIMA

Na 576 ploča hranjivih podloga prebrojane su bakterije i dobiveno je sljedeće:

- na 229 pločica nije nađena niti jedna bakterija
- na 211 pločica nađena je 1 bakterija
- na 93 pločice nađene su 2 bakterije
- na 35 pločica nađene su 3 bakterije
- na 7 pločica nađene su 4 bakterije
- na jednoj pločici nađeno je 7 bakterija.

Odgovara li rast bakterija Poissonovoj raspodjeli?

vrijedi:

$$\frac{p(x)}{p(x-1)} = \frac{\mu}{x} \Rightarrow p(x) = \frac{\mu}{x} p(x-1)$$

x	f_x	$x \cdot f_x$	μ/x	p(x)	$f_{ex}=N \cdot p(x)$	f_{ex}
0	229	0	-	0.39365	226.74	227
1	211	211	0.9323	0.36700	211.39	211
2	93	186	0.4662	0.17108	98.54	99
3	35	105	0.3108	0.05317	30.63	31
4	7	28	0.2331	0.01239	7.14	7
5	0	0	0.1865	0.00231	1.33	1
6	0	0	0.1554	0.00036	0.21	0
7	1	7	0.1332	0.00005	0.03	0
	576	537				576

$$\mu = \overline{x} = \frac{1}{N}\sum_i x_i f_{xi} = \frac{537}{576} = 0.9323;$$

$$P(0) = e^{-\mu} = e^{-0.9323} = 0.3936$$

NORMALNA RASPODJELA $X \sim N(\mu,\sigma^2)$

- slučajna varijabla X ima normalnu raspodjelu ako je područje njenih vrijednosti <-∞,+∞>, a funkcija vjerojatnosti

$$\boxed{f(x) = \frac{1}{b\sqrt{2\Pi}} e^{-\frac{1}{2}\left(\frac{x-a}{b}\right)^2}}$$

FUNKCIJA VJEROJATNOSTI NORMALNE RASPODJELE

NORMALNA RASPODJELA X~N(μ,σ^2) .... svojstva ....

OČEKIVANJE $\boxed{E(X) = \mu = a}$

VARIJANCA $\boxed{V(X) = \sigma^2 = b^2}$

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\Pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}$$

FUNKCIJA VJEROJATNOST I NORMALNE RASPODJELE

- jednoznačno je određena parametrima

μ, σ^2

koeficijent asimetrije

$$\boxed{\alpha_3 = 0}$$

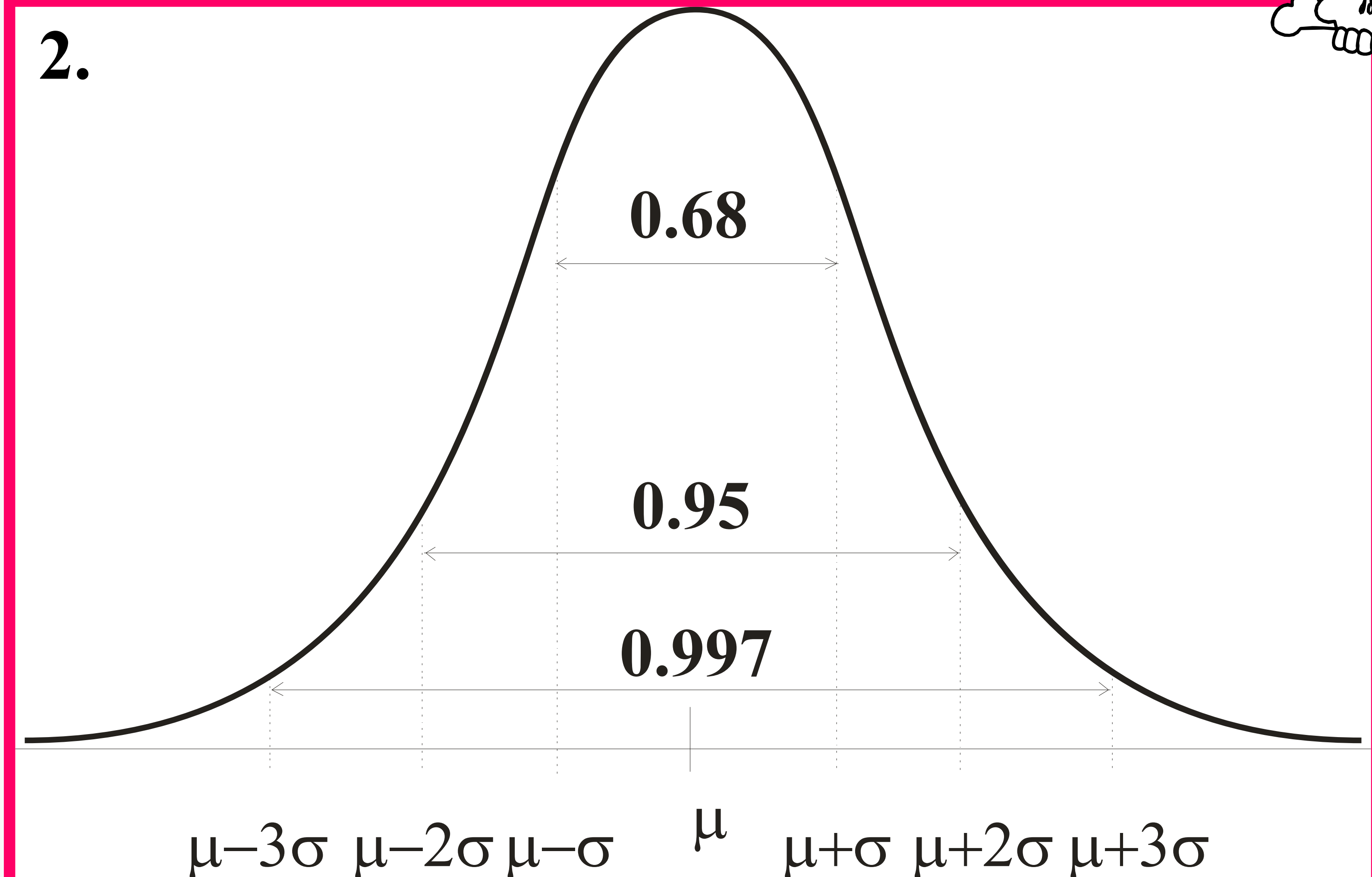
- simetrična s obzirom na pravac x = μ

koeficijent spljoštenosti $\boxed{\alpha_4 = 3}$

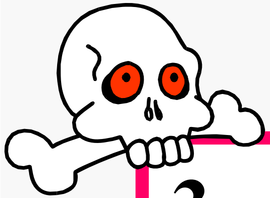
NORMALNA RASPODJELA $X \sim N(\mu,\sigma^2)$ ... svojstva ....

1. $\mu = Me = Mo$

2.



NORMALNA RASPODJELA X~N(μ,σ^2) .... svojstva ....



3. Površina ispod krivulje normalne raspodjele u intervalu između dvije vrijednosti koje su definirane udaljenošću od aritmetičke sredine izražene u standardnim devijacijama je KONSTANTNA bez obzira na stvarne vrijednosti aritmetičke sredine i standardne devijacije u pojedinom slučaju

Npr:

a)

$x_1 = 2.5$

$x_2 = 3.5$

$\mu = 3$

$\sigma = 0.5$

$P(x_1<x<x_2)=?$

$x_1 = 2.5 = \mu - \sigma$

$x_2 = 3.5 = \mu + \sigma$

$\mathbf{P(x_1<x<x_2)} = P(\mu-\sigma<x<\mu+\sigma) = \mathbf{0.68}$

b)

$x_1 = 10$

$x_2 = 14$

$\mu = 12$

$\sigma = 2$

$P(x_1<x<x_2)=?$

$x_1 = 10 = \mu - \sigma$

$x_2 = 14 = \mu + \sigma$

$\mathbf{P(x_1<x<x_2)} = P(\mu-\sigma<x<\mu+\sigma) = \mathbf{0.68}$

$$z = \frac{x-\mu}{\sigma}$$

$$z_{a1} = \frac{x_{a1}-\mu_a}{\sigma_a} = \frac{2.5-3}{0.5} = -1$$

$$z_{a2} = \frac{x_{a2}-\mu_a}{\sigma_a} = \frac{3.5-3}{0.5} = 1$$

$$z = \frac{x-\mu}{\sigma}$$

$$z_{b1} = \frac{x_{b1}-\mu_b}{\sigma_b} = \frac{10-12}{2} = -1$$

$$z_{b2} = \frac{x_{b2}-\mu_b}{\sigma_b} = \frac{14-12}{2} = 1$$

$$\boxed{z = \frac{x - \mu}{\sigma}}$$

STANDARDIZIRANA VRIJEDNOST

(standardized deviate, z-value)

- supstitucijom u funkciju vjerojatnosti normalne raspodjele dobivamo:

$$\boxed{f(x) = \frac{1}{\sigma} \cdot \frac{1}{\sqrt{2\Pi}} e^{-\frac{1}{2}z^2}}$$

funkcija od z

STANDARDNA (jedinična) NORMALNA RASPODJELA X~N(0,1)

$$\boxed{\varphi(z) = \frac{1}{\sqrt{2\Pi}} e^{-\frac{1}{2}z^2}}$$

FUNKCIJA VJEROJATNOSTI STANDARDNE NORMALNE RASPODJELE

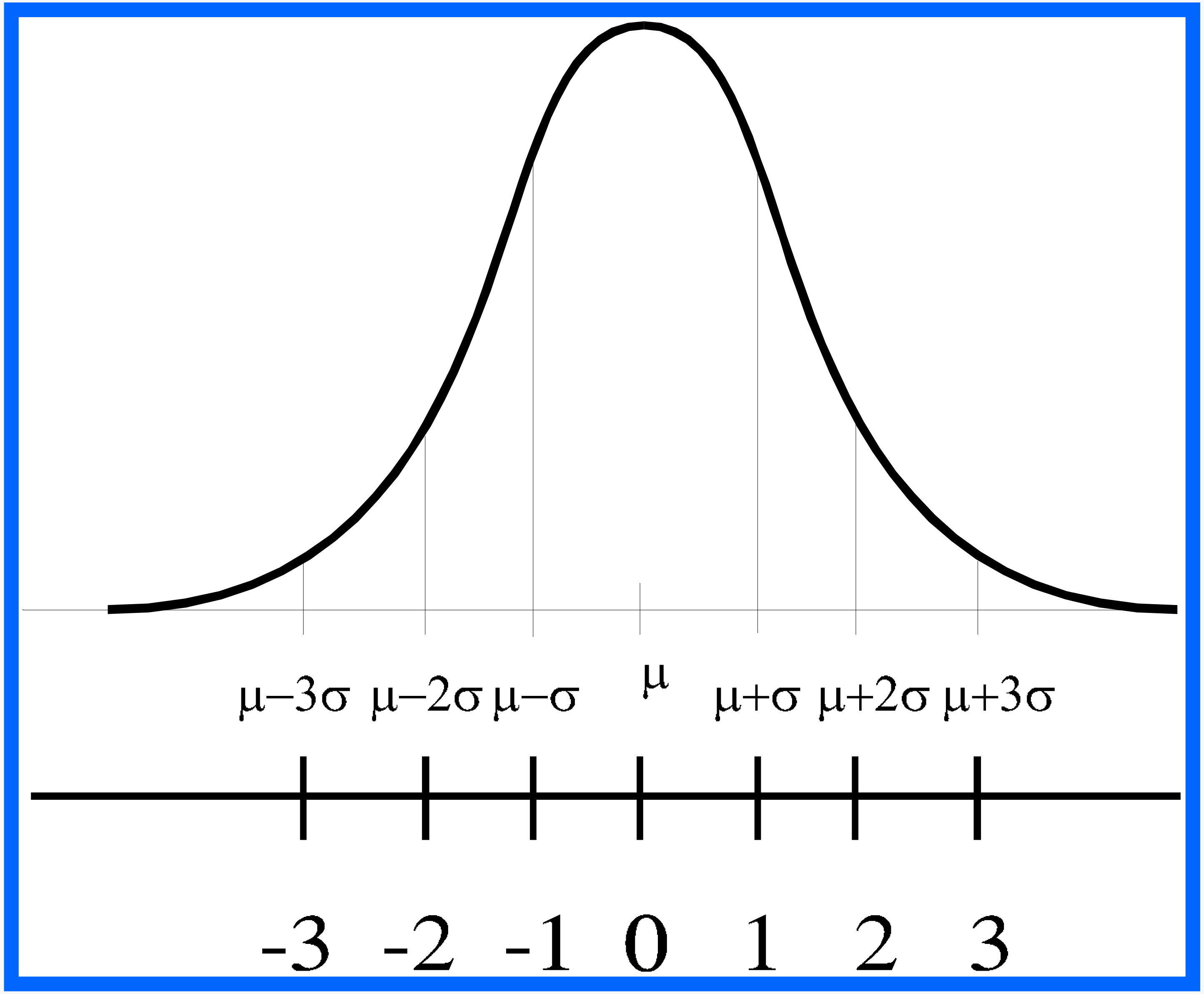
$$f(x) = \frac{1}{\sigma}\varphi(z)$$

OČEKIVANJE $\boxed{E(z) = 0}$

VARIJANCA $\boxed{V(z) = 1}$

X~N(0,1)

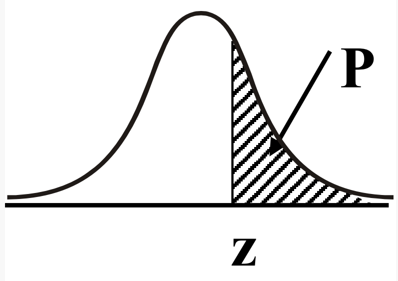
PRETVARANJE LJESTVICE MJERENJA U STANDARDIZIRANU Z-LJESTVICU



ljestvica mjerenja

z - ljestvica

TABLICA POVRŠINA ISPOD STANDARDNE NORMALNE RASPODJELE



- Tablica sadrži vrijednosti površina samo za nenegativne z vrijednosti ($z \geq 0$) i to iznad intervala <z, +∞>.
- Za odgovarajuće negativne z vrijednosti površina je jednaka, tj. $P_{-z} = P(-\infty < x < -z) = P(z < x < +\infty) = P_{+z}$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	[illegible]	[illegible]	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	[illegible]	[illegible]	[illegible]	[illegible]	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
0.8	0.2119	0.2090	0.2061	0.2033	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
0.9	0.1841	0.1814	0.1788	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.0	0.1587	0.1562	0.1539	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.1	0.1357	0.1335	0.1314	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	[illegible]	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681

površina ispod krivulje funkcije vjerojatnosti standardne normalne raspodjele iznad intervala <z, +∞>, za z=0.44

Na velikom uzorku izmjerena je visina desetogodišnjih dječaka. Aritmetička sredina visine bila je 137cm, a standardna devijacija 5cm. Kolika je visina od koje je 20% dječaka višlje?

P(z)=0.20

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611

z=0.84

iz $$z = \frac{x-\mu}{\sigma} \Rightarrow x = \sigma \cdot z + \mu = 5\text{cm} \cdot 0.84 + 137\text{cm} = 141.2\text{cm}$$

Koliki postotak dječaka ima visinu:

a) do 126cm

b) između 126cm i 134cm

c) između 134cm i 144cm

d) iznad 144cm

e) do 144cm?

a) $z_{126} = \frac{126-137}{5} = -2.2 \qquad P(z_{126}) = P_a = 0.0139 \Rightarrow 1.39\%$

b) $z_{134} = \frac{134-137}{5} = -0.6 \qquad P(z_{134}) = 0.2743$

$P_b = P(z_{134}) - P(z_{126}) = 0.2604 \Rightarrow 26.04\%$

c) $z_{144} = \frac{144-137}{5} = 1.4 \qquad P(z_{144}) = 0.0808$

$P_c = 1 - \left[P(z_{134}) + P(z_{144})\right] = 1 - (0.2743 + 0.0808) = 0.6449 \quad \Rightarrow 64.49\%$

d) $P_d = P(z_{144}) = 0.0808 \quad \Rightarrow 8.08\%$

e) $P_e = 1 - P_d = 1 - P(z_{144}) = 1 - 0.0808 = 0.9192 \quad \Rightarrow 91.92\%$

UZORAK I POPULACIJA

POPULACIJA

– osnovni skup
– skup svih jedinica promatranja (entiteta) opisanih varijablama (atributima)

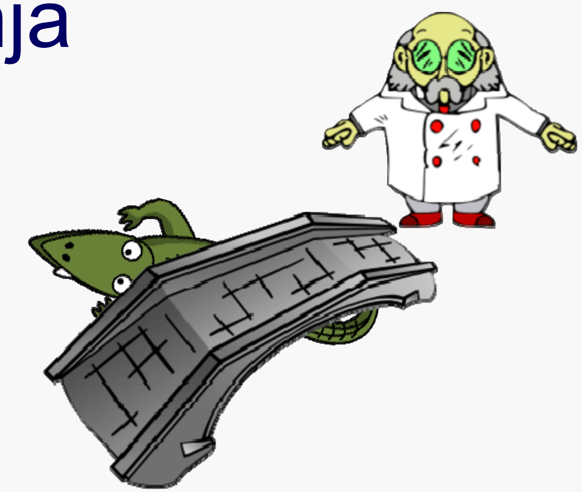
UZORAK - dio jedinica populacije (osnovnog skupa)

TEORIJA UZORAKA

– ustanovljava svojstva populacije iz svojstava uzorka
– procjenjuje parametre populacije na temelju podataka iz uzorka i ocjenjuje pouzdanost te procjene

UZORAK

- **zašto uzorak?**
  - brzina dobivanja rezultata
  - cijena istraživanja
  - dostupnost (materijala ili ispitanika)
  - stvarna nemogućnost ispitivanja populacije



temeljem podataka dobivenih iz uzorka

UOBIČAJENE OZNAKE

	OCJENA PARAMETRA (STATISTIKA)	PARAMETAR POPULACIJE
ARITMETIČKA SREDINA	$\overline{X}$	μ
STANDARDNA DEVIJACIJA	s	σ
PROPORCIJA	p	π

OSNOVNI POJMOVI

Koja je skupina na koju želimo generalizirati?

TEORETSKA POPULACIJA

Koja je populacija dostupna?

POPULACIJA KOJU ISTRAŽUJEMO

Na koji način možemo obuhvatiti populaciju?

OKVIR IZBORA

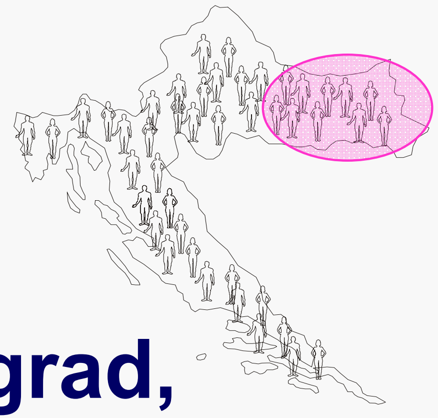
Tko je uključen u istraživanje?

UZORAK

TEOREṬSKA POPULACIJA

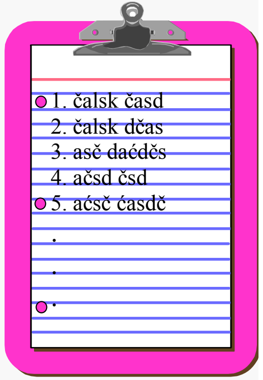
- **potpuna skupina na koju želimo generalizirati**
- **npr.:**
  - svi oboljeli od AIDS-a
  - svi oboljeli od dijabetesa
  - starije osobe oboljele od Alzheimerove bolesti zbrinute u ustanovi

POPULACIJA KOJU ISTRAŽUJEMO



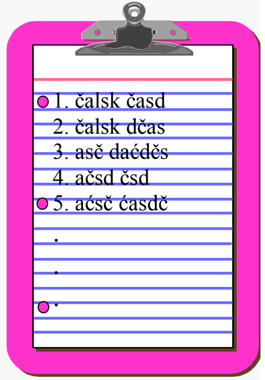
- **dostupna populacija**
- **može biti ograničena na instituciju, grad, županiju, državu, regiju**
- **npr.:**
  - svi oboljeli od AIDS-a u Osječko-baranjskoj županiji
  - svi oboljeli od dijabetesa registrirani u centru za dijabetes KBCO
  - starije osobe oboljele od Alzheimerove bolesti zbrinute u domovima za starije i nemoćne u OBŽ

OKVIR IZBORA



- popis svih članova (jedinica promatranja) dostupne populacije
- svaki član populacije treba biti identificiran i uključen u popis => svi trebaju imati jednaku šansu izbora u uzorak

UZORAK



- jedinice dostupne populacije izabrane iz okvira izbora nekom metodom uzorkovanja
- jedinice populacije koje će biti uključene u istraživanje (ispitanici ili objekti)

Kvaliteta ocjene parametara ovisi o:

- REPREZENTATIVNOSTI UZORKA
- ODABRANOJ VJEROJATNOSTI

REPREZENTATIVNI UZORAK

– uzorak koji dobro opisuje populaciju

Na reprezentativnost uzorka utječu:

1. Vrsta uzorka (prema metodi odabira)
2. Veličina uzorka
3. Varijabilnost promatranog obilježja

VRSTE UZORAKA

PROBABILISTIČKI (probability samples)

– svaka jedinica promatranja u populaciji ima ***jednaku vjerojatnost izbora*** u uzorak koja je ***različita od 0***

NEPROBABILISTIČKI (non-probability samples)

– ***vjerojatnost izbora*** jedinica promatranja iz populacije je ***različita i nepoznata***

(može biti i 0)

PROBABILISTIČKI UZORCI

JEDNOSTAVNI SLUČAJNI

SUSTAVNI SLUČAJNI (SISTEMATSKI)

SLOJEVITI (STRATIFICIRANI)

UZORAK SKUPINE (*"CLUSTER"*, GROZD)

VIŠEFAZNI

JEDNOSTAVNI SLUČAJNI UZORAK (*simple random sample*)

svojstva:

– ***svaki element*** populacije ima ***jednaku šansu*** da bude izabran
– ***svaki uzorak*** ima ***jednaku šansu*** da bude izabran

način izbora:

– lutrijska metoda
– pomoću tablice slučajnih brojeva
– pomoću programske podrške koja ima funkciju generatora slučajnih brojeva

SUSTAVNI SLUČAJNI UZORAK (*systematic sample*)

– jedinice koje ulaze u uzorak odabiru se po nekakvom pravilu

postupak:

– numerirati jedinice populacije od 1 do N
– odrediti potrebnu veličinu uzorka (n)
– odrediti veličinu intervala k = N/n
– slučajno odabrati broj između 1 i k (početna jedinica)
– uzimati svaku k-tu jedinicu

SLOJEVITI (STRATIFICIRANI) UZORAK

- primjenjuje se u slučajevima kad je promatrano obilježje heterogeno u populaciji
- dobiva se uzimanjem jednostavnih slučajnih uzoraka iz stratuma određenih obzirom na promatrano obilježje

postupak:

- podijeliti populaciju na disjunktne skupine od

  $n_1, n_2, \ldots n_s$

  jedinica, pri čemu je

  $$n_1+n_2+\ldots+n_s = N$$

- uzeti jednostavni slučajni uzorak od

  $$f_i = n_i/N$$

  jedinica iz svake skupine

UZORAK SKUPINE (CLUSTER)

- primjenjuje se u slučajevima kada treba uzeti uzorak iz populacije koja se sastoji od skupina jedinica (ulice, popisni krugovi, škole, općine, ...)

postupak:

- podijeliti populaciju na skupine jedinica
- jednostavnim slučajnim izborom odabrati skupine
- ispitati SVE jedinice unutar odabranih skupina

VIŠEFAZNI UZORAK

– uzimanje "uzorka iz uzorka"
– kombinacija više metoda odabiranja uzorka

Npr. jedan od mogućih načina dobivanja uzorka iz populacije učenika osnovnih škola u Hrvatskoj:

– podijeliti osnovne škole u stratume s obzirom na županijsku pripadnost
– iz svakog stratuma jednostavnim slučajnim izborom odabrati škole (prva faza)
– unutar odabranih škola, jednostavnim slučajnim izborom odabrati razrede (druga faza)
– unutar odabranih razreda, jednostavnim slučajnim izborom odabrati učenike (treća faza)

NEPROBABILISTIČKI UZORCI

PRIGODNI (convenience)

UZORAK KOJI SLUŽI SVRSI (purposive)

UZORAK UDJELA (quota)

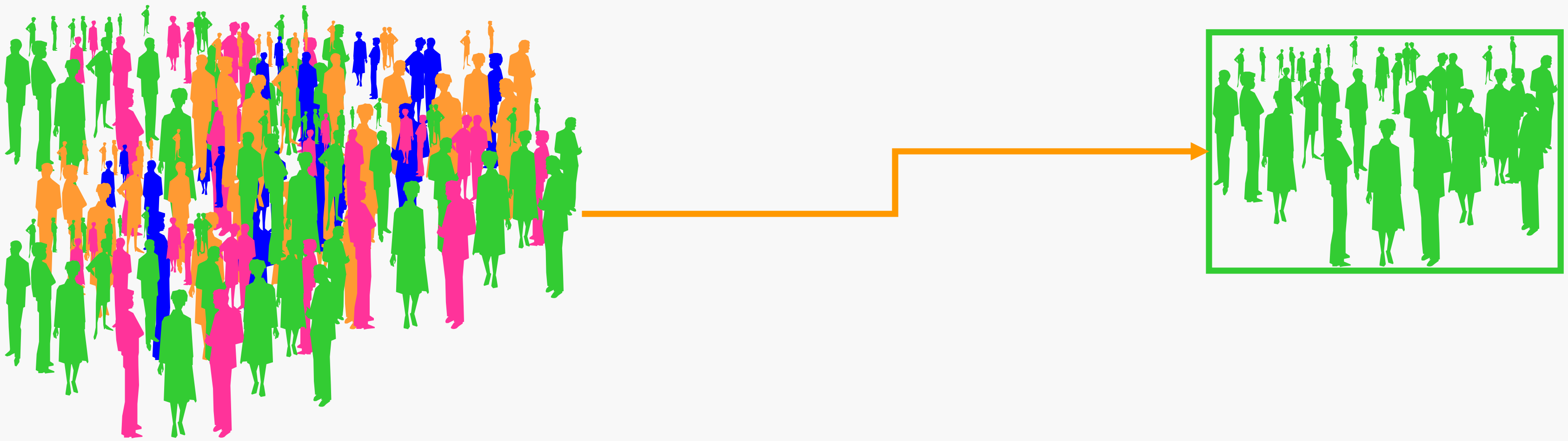
UZORAK SNJEŽNE GRUDE (snowball)

PRIGODNI UZORAK (convenient sample)

- u uzorak se biraju jedinice populacije koje su “pri ruci”:
  - prolaznici
  - pozvani dobrovoljci
  - prvih 50 pacijenata u nekoj ambulanti

UZORAK KOJI SLUŽI SVRSI (purposive sample)

– u uzorak se biraju jedinice populacije koje imaju traženo svojstvo



UZORAK UDJELA (quota sample)

– u uzorak se bira određeni broj jedinica odabranih dijelova populacije

POPULACIJA

Traži se uzorak:

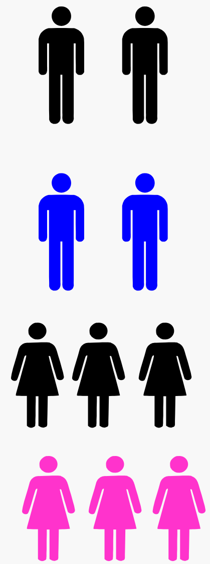
20 %

20 %

30 %

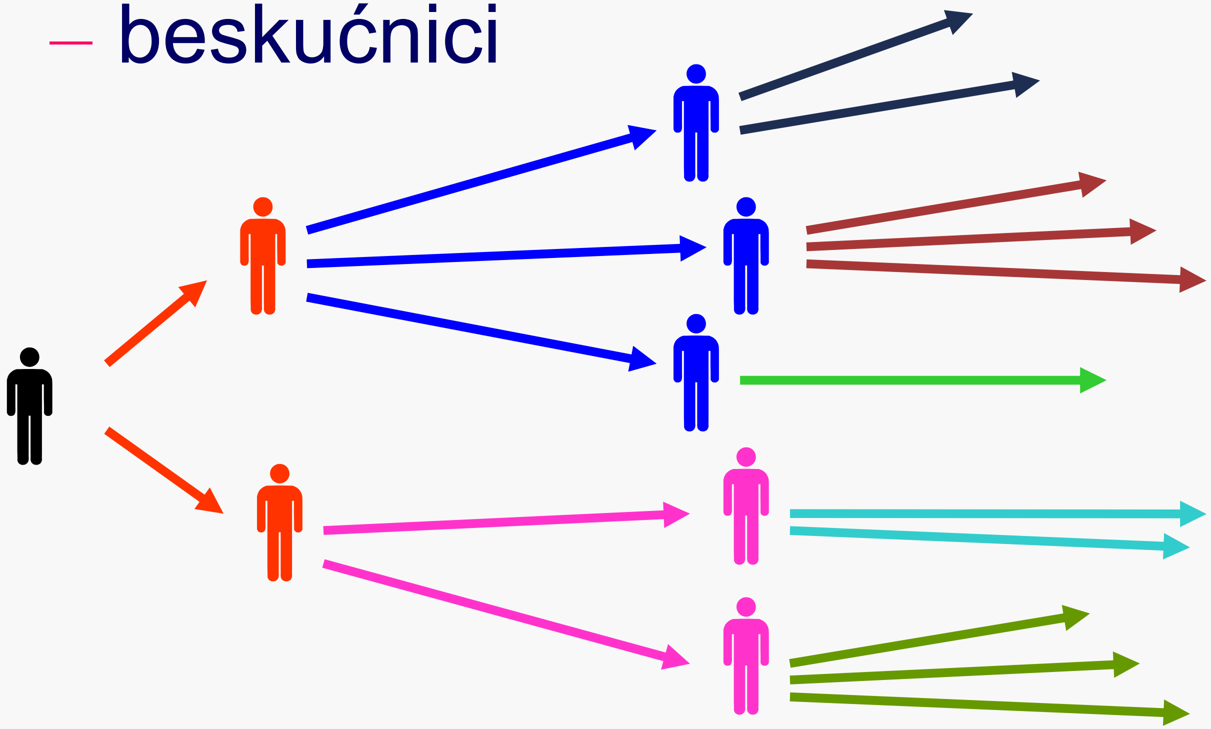
30 %

Uzorak veličine 10:



UZORAK SNJEŽNE GRUDE (snowball sample)

- ispitanici regrutiraju daljnje ispitanike
- zatvorene ili teško dostupne skupine:
  - oboljeli od AIDSa
  - ovisnici
  - beskućnici



U kojem od sljedećeg se koristi jednostavni slučajni uzorak:

a) igra “Bingo”,

b) popis stanovništva,

c) izbori za lokalnu samoupravu?

Koje metode odabira uzorka se koriste u ovim primjerima?

a) igra “Bingo” – jednostavni slučajni uzorak

b) popis stanovništva - ne koristi jednostavni slučajni uzorak jer SVE jedinice populacije moraju biti obuhvaćene popisom

c) izbori za lokalnu samoupravu – neprobabilistički uzorak; na izbore izlaze oni koji to žele (“dobrovoljci”)

VELIČINA UZORKA

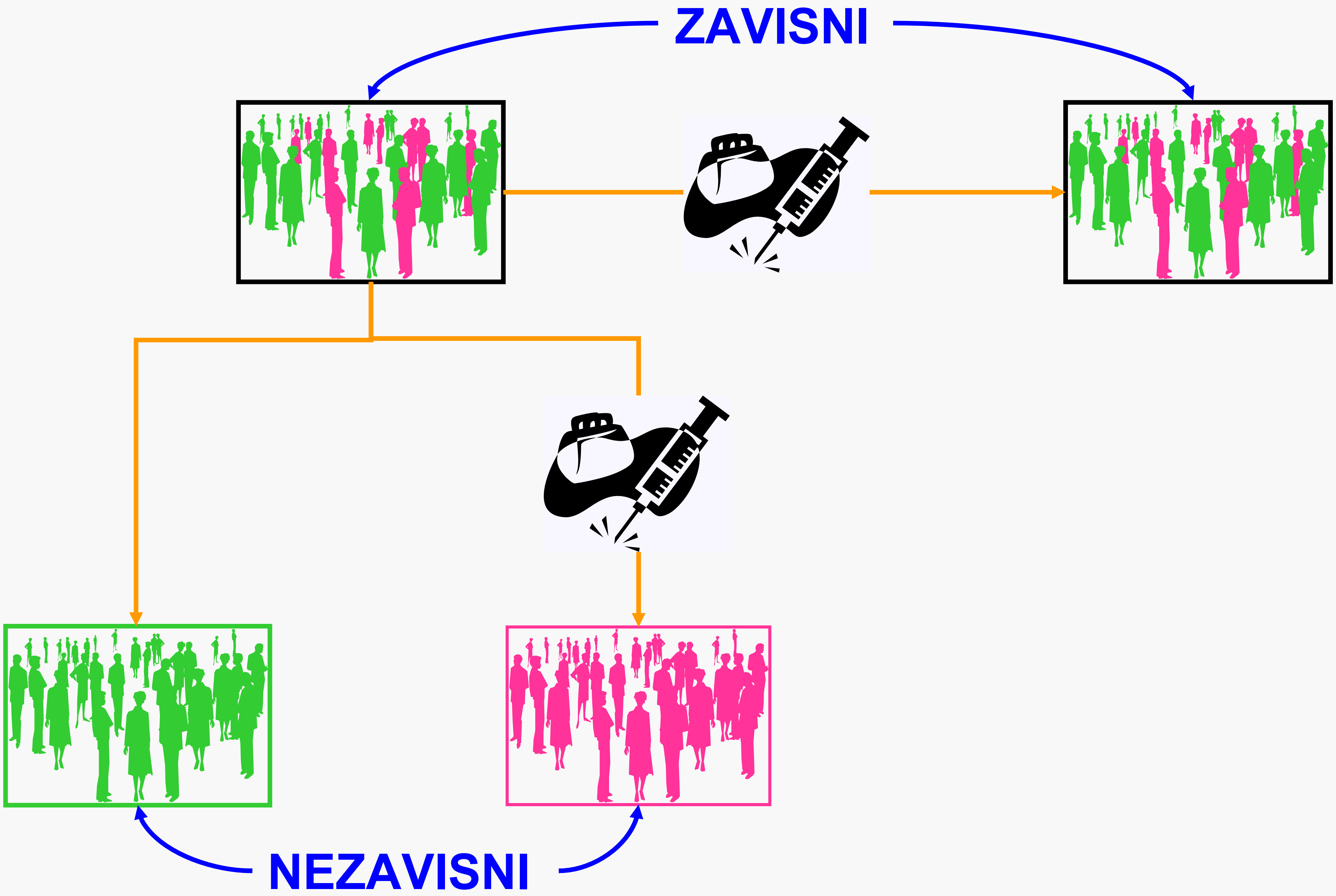
dovoljno veliki uzorak:

– uzorak pomoću kojeg s razumnom pouzdanošću možemo prihvatiti ili odbaciti neku hipotezu i ocijeniti parametar populacije

ovisit će o:

– homogenosti populacije s obzirom na promatrano obilježje
– učestalosti promatranog obilježja (obrnuto proporcionalno)

ZAVISNI I NEZAVISNI UZORCI



Unos podataka o mjerenjima na nezavisnim skupinama

- **nezavisne skupine = različiti ispitanici**

  (ispitanici koji pripadaju nekoj skupini ne pripadaju niti jednoj od preostalih skupina)

- za unos podataka o nekom mjerenju na nezavisnim skupinama ispitanika UVIJEK imamo **2 varijable** (bez obzira koliko je skupina ispitanika):

  1. varijabla koja određuje **pripadnost** ispitanika pojedinoj **skupini**

  2. varijabla u koju unosimo **vrijednost mjerenja** za danog ispitanika

Unos podataka o mjerenjima na nezavisnim skupinama

- npr. mjerenje dobi; skupine po spolu
  - broj mogućih skupina: 2

varijabla koja sadrži vrijednost mjerenja

varijabla koja definira pripadnost skupini

	Dob	Spol	......
ispitanik1	35	M	
ispitanik2	37	M	
ispitanik3	32	M	
ispitanik4	33	Z	

Unos podataka o mjerenjima na nezavisnim skupinama

- npr. mjerenje visine; skupine po razredu (osnovna škola)
  - broj mogućih skupina: 8

varijabla koja sadrži vrijednost mjerenja

varijabla koja definira pripadnost skupini

	Visina	Razred	.....
ispitanik1	110	2	
ispitanik2	140	2	
ispitanik3	100	1	
ispitanik4	176	7	

Unos podataka o mjerenjima na zavisnim skupinama

- zavisne skupine = ponavljana mjerenja na ISTIM ispitanicima

- SVAKO mjerenje = JEDNA varijabla

koliko mjerenja toliko varijabli

Unos podataka o mjerenjima na zavisnim skupinama

- npr. praćenje dnevnih varijacija sistoličkog tlaka; mjerenja u 6h, 10h, 14h, 18h, 22h

po jedna varijabla za svako mjerenje

	ST6	ST10	ST14	ST18	ST22
ispitanik1	120	135	140	180	160
ispitanik2	115	120	120	125	120
ispitanik3	140	145	150	150	180
ispitanik4	118	110	110	115	120

UTJECAJ VARIJABILNOSTI

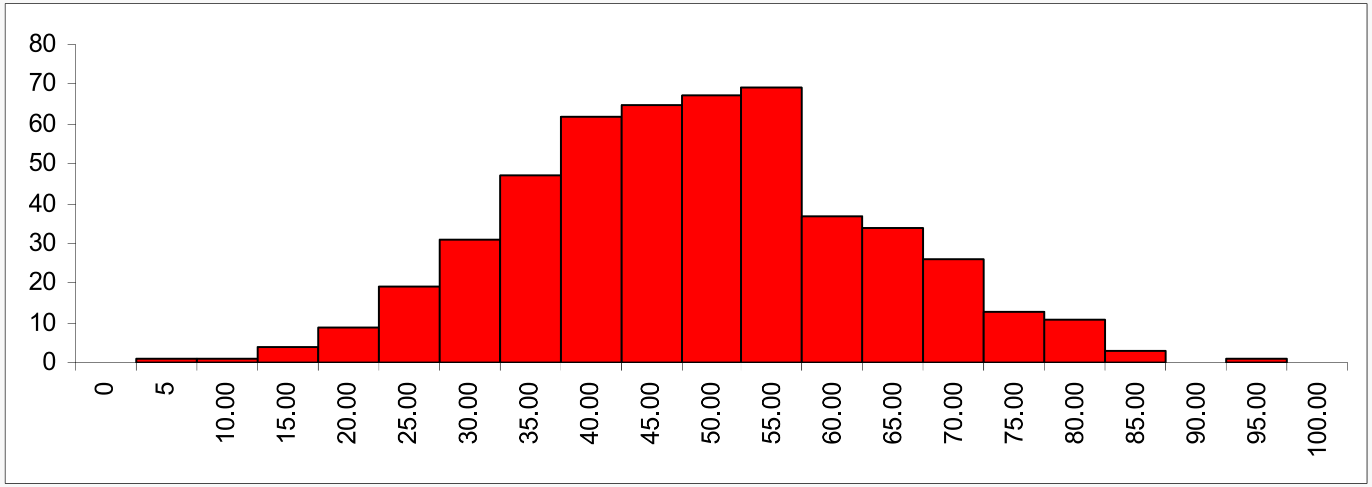
- varijabilnost je često nepoznata
- poznata, a velika varijabilnost ugrožava reprezentativnost uzorka
- utjecaj varijabilnosti se smanjuje s povećanjem uzorka

STANDARDNA POGREŠKA

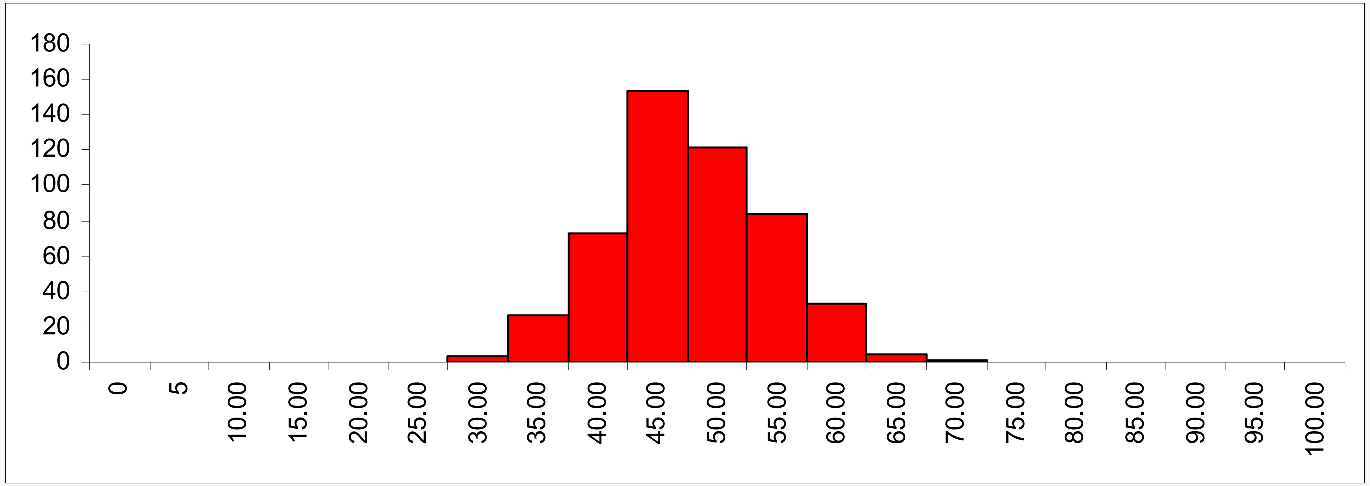
POKUS 1. Napraviti razdiobe aritmetičkih sredina 500 uzoraka veličine n = 4, n = 20 i n = 50 osnovnog skupa N = 101 brojeva od 0 do 100.

aritmetička sredina osnovnog skupa $\mu = 50$

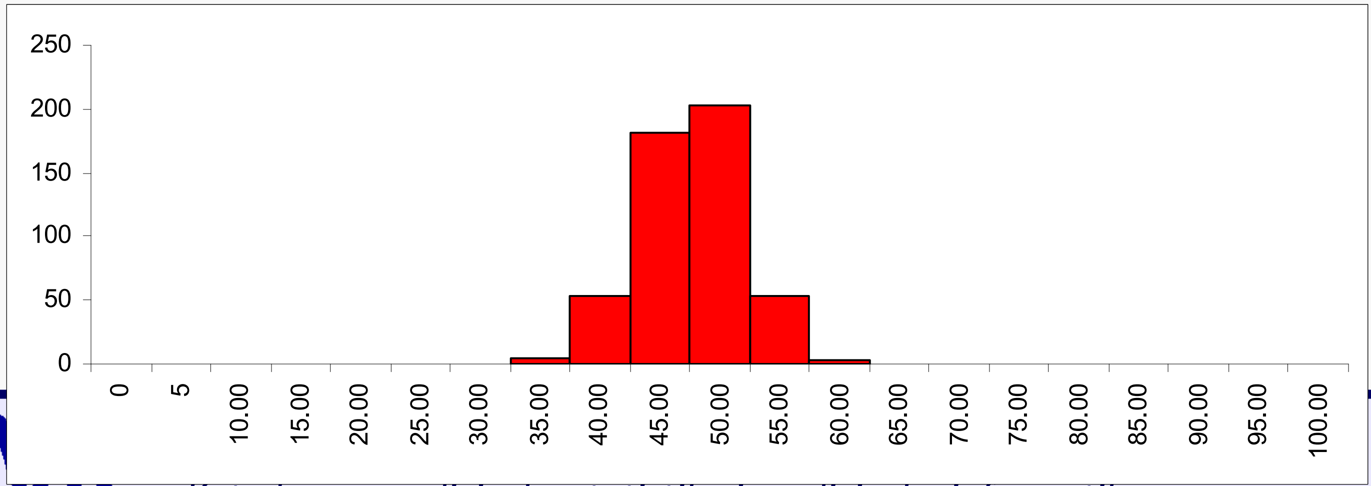
standardna devijacija osnovnog skupa $\sigma = 29{,}15$



$$n = 4$$
$$\overline{x} = 50{,}96$$
$$s = 14{,}451$$



$$n = 20$$
$$\overline{x} = 50{,}10$$
$$s = 6{,}639$$



$$n = 50$$
$$\overline{x} = 50{,}07$$
$$s = 4{,}189$$

N .... veličina osnovnog skupa

n .... veličina slučajnih uzoraka

$\binom{N}{n}$ broj svih mogućih uzoraka veličine **n** uzetih iz osnovnog skupa veličine N

1. uzorak $x_{11}, x_{12}, \ldots, x_{1n}$ sa sredinom $\bar{x}_1$

2. uzorak $x_{21}, x_{22}, \ldots, x_{2n}$ sa sredinom $\bar{x}_2$

3. uzorak $x_{31}, x_{32}, \ldots, x_{3n}$ sa sredinom $\bar{x}_3$

....

k-ti uzorak $x_{k1}, x_{k2}, \ldots, x_{kn}$ sa sredinom $\bar{x}_k$

$$\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots, \bar{x}_k$$

slučajna varijabla
(*sampling distribucija)*

OČEKIVANJE

$$\boxed{E(\bar{x}) = \mu}$$

VARIJANCA

$$\boxed{V(\bar{x}) = \frac{\sigma^2}{n}}$$

standardna devijacija

$$\boxed{s_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$$

STANDARDNA POGREŠKA (SE, standard error)

- za slučajne i dovoljno velike uzorke

$$\boxed{s_{\bar{x}} = \frac{s}{\sqrt{n}}}$$

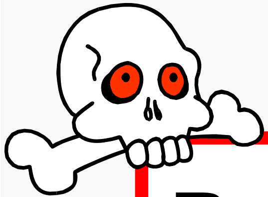
STANDARDNA POGREŠKA (SE, standard error)

- standardna pogreška aritmetičke sredine (SEM, *standard error of the mean*)
- pogreška kojoj se izlažemo pri zaključivanju o populaciji na temelju uzorka

$s \uparrow \Rightarrow s_{\bar{x}} \uparrow$ ***povećava se*** s povećanjem *varijabilnosti obilježja*

$n \uparrow \Rightarrow s_{\bar{x}} \downarrow$ ***smanjuje se*** s povećanjem *veličine uzorka*

CENTRALNI GRANIČNI TEOREM



Razdioba aritmetičkih sredina uzoraka teži normalnoj razdiobi s očekivanjem μ i varijancom $\sigma^2_{\bar{x}}$ $\left[N\left(\mu,\sigma^2_{\bar{x}}\right)\right]$ kada veličina uzorka *n* teži u beskonačnost.

$\Rightarrow$ ***za dovoljno velike uzorke*** razdioba aritmetičkih sredina uzoraka bit će ***normalna***, bez obzira na razdiobu vrijednosti promatranog obilježja za proporciju:

$$\boxed{s_p = \sqrt{\frac{p \cdot q}{n}}}$$

STANDARDNA POGREŠKA PROPORCIJE

$p \uparrow \Rightarrow s_{\bar{x}} \downarrow$ ***smanjuje se*** s povećanjem *homogenosti obilježja*

$n \uparrow \Rightarrow s_{\bar{x}} \downarrow$ ***smanjuje se*** s povećanjem *veličine uzorka*

STANDARDNA POGREŠKA
vs
STANDARDNA DEVIJACIJA

STANDARDNA POGREŠKA:

- procjenjuje "kvalitetu" ocjene parametra (statistike)
- velika standardna pogreška => ocjena parametra (ar. sredina, proporcija) je neprecizna

STANDARDNA DEVIJACIJA:

- opisuje varijabilnost podataka
- velika standardna devijacija => velika varijabilnost podataka

RASPON POUZDANOSTI

RASPON POUZDANOSTI

- confidence interval (CI)
- uobičajeno tumačenje:

  raspon unutar kojega se, s određenom vjerojatnošću, nalazi prava vrijednost (parametar) populacije

RASPON POUZDANOSTI ARITMETIČKE SREDINE

$$\boxed{\bar{x} - z \cdot s_{\bar{x}} \leq \mu \leq \bar{x} + z \cdot s_{\bar{x}}}$$

RASPON POUZDANOSTI PROPORCIJE

$$\boxed{p - z \cdot s_{p} \leq \Pi \leq p + z \cdot s_{p}}$$

z - standardizirana vrijednost normalne raspodjele

(ovisi o pretpostavljenoj vjerojatnosti)

PRIMJER. Koliki je raspon pouzdanosti ako želimo obuhvatiti μ sa:

a) 99 % pouzdanosti

b) 95 % pouzdanosti

c) 90 % pouzdanosti

uz pretpostavku normalne razdiobe?

a) $z_{0,005} = 2,576 \approx 2,58$ $$\bar{x} - 2,58 \cdot s_{\bar{x}} \leq \mu \leq \bar{x} + 2,58 \cdot s_{\bar{x}}$$

b) $z_{0,025} = 1,96$ $$\bar{x} - 1,96 \cdot s_{\bar{x}} \leq \mu \leq \bar{x} + 1,96 \cdot s_{\bar{x}}$$

c) $z_{0,05} = 1,65$ $$\bar{x} - 1,65 \cdot s_{\bar{x}} \leq \mu \leq \bar{x} + 1,65 \cdot s_{\bar{x}}$$

PRIMJER. Od 1000 ljudi koji su cijepljeni, 200 ih je pokazalo alergične reakcije. Koliku proporciju alergičnih očekujemo u populaciji cijepljenih uz vjerojatnost od 95 %?

$z_{0,025}=1{,}96$

$$p-1{,}96\cdot s_p \leq \Pi \leq p+1{,}96\cdot s_p$$

$p=0{,}20;$ $q=0{,}80;$

$$s_p=\sqrt{\frac{0{,}2\cdot 0{,}8}{1000}}=\sqrt{\frac{0{,}16}{1000}}=\sqrt{0{,}00016}=0{,}0126$$

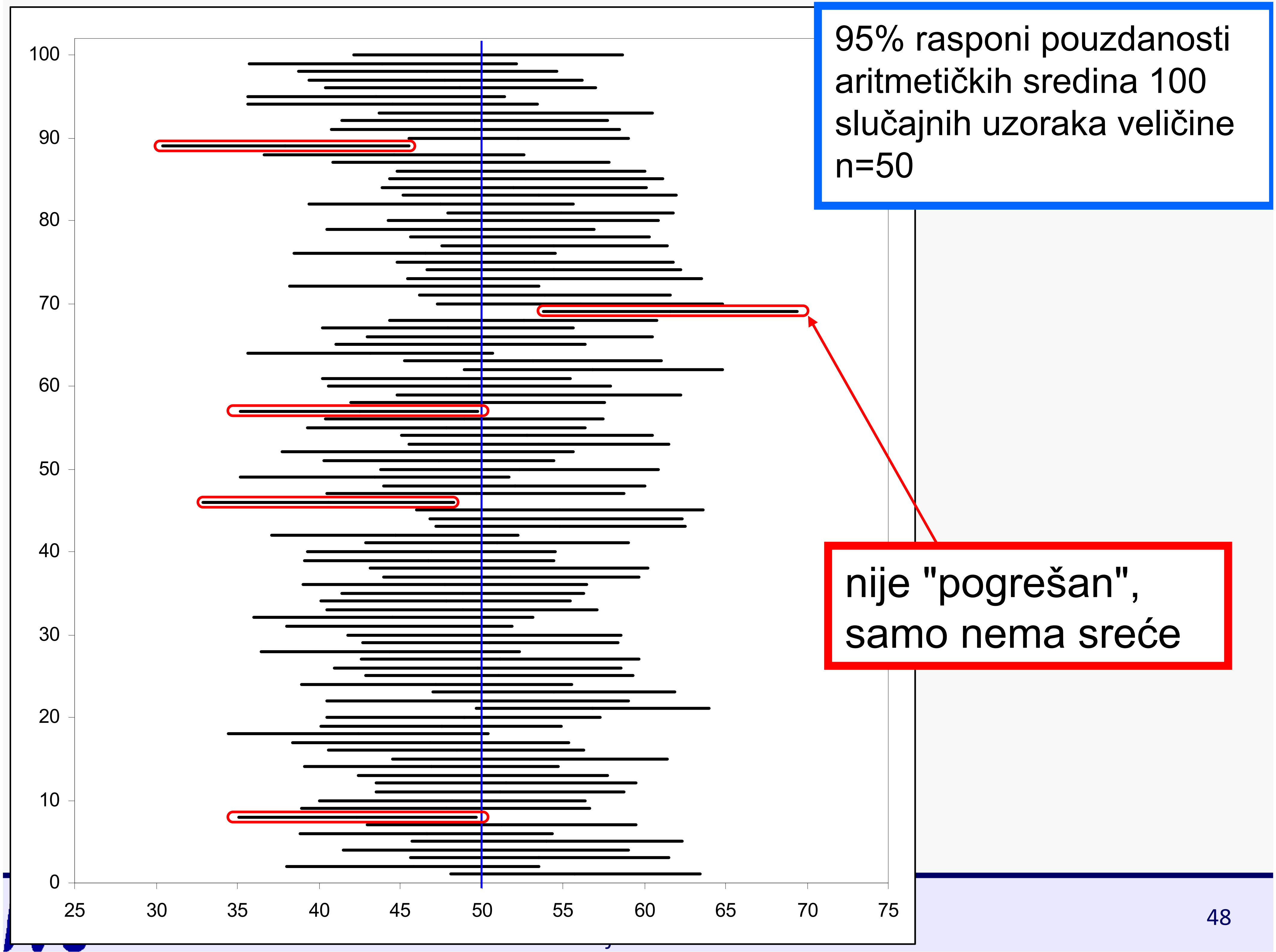
$$0{,}2-1{,}96\cdot 0{,}0126 \leq \Pi \leq 0{,}2+1{,}96\cdot 0{,}0126$$

$$0{,}2-0{,}025 \leq \Pi \leq 0{,}2+0{,}025$$

$$0{,}175 \leq \Pi \leq 0{,}225$$

95% raspon pouzdanosti aritmetičke sredine izračunat iz nekog uzorka:

- ***uobičajeno*** se tumači kao raspon vrijednosti unutar kojeg se s 95% pouzdanosti nalazi prava vrijednost aritmetičke sredine (aritmetička sredina populacije)
- ***u stvari znači*** da očekujemo da 95% takvih intervala dobivenih iz uzoraka iste veličine dane populacije uključuje pravu vrijednost aritmetičke sredine


95% rasponi pouzdanosti aritmetičkih sredina 100 slučajnih uzoraka veličine n=50
nije "pogrešan", samo nema sreće
0
10
20
30
40
50
60
70
80
90
100
25
30
35
40
45
50
55
60
65
70
75

95% raspon pouzdanosti:

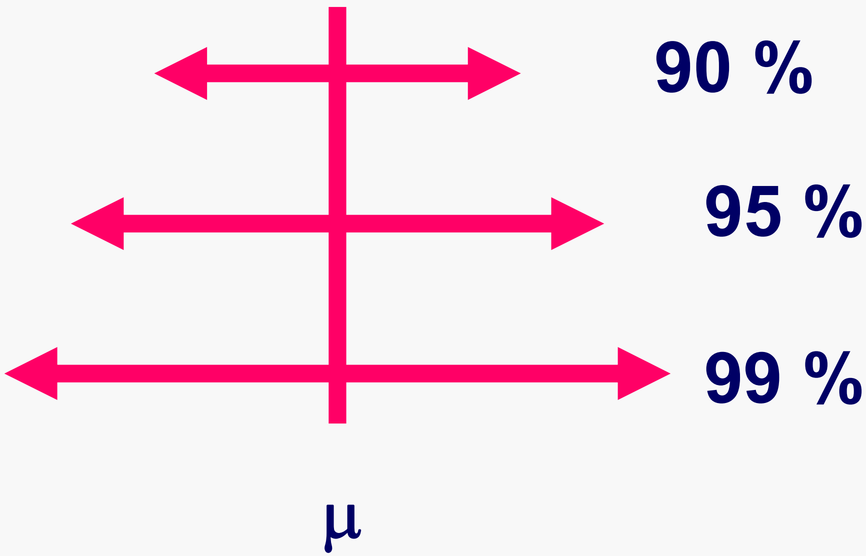
Slučajan interval čije granice se mogu izračunati iz podataka o uzorku, takav da 95 od svakih 100 takvih intervala obuhvaća pravu vrijednost parametra koji se procjenjuje.

- takodjer i raspon poželjnih vrijednosti parametra populacije (prihvatljiva nul-hipoteza)

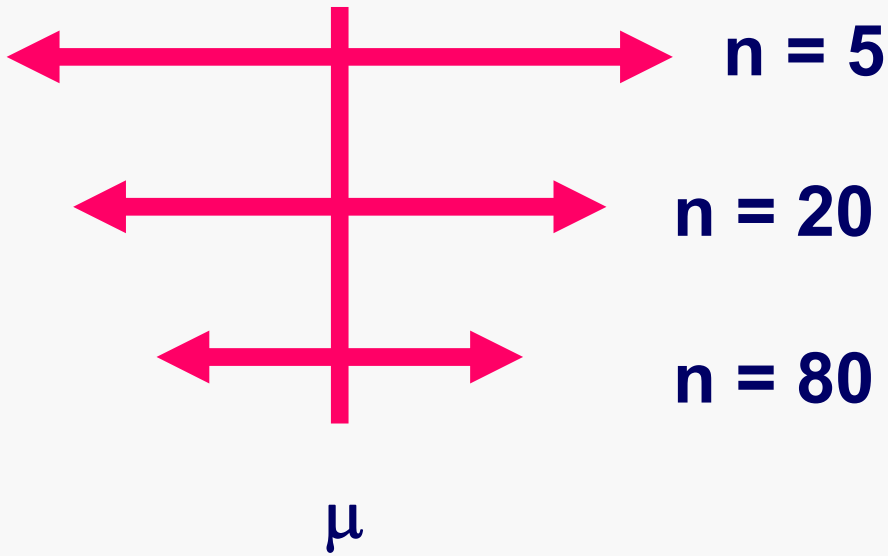
Širina raspona pouzdanosti ovisi o:

- **pretpostavljenoj vjerojatnosti**
- **varijabilnosti promatranog obilježja**
- **veličini uzorka**

širi se s povećanjem pouzdanosti



sužava se s povećanjem uzorka



POTREBNA VELIČINA UZORKA
za procjenu aritmetičke sredine

Ovisit će o:

- pogrešci procjene koju ćemo tolerirati
- stupnju pouzdanosti
- pretpostavljenoj varijabilnosti

$$\boxed{E = z \cdot s_{\bar{x}} = z \cdot \frac{s}{\sqrt{n}}}$$

POGREŠKA PROCJENE

$$\boxed{n = \left(\frac{z \cdot s}{E} \right)^2}$$

POTREBNA VELIČINA UZORKA

PRIMJER. Koliko ispitanika treba izabrati u uzorak kako bi se procijenila prosječna starost stanovnika nekog sela u 95 % rasponu pouzdanosti od 2 godine? Pretpostavlja se kako je standardna devijacija populacije 8 godina.

Koliki uzorak treba biti ako toleriramo pogrešku od najviše ± 10 mjeseci?

$z_{0,025} = 1{,}96 \qquad E = 1 \qquad \sigma = 8$

$$n = \left(\frac{z_{0,025} \cdot 8}{1}\right)^2 = \left(\frac{1{,}96 \cdot 8}{1}\right)^2 = 15{,}68^2 = 245{,}86 \approx 246$$

$z_{0,025} = 1{,}96 \qquad E = 10/12 = 0{,}83 \qquad \sigma = 8$

$$n = \left(\frac{z_{0,025} \cdot 8}{0{,}83}\right)^2 = \left(\frac{1{,}96 \cdot 8}{0{,}83}\right)^2 = 18{,}89^2 = 356{,}83 \approx 357$$

POTREBNA VELIČINA UZORKA
za procjenu proporcije

Ovisit će o:

- pogrešci procjene koju ćemo tolerirati
- stupnju pouzdanosti
- pretpostavljenoj proporciji

$$\boxed{E = z \cdot s_p = z \cdot \sqrt{\frac{p \cdot q}{n}}}$$

POGREŠKA PROCJENE

$$\boxed{n = \left(\frac{z}{E}\right)^2 \cdot p \cdot q}$$

POTREBNA VELIČINA UZORKA

PRIMJER. Studija provedena na Fakultetu javnog zdravstva na Harvardu utvrdila je da 19 % studenata nikada ne piju alkohol. Koliki uzorak vam je potreban za procjenu proporcije studenata koji ne piju alkohol na vašem fakultetu unutar raspona od 10 % uz pouzdanost od 95 %, vodeći se rezultatima harvardske studije?

$z_{0,025} = 1,96$ $\quad E = 0,1/2 = 0,05$ $\quad p = 0,19$

$$n = \left(\frac{z}{E}\right)^2 \cdot p \cdot q = \left(\frac{1,96}{0,05}\right)^2 \cdot 0,19 \cdot 0,81 = 39,3^2 \cdot 0,19 \cdot 0,81 = 236,49 \approx 237$$

p = 0,5 koristimo kada nemamo prethodnih saznanja o pretpostavljenoj proporciji

$z_{0,025} = 1,96$ $E = 0,1/2 = 0,05$ $p = 0,5$

$$n = \left(\frac{z}{E}\right)^2 \cdot p \cdot q = \left(\frac{1,96}{0,05}\right)^2 \cdot 0,5 \cdot 0,5 = 39,3^2 \cdot 0,25 = 384,15 \approx 385$$

STATISTIČKI TESTOVI

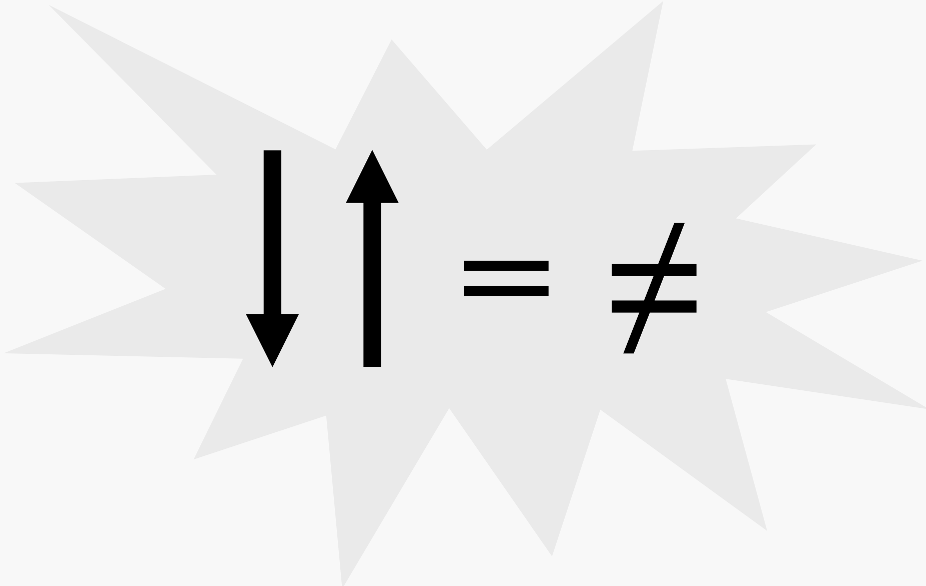
"Severely dependent patients had a longer duration of the disease ($p<0.001$) and a longer duration of stay at a nursing home ($p=0.001$) than mildly dependent patients."

"The addicts perceived their mothers as more rejecting ($p=0.018$ for total score), more aggressive ($p=0.007$), and showing more undifferentiated rejection ($p=0.001$) than non-addicts."

STATISTIČKI TEST

– postupak pomoću kojeg se dolazi do **odluke** o prihvaćanju ili odbacivanju **statističke hipoteze** uz **određenu vjerojatnost**

HIPOTEZA ISTRAŽIVANJA



STATISTIČKA HIPOTEZA

HIPOTEZA ISTRAŽIVANJA

– **pretpostavka (slutnja) o nekoj populaciji/populacijama koja motivira istraživanje**

HIPOTEZA ISTRAŽIVANJA ...cont.

- medicinske sestre/tehničari mlađih dobnih skupina imaju pozitivniji stav prema uvođenju IT u odnosu na medicinske sestre/tehničare starijih dobnih skupina
- oboljeli od KOBP uključeni u XY program rehabilitacije imaju veće funkcionalne sposobnosti od bolesnika na standardnom tretmanu KOBP
- osobe oboljele od dijabetesa imaju povišen sistolički tlak

STATISTIČKA HIPOTEZA

- izjava (tvrdnja) o nekoj karakteristici (parametru) populacije
- izvodi se iz hipoteze istraživanja
- matematički oblikovana

- može se vrednovati odgovarajućim statističkim postupcima
- prihvaćamo ju ili odbacujemo na osnovu informacija dobivenih iz podataka prikupljenih na uzorku.

STATISTIČKI TEST

NUL-HIPOTEZA
(H_0)

ALTERNATIVNA
HIPOTEZA
(H_1)

NUL-HIPOTEZA

- polazna hipoteza koja se testira
- "hipoteza o nepostojanju razlike"

$$\boxed{H_0 \;....\; \mu_1 = \mu_2}$$

ALTERNATIVNA HIPOTEZA

- negacija nul-hipoteze

$$\boxed{H_1 \;....\; \mu_1 \neq \mu_2}$$

POSTUPAK:

POSTAVLJANJE NUL-HIPOTEZE

I

ALTERNATIVNE HIPOTEZE

PRIKUPLJANJE PODATAKA

TESTIRANJE

DONOŠENJE ODLUKE

POSTAVLJANJE NUL- HIPOTEZE I ALTERNATIVNE HIPOTEZE

– odnose se na neki parametar populacije (sredina, varijanca,...)

– zajedno, moraju obuhvatiti sve moguće odnose promatranih parametara

npr.

H_0 .... $\mu_1 = \mu_2$

$\Rightarrow$ parametri populacija iz kojih su uzorci uzeti su jednaki

$\Rightarrow$ uzorci pripadaju istoj populaciji

H_1 .... $\mu_1 \neq \mu_2$

TESTIRANJE

= izračunavanje odgovarajuće test-statistike

vrijednost u uvjetima istinitosti nul-hipoteze

$$\textbf{test statistika} = \frac{\textbf{opažena vrijednost} - \textbf{hipotetska vrijednost}}{\textbf{standardna pogreška opažene vrijednosti}}$$

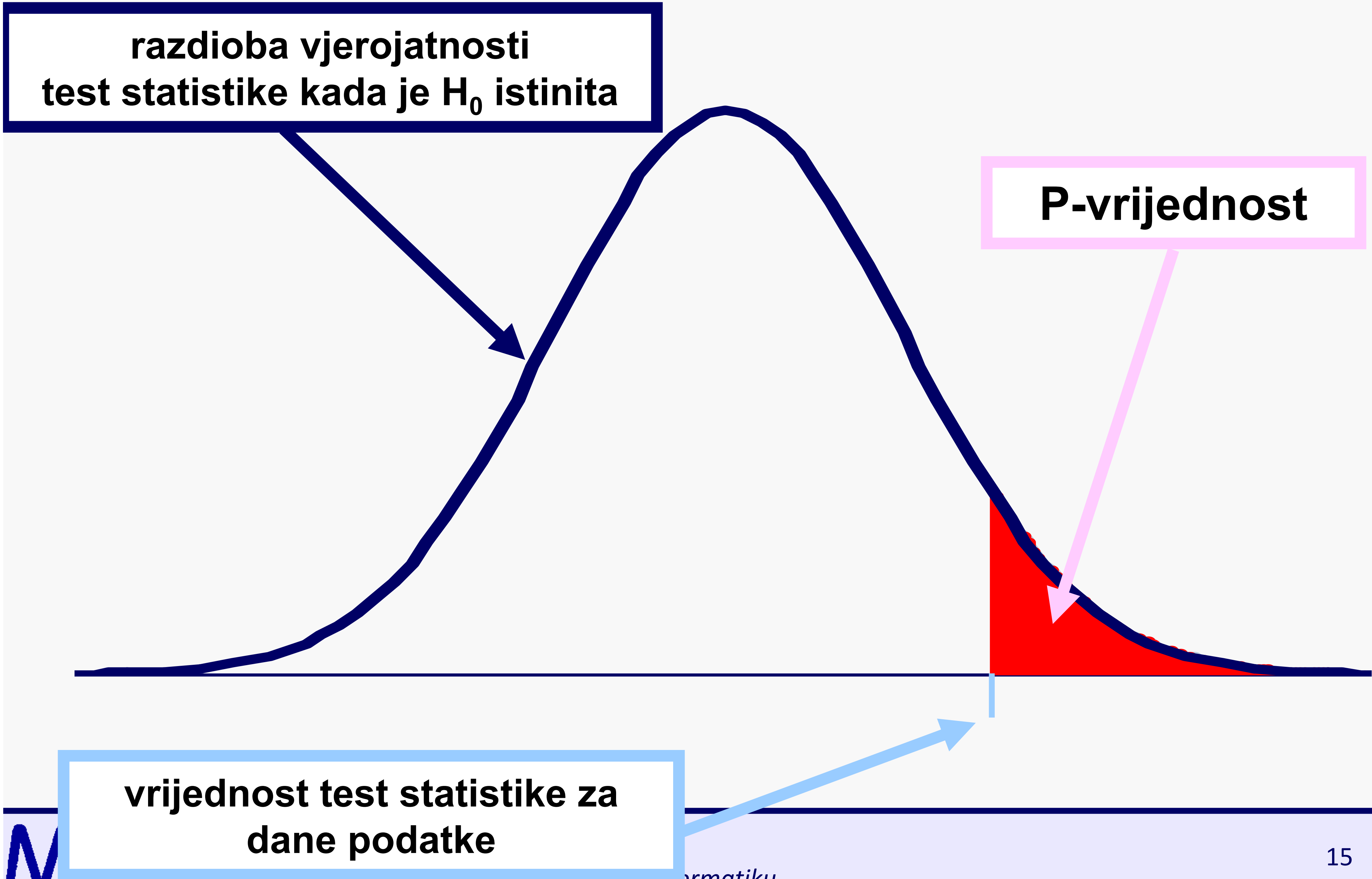
npr.

$$\boxed{\text{test statistika} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{SE(\bar{x}_1 - \bar{x}_2)}}$$

uz $H_0 \ldots\ldots \mu_1 = \mu_2 \quad \Rightarrow \quad \mu_1 - \mu_2 = 0$

$$\boxed{\text{test statistika} = \frac{(\bar{x}_1 - \bar{x}_2)}{SE(\bar{x}_1 - \bar{x}_2)}}$$

P-vrijednost



ŠTO JE P-VRIJEDNOST?

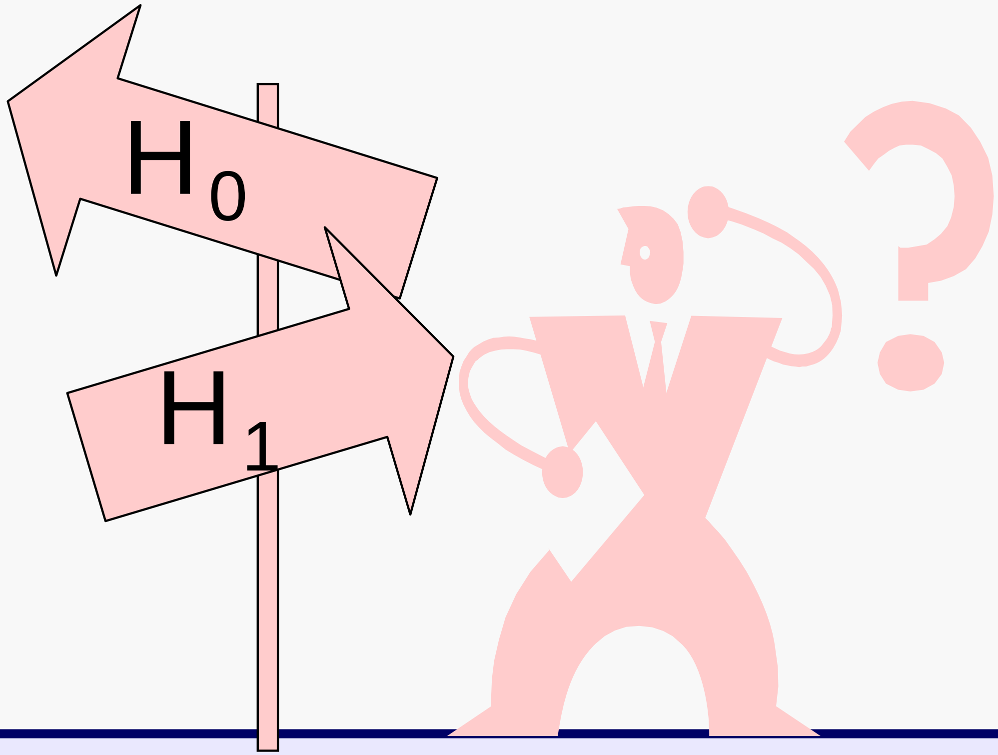
- NIJE vjerojatnost istinitosti nul-hipoteze (iako je vrlo slično)
- JESTE vjerojatnost dobivanja istih ili ekstremnijih rezultata kada je nul-hipoteza istinita

DONOŠENJE ODLUKE

o odbacivanju H_0

ili

prihvaćanju H_0



POGREŠKE PRI ODLUČIVANJU

		STVARNO STANJE	
		H_0 točna	H_1 točna
ODLUKA	PRIHVATI H_0	ISPRAVNO	POGREŠKA TIPA 2 (β)
	ODBACI H_0	POGREŠKA TIPA 1 (α)	ISPRAVNO

VJEROJATNOSTI POGREŠKE

Najveća vjerojatnost pogreške tipa 1 (α)

- *razina značajnosti*
- najmanja vjerojatnost uz koju još prihvaćamo H_0
- kada je $P < \alpha$, test sugerira odbacivanje H_0 (“statistički značajno”)
- određuje ju istraživač na temelju modela pokusa

Najveća vjerojatnost pogreške tipa 2 (β)

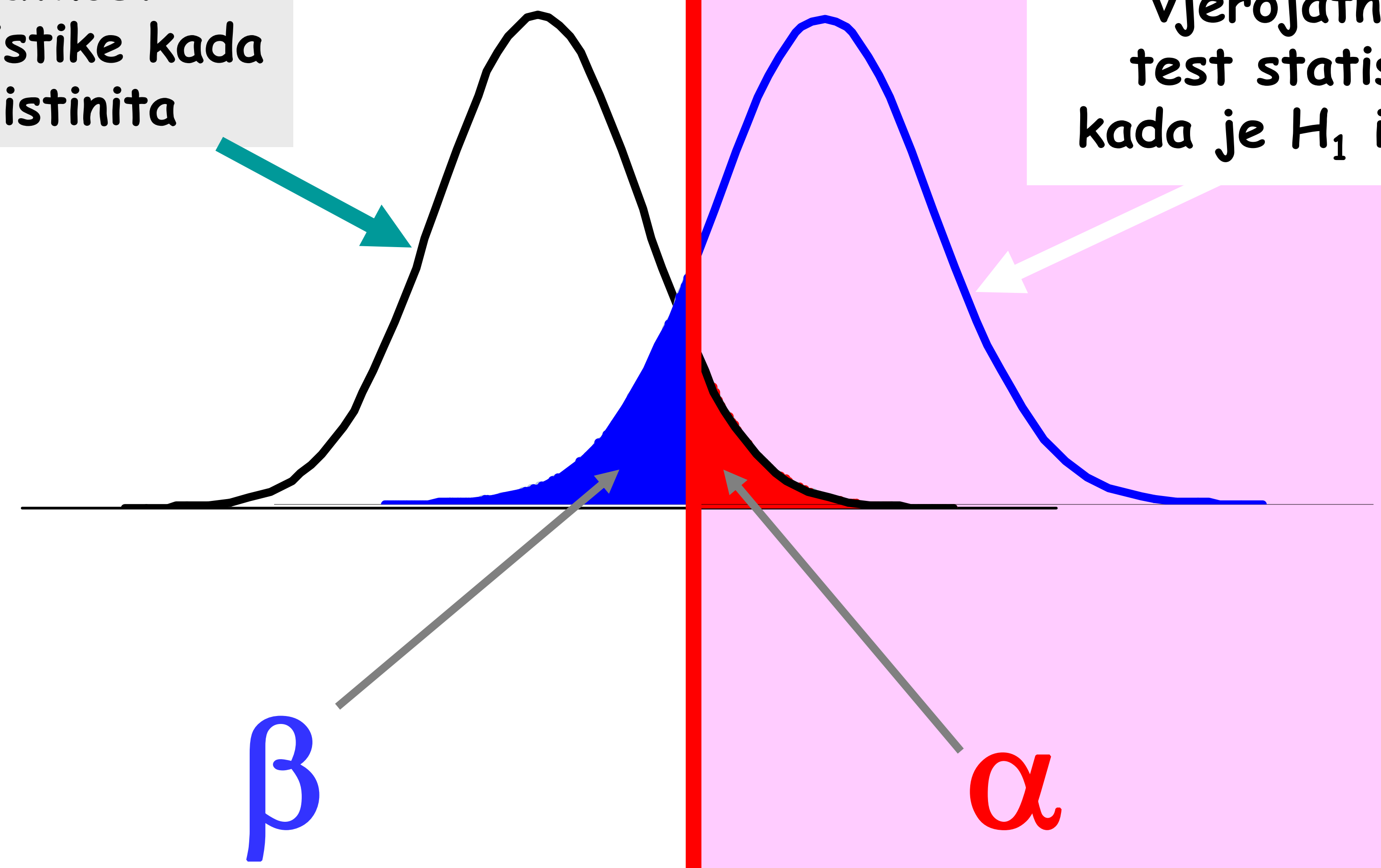
- djelomično je pod kontrolom
- ovisi o:
  - stvarnom stanju u populaciji (varijabilitet)
  - efektu od interesa
  - razini značajnosti a
- α i β su inverzno povezane (ali ne direktno)

PODRUČJE PRIHVAĆANJA H_0

PODRUČJE ODBACIVANJA H_0

razdioba vjerojatnosti test statistike kada je H_0 istinita

razdioba vjerojatnosti test statistike kada je H_1 istinita



ODABIR NIVOA ZNAČAJNOSTI

Pitanje štetnih posljedica pogreške:

1. Odluka/zaključak da razlike postoje onda kada ih u stvarnosti nema može prouzročiti štetne posljedice => ***smanjiti vjerojatnost nastajanja pogreške tipa 1, tj. odabrati manji*** α

2. Odluka/zaključak da nema razlike onda kada u stvarnosti razlika postoji može prouzročiti štetne posljedice => ***smanjiti vjerojatnost pogreške tipa 2, tj. odabrati veći*** α

Ispitivanja lijeka X pokazala su da njegovo korištenje izaziva vrlo štetne posljedice te je lijek X povučen iz uporabe.

Ispitan je novi alternativni lijek Y i ustanovljeno je smanjenje štetnog utjecaja u odnosu na lijek X.

Koju razinu značajnosti treba upotrijebiti za ocjenu značajnosti smanjenja štetnog utjecaja lijeka Y u odnosu na lijek X?

STVARNO STANJE: Oba lijeka jednako su štetna.

ODLUKA: Lijek Y ima manje štetne posljedice od lijeka X.

α

STVARNO STANJE: Lijek Y manje je štetan od lijeka X.

ODLUKA: Lijek Y ima jednako štetne posljedice kao i lijek X.

β

STVARNO STANJE: Oba lijeka jednako su štetna.

ODLUKA: Lijek Y ima manje štetne posljedice od lijeka X.

STVARNO STANJE: Lijek Y manje je štetan od lijeka X.

ODLUKA: Lijek Y ima jednako štetne posljedice kao i lijek X.

α

β

Na slučajnom uzorku vozača ispitivan je utjecaj alkohola na vrijeme reagiranja. Mjerenja vremena reakcije prije i nakon konzumacije određene količine alkohola pokazala su prosječno povećanje vremena reakcije nakon konzumacije alkohola.

Koju razinu značajnosti treba upotrijebiti za ocjenu značajnosti pronađene razlike?

STVARNO STANJE:
Alkohol ne utječe na vrijeme reakcije.

ODLUKA:
Alkohol produljuje vrijeme reakcije

STVARNO STANJE:
Alkohol produljuje vrijeme reakcije.

ODLUKA:
Alkohol ne utječe na vrijeme reakcije.

α

β


STVARNO STANJE: Alkohol ne utječe na vrijeme reakcije.
ODLUKA: Alkohol produljuje vrijeme reakcije
STVARNO STANJE: Alkohol produljuje vrijeme reakcije.
ODLUKA: Alkohol ne utječe na vrijeme reakcije.
α
β

POSTAVKE DIZAJNA

- općenito testove treba dizajnirati tako da imaju

$$\beta \geq \alpha$$

a gdje je odabrani β 0.2 ili 0.1

- izraz

$$100 \cdot (1-\beta)\%$$

naziva se *(statistička) SNAGA TESTA*

SNAGA TESTA

- šansa da se detektira određena alternativna hipoteza kada je stvarno točna
- **NEETIČNO je (a i skupo!) raditi istraživanja male snage !**

ŠTO I KAKO UTJEČE NA SNAGU TESTA

Statistička značajnost

NIJE isto što i

klinička važnost!

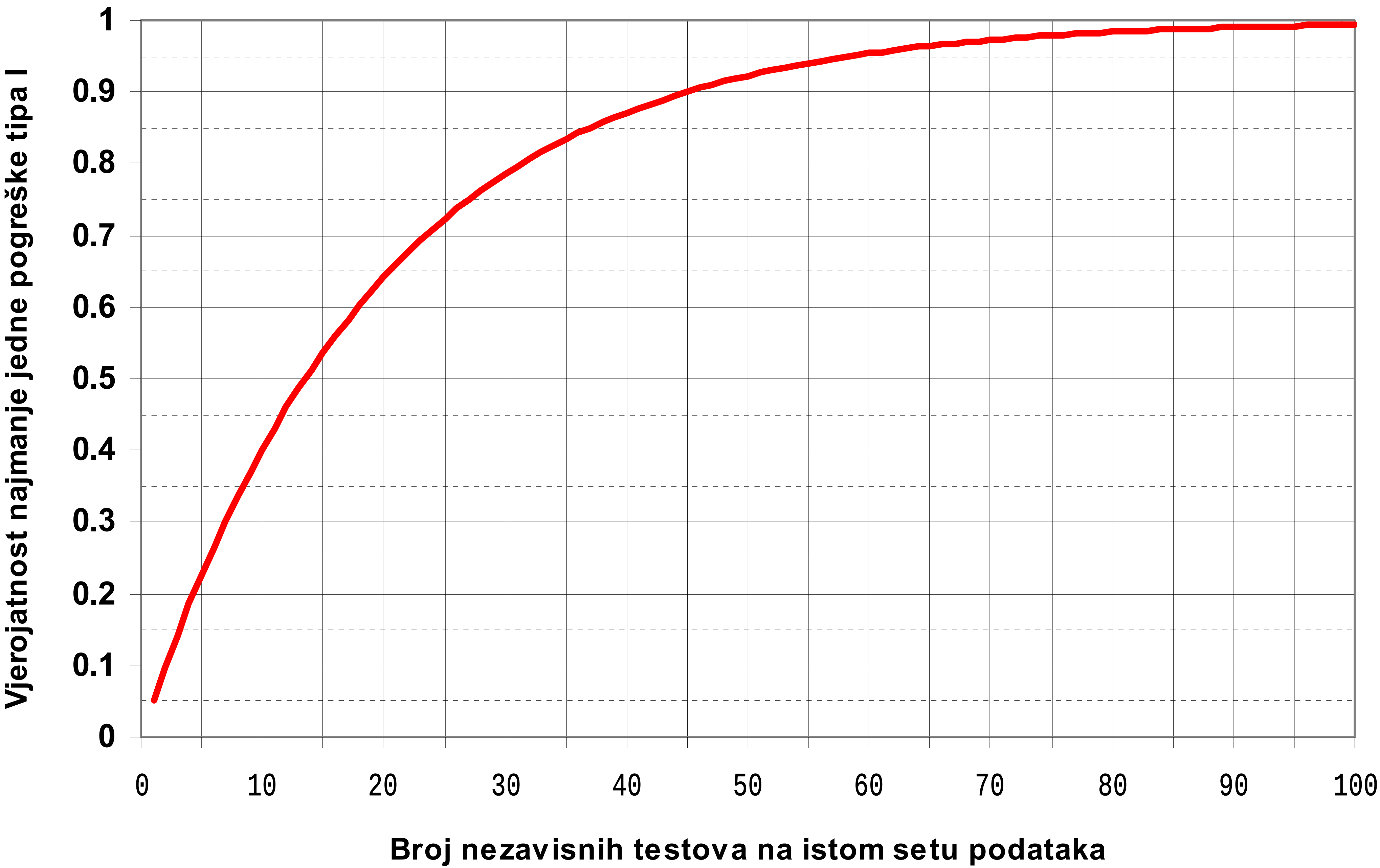
VIŠESTRUKA TESTIRANJA

- ugrožavaju valjanost
- povećavaju pogrešku tipa 1

$$\boxed{\alpha_r = 1 - (1 - \alpha)^r}$$

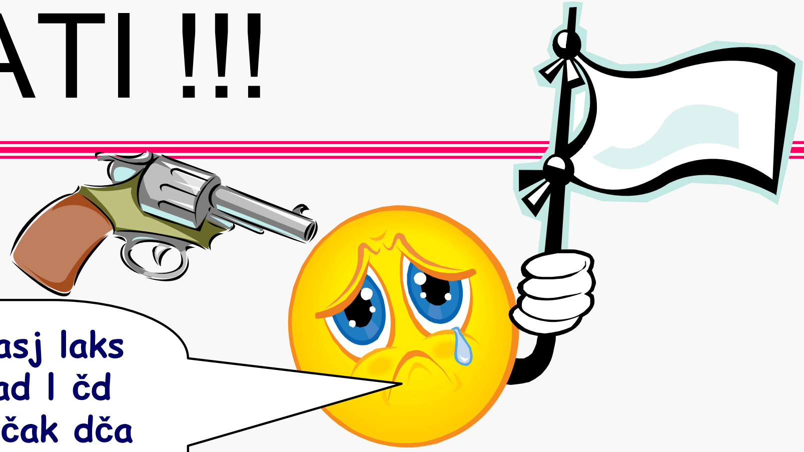
za $\alpha = 0.05$ i **r** višestrukih testova

r	**1**	**2**	**3**	**4**	**5**	**10**	**15**	**20**
α_r	**0.050**	**0.098**	**0.143**	**0.185**	**0.226**	**0.401**	**0.537**	**0.642**

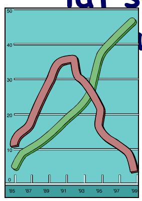


Vjerojatnost najmanje jedne pogreške tipa I
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
10
20
30
40
50
60
70
80
90
100
Broj nezavisnih testova na istom setu podataka

AKO MUČITE PODATKE DOVOLJNO DUGO ONI ĆE NAPOSLIJETKU PRIZNATI !!!



akl lk laj la dai jdapo d a p čjd asj laks
lai sa da ljd aj dlajd lajd lajd lad l čd
lk dčadk ač dčad čak dča sčak dča
sčdak sdakć sdčka ćsd aćsd ćadsl
ćal dsćalds ćal dsćalds ća da



P<0,05

VIŠESTRUKA TESTIRANJA ....cont.

- rješenje:
  - prilagodba P-vrijednosti u cilju održavanja općeg nivoa značajnosti (Bonferroni, Sidak, Hochberg...)
  - primjena složenijih metoda analize (npr. ANOVA, multivarijatne metode)

IZBOR STATISTIČKOG TESTA

Ne ovisi u velikoj mjeri o veličini uzorka nego:

- **prirodi (tipu i raspodjeli) varijabli**
- **broju uzoraka (1, 2 ili više)**
- **jesu li su uzorci zavisni ili ne**

		VARIJABLA		
BROJ UZORAKA		NOMINALNA	ORDINALNA ILI NUMERIČKA KOJA NIJE NORMALNO DISTRIBUIRANA	NUMERIČKA NORMALNO DISTRIBUIRANA
JEDAN		χ^2-test	Kolmogorov-Smirnov test	t-test
DVA	NEZAVISNI	χ^2-test Fisherov egzaktni test	Mann-Whitney U test Medijan test	Studentov t-test
	ZAVISNI	McNemarov test	Wilcoxonov test	t-test diff.
VIŠE OD 2	NEZAVISNI	χ^2-test	Kruskall-Wallis test	ANOVA
	ZAVISNI	Cochran Q Stuart-Maxwell	Friedmanov test	ANOVA za ponavljana mjerenja
POVEZANOST DVIJU VARIJABLI		Koef. kontingencije Kappa koef.	Spermanov r Kendalov τ	Pearsonov r

POSTAVLJANJE H_0 i H_1

izbor α i $(1-\beta)$

određivanje veličine uzorka POTREBNE da se uz ODABRANE α i $(1-\beta)$ detektira ŽELJENI efekt

prikupljanje primjerenih podataka

izbor odgovarajućeg testa

računanje test statistike

određivanje odgovarajuće P-vrijednosti

DONOŠENJE ODLUKE - TUMAČENJE