

# Uvod u statistiku

# Osnovni pojmovi

- Statistika – nauka o podacima
  - Uključuje prikupljanje, klasifikaciju, prikaz, obradu i interpretaciju podataka
- Statistička jedinica – objekt kome se mjeri neko svojstvo. Svi objekti zajedno čine statistički skup
- Mjereno svojstvo se javlja u dva ili više oblika – modaliteta
  - Na primjer za pol imamo dva modaliteta: muški i ženski
- Obilježje ili varijabla – svojstvo koje se dobija opažanjem ili mjerenjem na statističkom skupu. Ono svojim oblikom varira od jedinice do jedinice statističkog skupa
- Rezultat svakog mjerenja je podatak

# Primjer 1

- Na Univerziteti se željelo ispitati da li studenti podržavaju pokretanje plesnog kursa. Od 650 ispitanih studenata, 435 je podržalo takvu ideju.
  - Šta je statistička jedinica u ovom primjeru?
    - a) Odgovor svakog studenta
    - b) Studenti koji podržavaju
    - c) Jedan student
  - Šta je varijabla?
    - a) Broj studenata koji podržavaju opciju
    - b) Broj studenata koji ne podržavaju opciju
    - c) Odgovor pojedinog studenta

# Osnovni pojmovi - rezime

- Statistički skup:  $\Omega$
- Statistička jedinica:  $\omega$
- Varijabla:  $X : \Omega \rightarrow K$  (skup mogućih oblika pridruženih obilježju  $X$ )
- Mjerenjem (opažanjem) varijable  $X$  dobijamo podatke o nekom svojstvu na posmatranom statističkom skupu.

# Vrste podataka i varijabli

- Prema mjernoj skali obilježja (varijable) dijelimo na:
  - kvantitativne – mjerene na numeričkoj skali, dopuštene brojčane operacije (visina, težina, primanja, broj bodova na kolokvijumu, ...)
  - kvalitativne – vrijednosti su kategorije. Nisu dopuštene brojčane operacije (boja kose, pol, ...). U pojedinim slučajevima dopuštene su operacije upoređivanja:  $<$ ,  $>$ ,  $=$  (npr. ocjena na ispitu). Analogno govorimo i o kvantitativnim i kvalitativnim podacima (dobivenim mjerenjima odgovarajućih varijabli).
- Primjer 2. Odredite kojeg tipa su sljedeći podaci:
  - Visina studenta u razredu
  - Vrijeme koje svako od 30 pacijenata mora provesti u bolnici
  - Politička stranka pojedinog poslanika u Skupštini Crne Gore
  - Veličina stana (u m<sup>2</sup>).

# Vrste podataka

- Razmotrimo opet prvi primjer.
- Primjer 3. Bila su ponuđena tri odgovora, označena sa a), b) i c), te je zabilježen odgovor svakog studenta.
  1. Kakav je to tip varijable?
  2. Zamislimo da smo u istom primjeru ponuđene moguće odgovore označili sa 1), 2) i 3), te ponovno bilježili odgovor svakog studenta. Je li u tom slučaju riječ o kvantitativnoj ili kvalitativnoj varijabli?

# Osnovni elementi statističke analize

- Populacija ili osnovni skup – skup svih podataka koji opisuju neki fenomen koji nas zanima.
- Primjer: Zanima nas vrijednost prosječnog primanja u Crnoj Gori. Šta je populacija u ovom slučaju? Sta je statistička jedinica?
- Uzorak - podskup podataka sakupljenih iz populacije.
- Primjer 4. Menadžer u nekoj kompaniji želi ispitati starosnu dob kupaca koji kupuju njihove proizvode. U tu svrhu na slučajan način izaberu 742 kupca i utvrde da je prosjek njihovih godina 42. Šta je interesna populacija u ovom slučaju?
  - Prosjek godina svih kupaca koji kupuju proizvode
  - Prosjek godina 742 ispitanih kupaca
  - Starosna dob svih kupaca koji kupuju proizvode

# Primjer 4

- Šta je uzorak?
  - Prosjek godina svih kupaca proizvoda
  - Starosna dob 742 ispitanih kupaca.
  - 742 ispitana kupca.
- Šta je statistička jedinica?
  - Proizvod koji se kupuje
  - Starosna dob svakog kupca
  - Pojedini kupac
  - Prosjek godina svih kupaca koji kupuju proizvode
- Šta je varijabla?
  - Proizvod
  - Starosna dob kupca
  - Starosna dob 742 ispitanih kupaca
  - Prosjek godina svih kupaca koji kupuju proizvode



# Grane statistike

- Dizajn eksperimenta (experimental design) – bavi se planiranjem eksperimenta i prikupljanjem podataka
- Opisna statistika (descriptive statistics) – grana statistike koja se bavi obradom i prikazom podataka
- Statističko zaključivanje (inferential statistics) – na osnovu uzorka donosi zaključke o cijeloj populaciji. Zasniva na nepotpunim podacima, te u sebi sadrži određenu komponentu greške.

# Primjer 5

- Vratimo se na primjer 4. Šta od sljedećeg bi bio primjer statističkog zaključivanja?
  - 742 kupca su slučajno izabrana
  - Prosjek godina 742 ispitanih kupaca je 42
  - Na osnovu uzorka, možemo sa X% sigurnosti zaključiti da je prosjek kupaca između 40 i 44 godine.

# Prikupljanje podataka

- Reprezentativni uzorak – posjeduje svojstva slična onima iz cijele populacije.
- Kako izabrati takav uzorak?
- Slučajni uzorak (random sample) – od  $N$  statističkih jedinica svaka jedinica ima jednaku vjerovatnoću da bude izabrana za uzorak
- Primjeri:
  - Loto 6/45
  - generator slučajnih brojeva
  - Izvlačenje iz šešira

# Grafički prikaz kvalitativnih podataka

- Kvalitativni podaci – nemaju numeričko značenje, mogu se samo razvrstavati u kategorije.
- Primjer 6. Uzet je uzorak od 176 zaposlenih jednog poduzeća, te je bilježena vrsta njihovog obrazovanja. Mogući odgovori su bili: srednja škola (S), specijalista (B), magistar (M), doktor (D), ostalo (O). Što su kategorije u ovom primjeru?

# Grafički prikaz kvalitativnih podataka 2

- Šta nas zanima?
- Frekvencija ili učestalost pojedine kategorije je broj podataka koji pripadaju toj kategoriji.
- Što nas još zanima?
- Relativna frekvencija ili učestalost pojedine kategorije je procenat od ukupnog broja podataka koji pripada toj kategoriji. On je jednak količniku frekvencije i ukupnog broja podataka.
  - Relativna frekvencija =  $\text{frekvencija} / \text{ukupan broj podataka}$

# Grafički prikaz kvalitativnih podataka 3

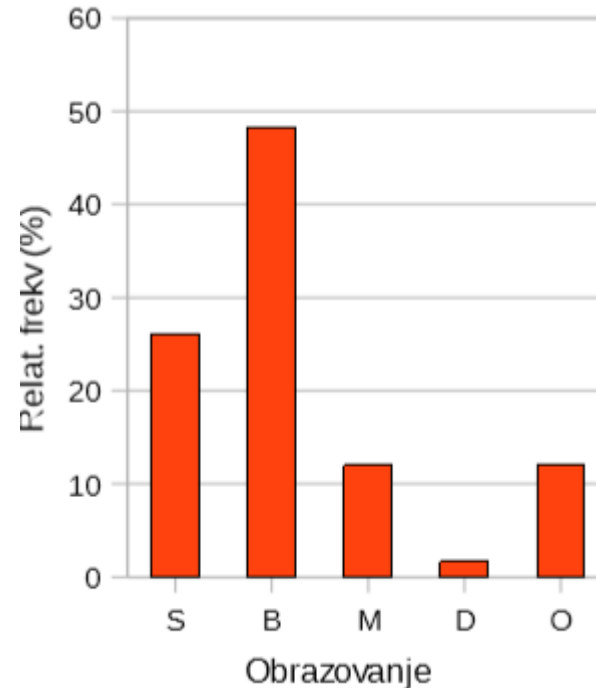
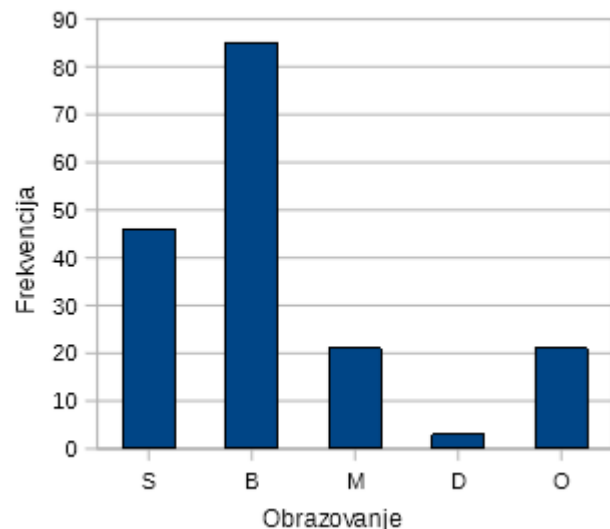
- Tablica frekvencija

Kategorija	Frekvencija	Relativna frekvencija
Srednja škola	46	26
Specijalista	85	48
Magistar	21	12
Doktor3	3	2
Ostalo	21	12

- Dvije najčešće korišćene grafičke metode za prikaz kvalitativnih podataka su:
  - stubičasti dijagrami (frekvencija ili relativnih frekvencija)
  - strukturni dijagrami (prikazuju relativnu frekvenciju svake kategorije)

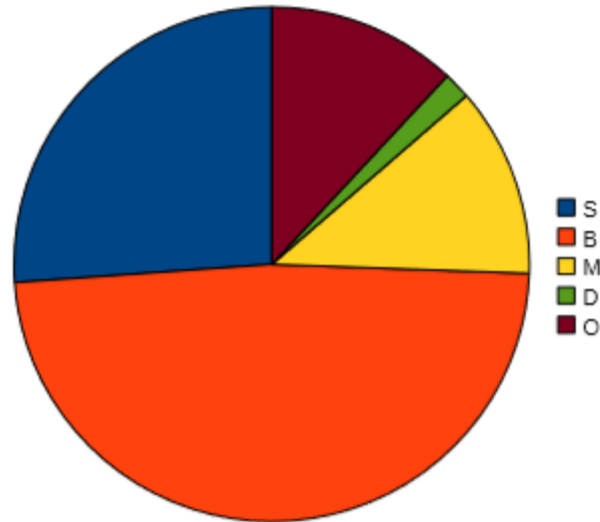
# Konstrukcija stubičastog dijagrama

- Nacrtajte horizontalnu i vertikalnu osu na papiru. Vertikalna osa predstavlja (relativnu) frekvenciju. Horizontalna osa predstavlja kategorije. Nacrtajmo stupce jednakih širina iznad svake kategorije. Visina stupca mora biti srazmjerna frekvenciji, odnosno relativnoj frekvenciji.



# Konstrukcija strukturnog dijagrama

- Pie chart sastoji se od kruga razdijeljenog na komade, od kojih svaki predstavlja pojedinu kategoriju. Veličina (ugao) pojedinog komada srazmjerna je relativnoj frekvenciji.





# Grafički prikaz kvantitativnih podataka

- Dvije najpopularnije vrste grafičkog prikaza kvantitativnih podataka su:
  - S-L prikaz (stablo i list, eng. stem and leaf)
  - Histogram
- Oni prikazuju:
  - broj podataka - frekvenciju
  - relativnu frekvenciju

# Histogram

- Histogrami frekvencija, odnosno relativnih frekvencija
  - podesniji za veće skupove podataka
  - slični stubičastim dijagramima
- Primjer 7. Data su 20 mjerenja slučajne varijable. Nacrtajte histogram frekvencija.

26 21 32 28 17 22 26 25 30 27 34 32 36 38 39 12 39 31 23 19

# Histogram 2

- Odredimo najmanju i najveću vrijednost: 12, 39
- Odredimo interval nad kojim ćemo crtati histogram:
  - početna tačka intervala < najmanjeg podatka
  - Završna tačka intervala > najvećeg podatka[10, 40]
- Odredimo podintervale. Pravićemo podintervale jednakih širina:
  - širina intervala = širina cijelog intervala / broj intervala
  - širina intervala =  $40 - 10 = 30$
  - Broj intervala – uzmimo ih 6 (Štrugesovo pravilo  $k \approx 1 + 3.3 \log N$ )
  - Širina razreda =  $30 / 6 = 5$

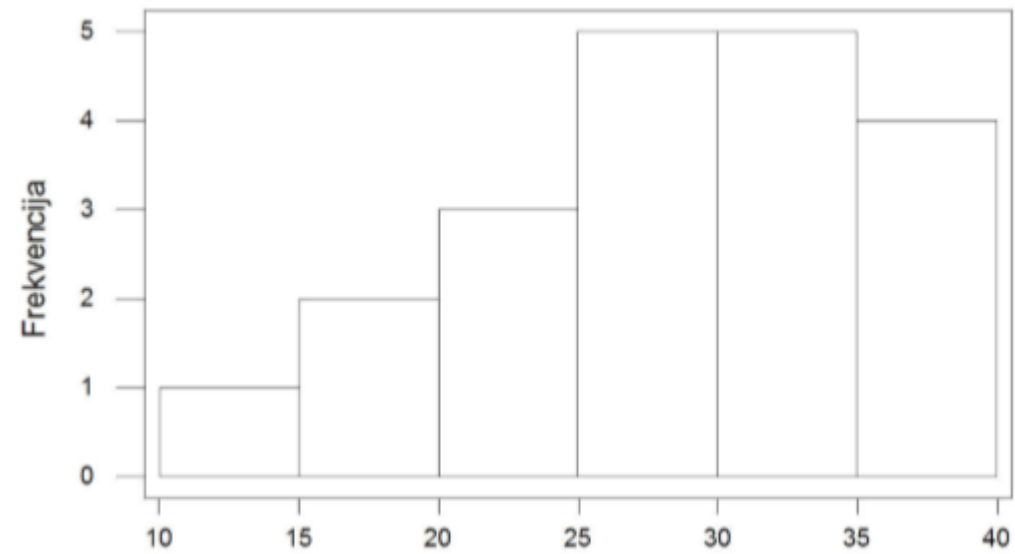
# Histogram 3

- Prvi razred: 10–15
- Drugi razred: 15–20
- Treći razred: 20–25 itd.
- Svaki podatak se mora razvrstati tačno u jednu kategoriju
- Precizniji zapis bi bio:
  - Prvi razred:  $[10,15)$
  - Drugi razred:  $[15,20)$  itd.
  - Svi razredi su poluotvoreni intervali, osim zadnjeg koji je zatvoren

# Histogram 4

Razred	Interval	Brojač	Frekvencija	Relat. frekv.(%)
1	10-15	/	1	5
2	15-20	//	2	10
3	20-25	///	3	15
4	25-30	////	5	25
5	30-35	////	5	25
6	35-40	////	4	20
		Ukupno:	20	100

# Histogram 5



# Primjer 9

- Nacrtajte histogram relativnih frekvencija za sljedeće podatke.

59 53 16 74 86 12 21 40 73 84 89 67 45 63 76 97 35 11 43 33 84 16 82  
65 11 50 94 64