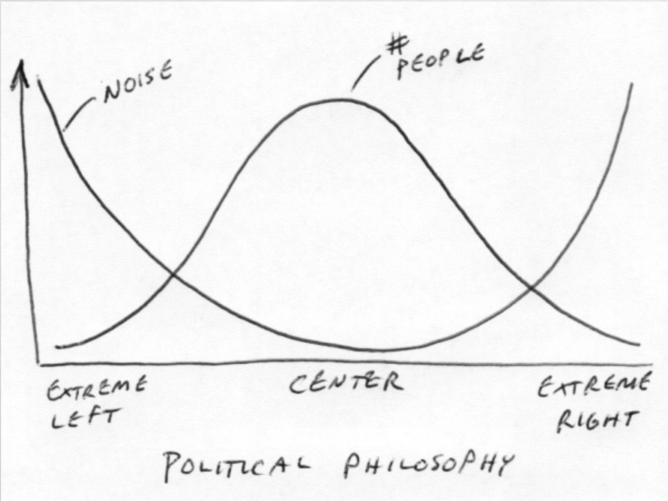
STATISTIKA I OSNOVE FIZIKALNIH MJERENJA

ŽELJKO SKOKO

PREDAVANJA: PONEDJELJAK, 13-15 h, F25

VJEŽBE: PONEDJELJAK, 15-16 h, F25

TEODORO KLASER



KONZULTACIJE: prema dogovoru

e-mail: zskoko@phy.hr

soba 224, tel.: 4605 518

STATISTIKA I OSNOVE FIZIKALNIH MJERENJA

LITERATURA:

WALPOLE ET AL.: Probability & Statistics For Engineers & Scientists; Pearson, Prentice Hall, 2007.

I. ŠOŠIĆ, V. SERDAR: Uvod u Statistiku; ŠK, Zagreb 1994. (Ekonomski fakultet).

V. VRANIĆ: Vjerojatnost i statistika, Tehnička knjiga, Zagreb 1971.

I.PAVLIĆ: Statistička teorija i primjena, Tehnička knjiga, Zagreb 1970 (Strojarski fakultet).

N. SARAPA: Teorija vjerojatnosti, ŠK, Zagreb 1992. (PMF-MO)

D. UGRIN-ŠPARAC: Primijenjena teorija vjerojatnosti I., II., Sveučilišna naklada - Liber, Zagreb, 1975/1976. (FER).

STATISTIKA I OSNOVE FIZIKALNIH MJERENJA

http://www.phy.pmf.unizg.hr/~zskoko/

- nastavni materijali
- obavijesti
- ispitna pitanja
- termini ispita, rezultati
- itd...

Pravila kolegija:

- predavanja nisu obavezna ali iskustvo pokazuje da su vrlo korisna kod polaganja ispita
- u toku semestra bit će održana 2 kolokvija (neobavezna)
- ukoliko je zbroj na oba kolokvija veći od 50% oslobađanje pismenog dijela ispita
- ukoliko kolokviji nisu položeni obavezan je izlazak na pismeni dio ispita (zadaci kao na vježbama)
- usmeni dio ispita je obavezan (pitanja na webu)
- usmeni: tri teorijska pitanja, na svako barem nešto odgovoriti
- zadaće nose dodatne bodove
- ispitni rokovi u lipnju (2x), rujnu (2x), studenom (1x), veljači (2x)
- prolaznost: ak. god. 2017/2018 položilo 25 od 54 studenata (46%) (ak. god. 2016/17 - 71%)

ČEMU SLUŽI TEORIJA VJEROJATNOSTI I STATISTIKA

- IGRE NA SREĆU
- POLITIKA I MARKETING (DOVOLJAN JE UZORAK < 0.1% POPULACIJE)
- GOSPODARSTVO
- ZDRAVSTVO
- GENETIKA
- FIZIKA
  - TEORIJSKA FIZIKA
    - STATISTIČKA FIZIKA
    - KVANTNA FIZIKA
  - EKSPERIMENTALNA FIZIKA
    - OBRADA REZULTATA MJERENJA - PROCJENA POGREŠAKA
    - PROVJERA HIPOTEZA

UVOD

- MNOŠTVO PODATAKA - Statistika nam daje metode za njihovo organiziranje i sažeto prikazivanje te izvlačenje zaključaka na osnovu informacija sadržanih u tim podacima
- Od latinskog "STATUS" - stanje, situacija
- Povijesno, državni interes - popisi vojnih obveznika u Kini, Perziji, GrčkoJ...
- Državni zavod za statistiku http://www.dzs.hr

  http://www.dzs.hr/StatInfo/Zaposl.htm

  http://www.dzs.hr/StatInfo/RADSNAGA.htm (anketa na reprezentativnom uzorku)
- Primjeri u meteorologiji: http://meteo.hr

KOMBINATORIKA

(1) PERMUTACIJE BEZ PONAVLJANJA

- N RAZLIČITIH ELEMENATA A1, ..., AN MOGUĆE JE POREDATI U NIZ NA VIŠE RAZLIČITIH NAČINA
- BROJ TIH NAČINA ZOVEMO PERMUTACIJE BEZ PONAVLJANJA
- ODABEREMO JEDAN ELEMENT $\longrightarrow$ N IZBORA
- ODABEREMO SLJEDEĆI ELEMENT $\longrightarrow$ N - 1 IZBORA
- ODABEREMO TREĆI ELEMENT $\longrightarrow$ N - 2 IZBORA

$$\boxed{P^{(N)} = N\cdot(N-1)\cdot(N-2)\cdot\ldots\cdot2\cdot1 = N!}$$

PO DEFINICIJI: $0! = 1$

Primjer: Na koliko načina možemo poredati brojeve 1, 2 i 3?

123, 132, 213, 231, 312, 321 $\longrightarrow$ 6 načina

$P(3) = 3! = 3 \cdot 2 \cdot 1 = 6$

(2) PERMUTACIJE S PONAVLJANJEM

- n elemenata unutar kojih je r_1, r_2, ..., r_k jednakih

$$\sum_i r_i = n$$

- na koliko načina možemo poredati te elemente u niz?

- zamislimo da su svi elementi različiti; tada bi bilo n! načina; zamislimo da su svi elementi različiti osim njih r_1 $\longrightarrow$ permutacije tih istih elemenata ne pridonose ništa: $\dfrac{n!}{r_1!}$ načina

- ako su r_1 i r_2 elemenata isti: $\dfrac{n!}{r_1!r_2!}$ načina

OPĆENITO: $$P^{(n)}_{r_1, r_2, \ldots, r_k} = \frac{n!}{r_1! \cdot r_2! \cdot \ldots \cdot r_k!}$$

Primjer: Na koliko načina možemo poredati brojeve 1, 1 i 2?

112, 121, 211 3 načina $n = 3, r_1 = 2$

$$P_2^{(3)} = \frac{3!}{2!} = 3$$

Primjer: Na koliko načina možemo poredati slova a,a,b,b,b,c?

60 načina

$n = 6, r_1 = 2, r_2 = 3$

$$P_{2,3}^{(6)} = \frac{6!}{2!3!} = 60$$

WolframAlpha

permutations of {a,a,b,b,b,c}

Input interpretation

permutations	{a, a, b, b, b, c}

Number of distinct permutations

60

Permutations:

{a, a, b, b, b, c} | {a, a, b, b, c, b} | {a, a, b, c, b, b} | {a, a, c, b, b, b} | {a, b, a, b, b, c} | {a, b, a, b, c, b} | {a, b, a, c, b, b} | {a, b, b, a, b, c} | {a, b, b, a, c, b} | {a, b, b, b, a, c} | {a, b, b, b, c, a} | {a, b, b, c, a, b} | {a, b, b, c, b, a} | {a, b, c, a, b, b} | {a, b, c, b, a, b} | {a, b, c, b, b, a} | {a, c, a, b, b, b} | {a, c, b, a, b, b} | {a, c, b, b, a, b} | {a, c, b, b, b, a} | {b, a, a, b, b, c} | {b, a, a, b, c, b} | {b, a, a, c, b, b} | {b, a, b, a, b, c} | {b, a, b, a, c, b} | {b, a, b, b, a, c} | {b, a, b, b, c, a} | {b, a, b, c, a, b} | {b, a, b, c, b, a} | {b, a, c, a, b, b} | {b, a, c, b, a, b} | {b, a, c, b, b, a} | {b, b, a, a, b, c} | {b, b, a, a, c, b} | {b, b, a, b, a, c} | {b, b, a, b, c, a} | {b, b, a, c, a, b} | {b, b, a, c, b, a} | {b, b, b, a, a, c} | {b, b, b, a, c, a} | {b, b, b, c, a, a} | {b, b, c, a, a, b} | {b, b, c, a, b, a} | {b, b, c, b, a, a} | {b, c, a, a, b, b} | {b, c, a, b, a, b} | {b, c, a, b, b, a} | {b, c, b, a, a, b} | {b, c, b, a, b, a} | {b, c, b, b, a, a} | {c, a, a, b, b, b} | {c, a, b, a, b, b} | {c, a, b, b, a, b} | {c, a, b, b, b, a} | {c, b, a, a, b, b} | {c, b, a, b, a, b} | {c, b, a, b, b, a} | {c, b, b, a, a, b} | {c, b, b, a, b, a} | {c, b, b, b, a, a} (total: 60)

(3) VARIJACIJE BEZ PONAVLJANJA

- n različitih elemenata A1, ..., An i želimo od njih stvoriti skup od r elemenata, s time da se isti element može pojaviti samo jednom i da je poredak elemenata važan!
- odaberemo prvi član skupa $\longrightarrow$ n načina
- odaberemo drugi član skupa $\longrightarrow$ n - 1 načina
- ...
- odaberemo r-ti član skupa $\longrightarrow$ (n - r+1)
- slično kao kod permutacija bez ponavljanja

$$\boxed{n(n-1)(n-2)\ldots[n-(r-1)]=\frac{n!}{(n-r)!}}$$

- iz binomnog poučka imamo:

$$(a+b)^n=\sum_{x=0}^{n}\binom{n}{x}a^{n-x}b^x;\quad \binom{n}{x}=\frac{n!}{(n-x)!x!}$$

VARIJACIJE BEZ PONAVLJANJA:

$$V_r^{(n)} = \frac{n!}{(n-r)!} = \frac{n!}{(n-r)!r!} \cdot r! = \binom{n}{r} \cdot r!$$

Primjer: Na koliko načina možemo od brojeva 1, 2, 3, 4 formirati skupove od po dva broja, ali tako da se svaki broj može pojaviti samo jednom, i da je poredak brojeva u skupu važan?

1 2, 1 3, 1 4, 2 3, 2 4, 3 4

2 1, 3 1, 4 1, 3 2, 4 2, 4 3

12 NAČINA!

$$V_2^{(4)} = \frac{4!}{(4-2)!} = \frac{4!}{2!} = 12$$

(3) VARIJACIJE S PONAVLJANJEM

- n različitih elemenata A1, ..., An i želimo od njih stvoriti skup od r elemenata, poredak elemenata je važan, isti element se može pojaviti više puta!
- odaberemo prvi član skupa $\longrightarrow$ n načina
- odaberemo drugi član skupa $\longrightarrow$ n načina
- ...
- odaberemo r-ti član skupa $\longrightarrow$ n načina
- slično kao i prije

$$n \cdot n \cdot n \cdot n \ldots \cdot n = n^r \quad (r \text{ elemenata})$$

$$\boxed{\bar{V}_r^{(n)} = n^r}$$

Primjer: Na koliko načina možemo od brojeva 1, 2, 3, 4 formirati skupove od po dva broja, tako da se svaki broj može pojaviti više puta, i da je poredak brojeva u skupu važan?

1 2, 1 3, 1 4,   2 3, 2 4,   3 4

2 1, 3 1, 4 1,   3 2, 4 2,   4 3   16 NAČINA!

1 1, 2 2, 3 3, 4 4

$$\overline{V}_2^{(4)} = 4^2 = 16$$

(5) KOMBINACIJE BEZ PONAVLJANJA

- n različitih elemenata slažemo u skupove veličine r, svaki element smije se u skupu pojaviti samo jednom, poredak nije važan

- isto kao i varijacije bez ponavljanja, samo što poredak nije važan

$$\boxed{K_r^{(n)} = \frac{V_r^{(n)}}{r!} = \frac{n!}{r!(n-r)!} = \binom{n}{r}}$$

Primjer: Na koliko načina možemo od brojeva 1, 2, 3, 4 formirati skupove od po dva broja, tako da se svaki broj može pojaviti više puta, i da je poredak brojeva u skupu važan?

1 2, 1 3, 1 4, 2 3, 2 4, 3 4 6 NAČINA!

$$K_2^{(4)} = \frac{4!}{2!(4-2)!} = \frac{4!}{2!2!} = \frac{4\cdot 3\cdot 2\cdot 1}{2\cdot 1\cdot 2\cdot 1} = 6$$

(6) KOMBINACIJE S PONAVLJANJEM

- n različitih elemenata slažemo u skupove veličine r, svaki element smije se u skupu pojaviti više puta, poredak nije važan
- isto kao i varijacije sa ponavljanjem, samo što poredak nije važan

$$\boxed{\bar{K}_r^{(n)} = \binom{n+r-1}{r} = K_r^{(n+r-1)}}$$

OSNOVNI STATISTIČKI POJMOVI

POPULACIJA - dobro definiran skup objekata koji nas zanima

Populacija može biti konačna ili beskonačna, realna ili hipotetska

Primjeri:

1. svi zaposlenici na području Grada Zagreba u 2014.
2. svi stanovnici Republike Hrvatske 31. 03. 2014.
3. sve baterije proizvedene u nekoj tvornici u jednom mjesecu
4. svi mogući rezultati nekog mjerenja

Poznavanje cijele populacije - CENZUS

Cenzus rijetko možemo imati (preskupo ili neizvedivo)

OSNOVNI STATISTIČKI POJMOVI

UZORAK - podskup populacije koji uzimamo na unaprijed određen način

Primjeri:

1. iz populacije svih studenata uzimamo samo jednu godinu
2. u skladištu opeka uzimamo jedan red
3. obavimo deset mjerenja mase nekog predmeta

VARIJABLA - obilježje čija vrijednost se može mijenjati od objekta do objekta

Možemo promatrati jedno obilježje (npr. visina studentata) pa govorimo o 1-D skupu podataka ili više obilježja. Npr. promatramo li visinu i težinu studenata, imat ćemo 2-D skup podataka.

STATISTIKA

OPISNA

Istraživač je sakupio podatke i želi ih organizirati, te sažeto opisati važna svojstva

Grafičke metode: histogrami, krivulje, kružni dijagrami...

Numeričke vrijednosti: prosjek, standardna devijacija, korelacijski koeficijent,...

INDUKTIVNA

Statističko zaključivanje

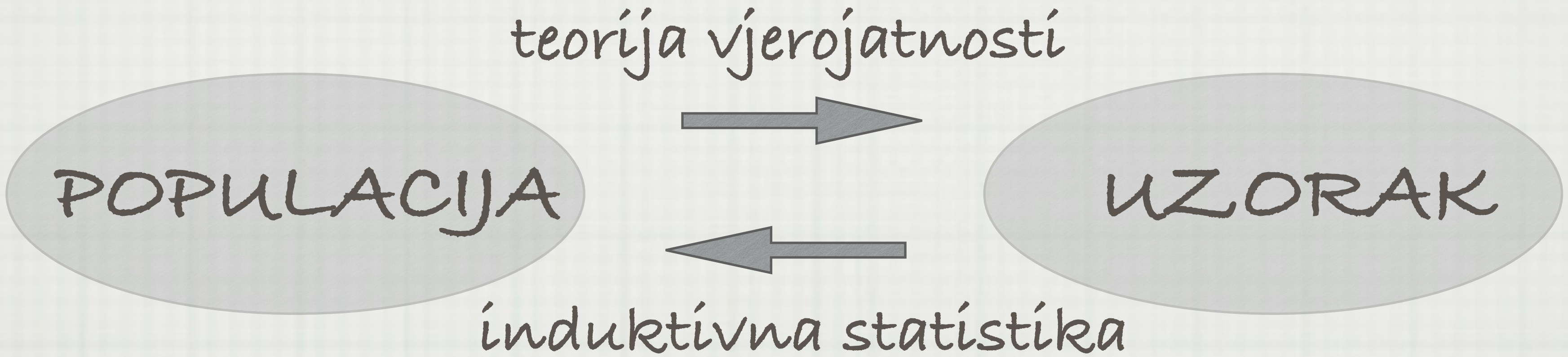
Istraživač je proučio uzorak i iz njegovih svojstava želi izvesti zaključke o cijeloj populaciji

Nužno je razumjeti neodređenosti povezane s uzimanjem uzorka iz populacije

STATISTIKA

TEORIJA VJEROJATNOSTI

pretpostaka - znamo svojstva populacije i želimo predvidjeti svojstva uzorka



U OVOM KOLEGIJU NAJPRIJE ĆEMO UKRATKO NAUČITI METODE OPISNE STATISTIKE, ZATIM ĆEMO PROUČAVATI TEORIJU VJEROJATNOSTI, I NAKON TOGA INDUKTIVNU STATISTIKU

GRAFIČKE I TABLIČNE METODE

S-L DIJAGRAM (STEM AND LEAF, STABLJIKA I LIST)

primjer: Na 140 američkih koledža provedena je anketa o "tulum-pijancima". Kao tulum-pijanac opisan je onaj student koji u nizu popije 4 ili više alkoholnih pića. Postotak tulum-pijanaca dan je u sljedećem popisu:

37	27	36	43	49	57
61	27	37	44	49	46
11	28	68	55	50	26
14	29	37	44	50	58
17	29	64	45	51	59
29	4	37	66	51	60
18	29	38	45	52	42
51	30	59	46	52	61
19	64	57	59	52	62
21	31	16	57	18	64
22	39	38	46	53	65
66	31	39	13	37	43
22	32	39	46	53	65
23	56	39	47	54	46
24	32	41	67	47	66
52	33	41	47	55	67
25	48	42	58	56	48
56	44	61	48	56	39
28	34	42	49	41	36
26	35	46	33	56	44
61	35	42	58	48	33
48	35	18	26	57	
26	36	27	38	38	
34	36	58	42	15	

S-L DIJAGRAM

Da bismo iz ove "šume" podataka stekli dojam o rasprostranjenosti pojave, izrađujemo S-L dijagram:

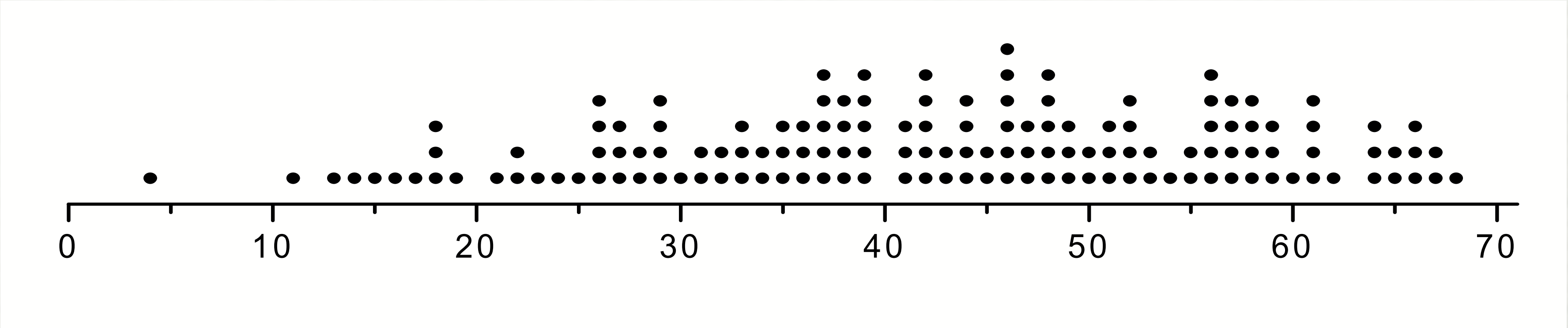
1. odaberemo jednu ili više početnih znamenki za 'stabljiku'
2. u stupac popišemo sve moguće 'stabljike'
3. desno od svake stabljike popišemo sve pripadajuće 'listove'

0	4
1	1345678889
2	1223456666777889999
3	0112233344555666677777888899999
4	1112222223344445566666677788888999
5	001112222233455666667777888899
6	01111244455666778

PRIKAZ POMOĆU TOČAKA

1. Nacrtamo dio brojevnog pravca koji obuhvaća sve moguće podatke
2. Svako opažanje prikažemo točkom iznad odgovarajućeg položaja

Za podatke iz gornjeg primjera točkasti prikaz izgleda ovako:



HISTOGRAMI

Postupak crtanja histograma različit je za diskretne i kontinuirane varijable

Diskretna varijabla – ona za koju je skup mogućih vrijednosti konačan ili prebrojiv

– najčešće se radi o prebrojavanju

Kontinuirana varijabla – skup mogućih vrijednosti jest cijeli interval na brojevnom pravcu

– najčešće rezultat mjerenja

PRIMJER DISKRETNE VARIJABLE

Starost djece koja se igraju u parku. Varijabla je X = starost djeteta

Rezultati ankete za 48 djece (starost u godinama):

5	5	8	9
10	11	11	12
3	5	8	9
4	6	18	10
9	11	9	9
11	6	8	3
4	4	1	7
14	10	8	10
4	9	9	5
10	6	11	12
5	7	9	5
8	10	4	20

Moguće vrijednosti varijable X su 1, 2, 3, ..., 18, 19, 20, ...

Dakle, radi se o diskretnoj varijabli

FREKVENCIJA

Def.: Frekvencija f_i neke određene vrijednosti x_i varijable X jest broj pojavljivanja te vrijednosti u promatranom skupu podataka.

Vrijedi: $\sum_i f_i = N$, gdje je N broj svih podataka u skupu

Tablični prikaz svih frekvencija naziva se raspodjela frekvencija

U našem primjeru raspodjela frekvencija izgleda ovako:

x_i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
f_i	1	0	2	5	6	3	2	5	8	6	5	2	0	1	0	0	0	1	0	1

Def.: Relativna frekvencija, f_{ri}, neke određene vrijednosti x_i varijable X jest frekvencija podijeljena s ukupnim brojem podataka $N = \sum f_i$,

$$f_{ri} = \frac{f_i}{N}$$

U našem primjeru je N = 48 pa je raspodjela relativnih frekvencija dana tablicom:

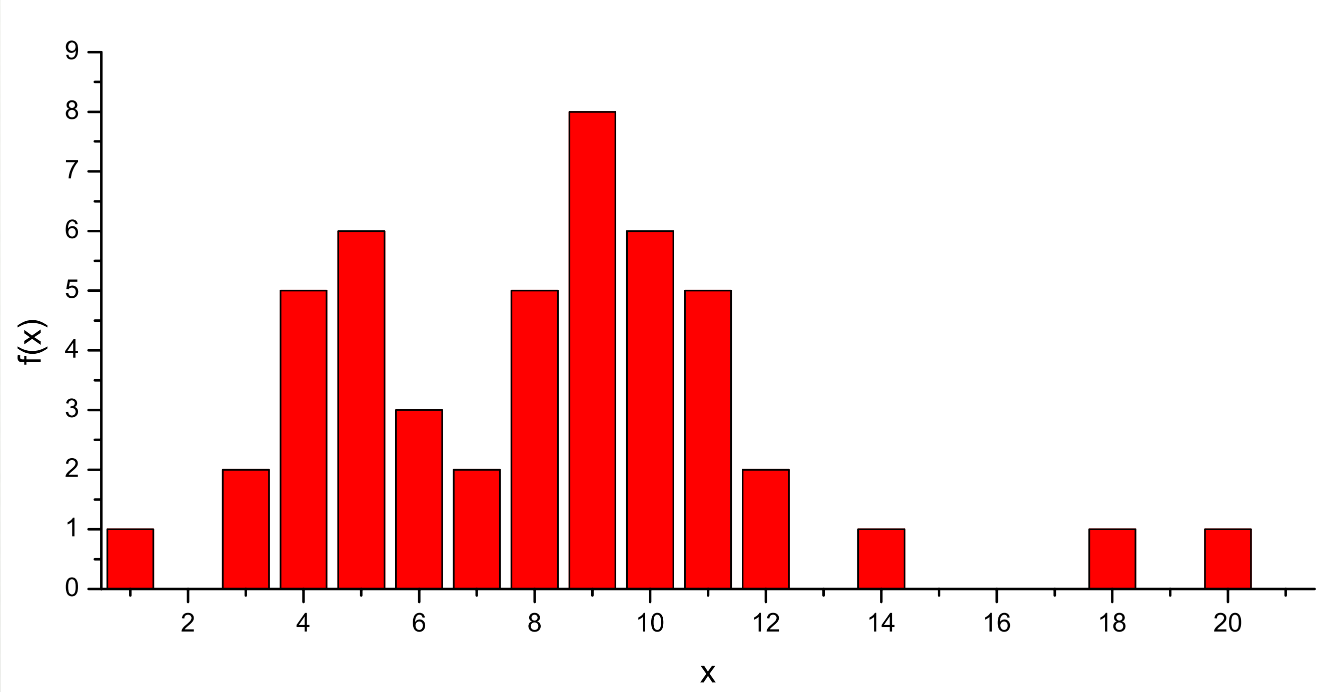
x_i	f_{ri}	f_{ri} (%)
1	0,0208	2,08
2	0	0,00
3	0,0417	4,17
4	0,1042	10,42
5	0,125	12,50
6	0,0625	6,25
7	0,0417	4,17
8	0,1042	10,42
9	0,1667	16,67
10	0,125	12,50
11	0,1042	10,42
12	0,0417	4,17
13	0	0,00
14	0,0208	2,08
15	0	0,00
16	0	0,00
17	0	0,00
18	0,0208	2,08
19	0	0,00
20	0,0208	2,08

Crtanje histograma

(za diskretnu raspodjelu):

1. odredimo frekvencije ili rel. frekvencije
2. na apscisi označimo moguće vrijednosti varijable x
3. nacrtamo pravokutnik visine f_i ili f_{ri}

"NAŠ" PRIMJER:



Konstrukcija histograma za kontinuiranu varijablu

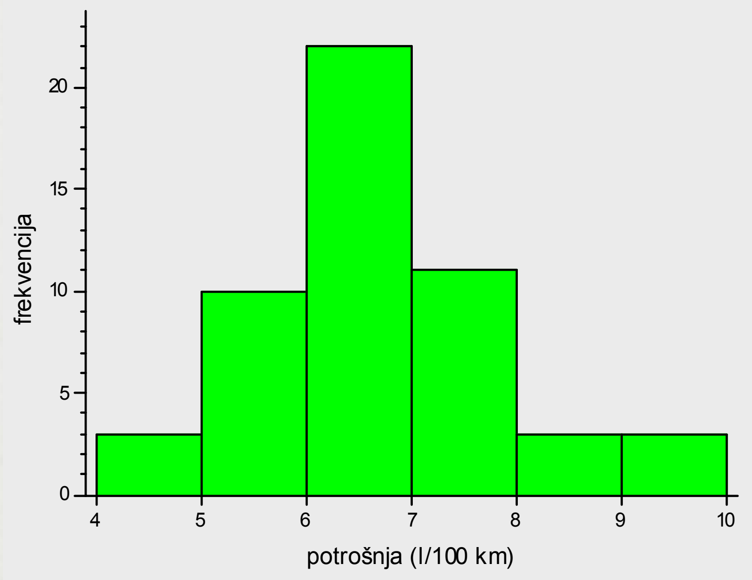
Primjer: potrošnja benzina (u litrama na 100 km):

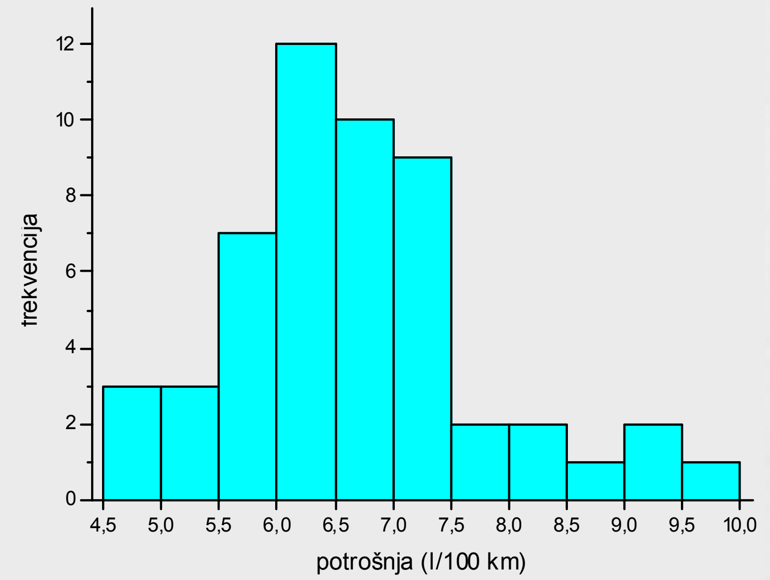
6,25	5,93	7,8	4,95	9,2	8,57
6,82	7,43	5,78	5,46	6,54	7,02
6,78	4,75	5,32	7,11	5,66	5,99
6,87	8,35	7,66	7,23	6,58	6,92
6,32	7,08	5,98	6,25	5,45	6,72
6,38	6,9	9,87	6,23	6,52	6,43
6,12	5,81	6,37	7,23	7,46	8,06
6,09	5,82	4,99	6,32	6,51	6,49
9,49	6,39				

Konstrukcija histograma:

1. podijelimo apscisu na prikladan broj razreda (ekvidistantno ili ne)
2. odredimo frekvencije
3. crtamo pravokutnike
   - za ekvidistantne razrede visina = f_{ri}
   - za neekvidistantne razrede visina = f_{ri}/širina

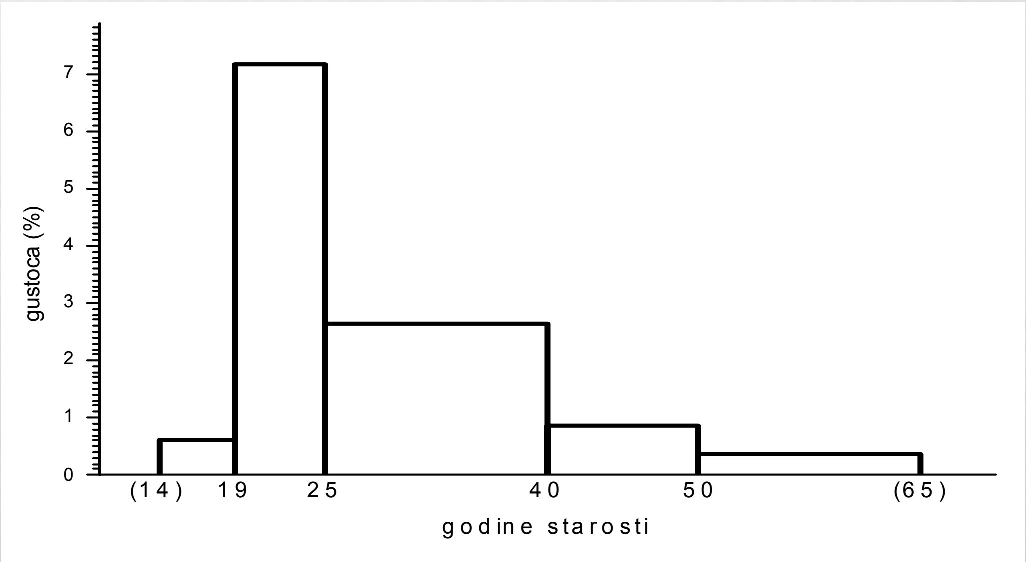
Primjer potrošnje benzina - ekvidistantni razredi





Primjer neekvidistantnih razreda:
osobe prijavljene zavodima za zapošljavanje, RH 1990:

dob	broj osoba (u tisućama)	f_{ri} (%)
(14)-19	6,1	3,12
19-25	84,0	42,99
25-40	77,5	39,66
40-50	16,6	8,50
50-(65)	11,2	5,73
UKUPNO	195,4	100,00



HISTOGRAMI - napomene:

- površina pravokutnika je frekvencija danog razreda
- ukupna površina relativnih frekvencija je 1

Prema obliku histogrami mogu biti:

- unimodalni
- bimodalni
- višemodalni

ili

- simetrični
- pozitivno nagnuti
- negativno nagnuti

MJERE POLOŽAJA

Grafičkim prikazom stekli smo impresiju, ali želimo izvući neke numeričke vrijednosti koji karakteriziraju skup podataka.

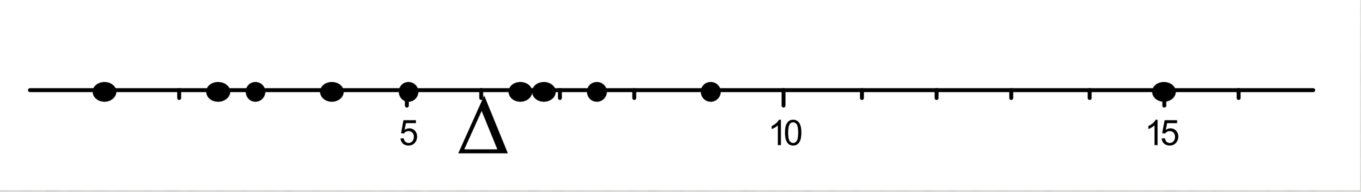
Neka se skup sastoji od brojeva $x_1, x_2, \dots x_n$. Najprije nas zanima položaj i-tog elementa, posebno središte.

Neka skup $x_1, x_2, \dots x_n$ predstavlja uzorak iz veće populacije!

Def: Srednja vrijednost, $\bar{x}$, uzorka $x_1, x_2, \dots x_n$ dana je s:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Fizikalno značenje – težište!



Def: Srednju vrijednost populacije označavamo s μ

U induktivnoj statistici najčešće želimo procijeniti μ na osnovu uzorka

---

Najveći nedostatak srednje vrijednosti kao mjere položaja - podliježe utjecaju ekstremnih vrijednosti u uzorku

Def: Medijan, $\tilde{\mu}$, određujemo tako da sva opažanja poredamo po veličini. Ako je broj opažanja neparan, medijan je vrijednost $[(n+1)/2]$ -tog opažanja. Ako je broj opažanja paran, medijan je srednja vrijednost $(n/2)$ -tog i $(n/2+1)$ -tog opažanja

Kvartili dijele skup opažanja na četiri dijela

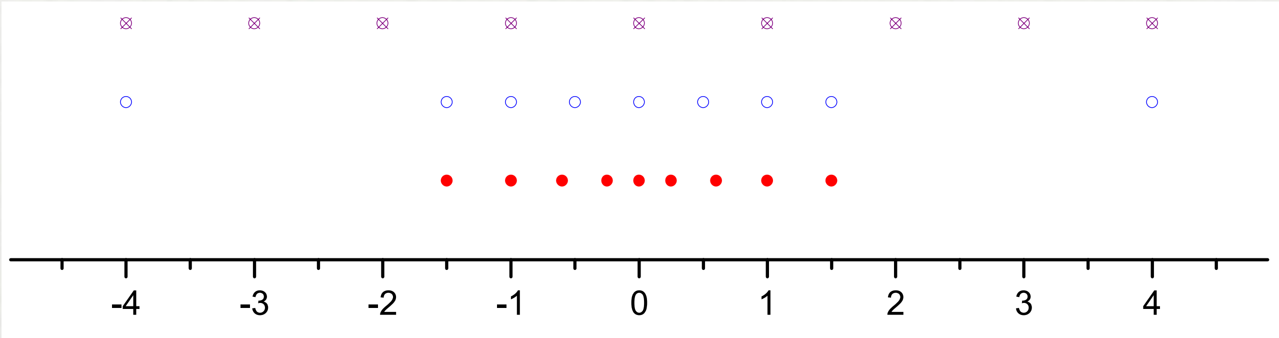
Percentili dijele skup opažanja na 100 dijelova

"Ošišani prosjek" - odbacimo jednak broj ekstrema s obje strane pa računamo srednju vrijednost

MJERE POLOŽAJA

Uzorci s istim mjerama položaja (prosjek, medijan) mogu se bitno razlikovati

Slika prikazuje tri različita skupa istog prosjeka i medijana:



Osim središta uzorka, zanima nas i koliko je raspršen

Najjednostavnija mjera raspršenja je:

Def: Raspon uzorka - razlika između najveće i najmanje vrijednosti u uzorku

Mana: ovisi samo o ekstremima

Def: Odstupanje od prosjeka pojedine vrijednosti u uzorku: $x_i - \bar{x}$

Želimo izraziti mjeru odstupanja u jednom broju

---

1. ideja: zbrojimo sva odstupanja: $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})$

No, to je jednako nuli (pozitivna i negativna odstupanja se poništavaju)

2. ideja: zbrojimo apsolutne vrijednosti svih odstupanja: $\frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$

Timo dobivamo mjeru odstupanja, ali i znatne matematičke komplikacije

3. ideja: zbrojimo sve kvadrate odstupanja: $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Def: Varijanca uzorka: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Def: standardna devijacija uzorka: $s = \sqrt{s^2}$

Pitanje! Zašto smo uzeli n-1, a ne n?

Def: Varijanca populacije: (gdje je μ prosjek populacije) $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$

Vrlo rijetko poznajemo μ pa pokušavamo iz uzorka odrediti μ i σ. No, prosječni kvadrat odstupanja $(x_i - \bar{x})^2$ u uzorku je manji od nepoznatog $(x_i - \mu)^2$. Da bismo to korigirali, u nazivnik izraza za varijancu stavljamo n-1 umjesto n. Na taj način možemo s pomoću varijance uzorka procijeniti varijancu populacije.

Korekcija dolazi zbog umanjenog broja stupnjeva slobode u uzorku.

To ćemo dokazati kasnije u semestru.