



Opisivanje podataka

Graficki prikazi, oblik distribucije i numeričke mjere

Deskriptivna statistika - od grafikona do brojeva

bb

Ekonomija

March 25, 2026

- 1 Graficki prikaz podataka
- 2 Oblik distribucije
- 3 Mjere centralne tendencije
- 4 Mjere disperzije
- 5 Izbor statistickog sazetka

Zašto grafički prikaz?

Osnovno pravilo

Uvijek počnite s grafikonom svojih podataka. Numeričke mjere opisuju distribuciju, ali ne mogu zamijeniti vizuelni uvid u njen cjelokupni oblik.

Za kategorijske podatke

- **Bar plot** — frekvencija ili udio po kategorijama
- **Pie chart** — udio svake kategorije u cjelini; dobar za prikaz %

Konvencija: kategorije na x-osi, frekvencija na y-osi.

Za kvantitativne podatke

- **Histogram** — najčešći; prikazuje distribuciju frekvencija
- **Stemplot** — za male uzorke; cuva sve vrijednosti
- **Timeplot** — x-os je uvijek vrijeme
- **Scatterplot** — veza između dviju varijabli

Ključna razlika

Kvantitativni podaci: **redosljed i razmaci** između vrijednosti nose bitnu informaciju — postavljamo ih na numerički pravu, ne tretiramo kao kategorije.

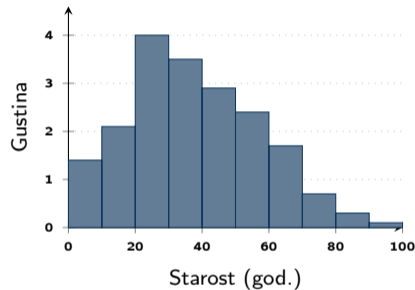
Kako je izgrađen histogram?

- x-osa: **intervali** (klase) mjerene varijable
- y-osa: **frekvencija** ili relativna frekvencija (%)
- Blokovi su **spojeni** — za razliku od bar plota
- **Površina bloka** \propto frekvenciji
- Ukupna površina = 100 %

Primjer izračuna

Interval 60–80 god., visina 0,7:

$20 \times 0,7 = 14\%$ populacije



Vrh: oko 20–30 god., $\approx 4\%$ gustine.

Tablica distribucije frekvencija

Interval	Frekv.	Rel. frekv.
[20, 30)	12	0,24
[30, 40)	18	0,36
[40, 50)	11	0,22
[50, 60)	9	0,18
Ukupno	50	1,00

Desna granica intervala nije uključena.

Relativna distribucija frekvencija

- Svaka frekvencija \div ukupan n
- Suma = 1,00 (ili 100%)
- Isti oblik histograma, promijenjena y-os

Konvencija granica

Interval [30, 40) uključuje 30, ali ne 40 — ta vrijednost ide u sljedeći interval. Bez ovog pravila neke vrijednosti ne bi imale interval.

Stemplot (stabljika–list)

- Prikazuje **svaku pojedinačnu vrijednost**
- Lijeva kolona = desetice, desno = jedinice
- Pogodan za **male uzorke** ($n < 50$)
- Rotirani stemplot = grubi histogram

1 | 2 5 7

2 | 3 4 4 8

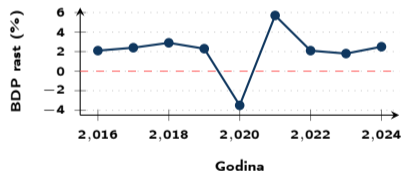
3 | 1 3 6 9

4 | 0 5

1/2 znaci vrijednost 12.

Timeplot (vremenski grafikon)

- x-os je **uvijek** vremenski period
- Prikazuje trend, sezonalnost, nagle promjene



Šta tražimo i čega se čuvamo

Na svakom grafikonu trazimo:

- 1 **Opšti obrazac** — centar, varijabilnost, oblik
- 2 **Outlieri** — opservacije van opšte slike
- 3 **Visestruki pikovi** — ukazuju na podgrupe
- 4 **Praznine** — nagli prekidi u distribuciji

Numeričke mjere ne otkrivaju pikove ni praznine.

Uobicajene grafičke zamke

- **3D grafikon** — uspoređujemo volumen umjesto visine
- **Presječena y-os** — dramtizira male razlike
- **Previše kategorija** u pie chart — tanki isjecci
- **Pogresni tip** — kategorije na kontinuojoj osi

Scatterplot — napomena

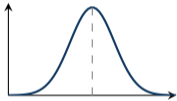
Za vezu izmedju **dviju** kvantitativnih varijabli. Svaka tačka = jedan pojedinac. Detaljno u kasnijim predavanjima.

Simetrične i iskošene distribucije

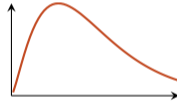
Definicija simetricne distribucije

Distribucija je **simetrična** ako su lijeva i desna strana histograma međusobno "ogledalne" slike.

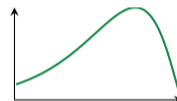
Simetrična



Iskošena udesno



Iskošena ulijevo



Iskošena udesno (pozitivna)

Desni rep je **duzi**. Rijetko ekstremno visoke vrijednosti.

Tipično za: prihode, cijene nekretnina, BDP

Iskošena ulijevo (negativna)

Lijevi rep je **duzi**. Rijetko ekstremno niske vrijednosti.

Tipično za: rezultate lakih testova, dob pri penzionisanju

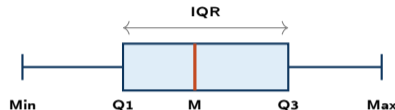
Boxplot: pet-brojčani sažetak vizuelno

Šta boxplot prikazuje?

Pet vrijednosti (manje info nego histogram, ali idealan za poređenje):

- **Min** (donji whisker)
- **Q1** — 25. percentil
- **M** — medijana, 50. percentil
- **Q3** — 75. percentil
- **Max** (gornji whisker)

IQR = $Q3 - Q1$ (raspon srednje 50%)



Kako čitati iskošenost?

- Veći dio *u kutiji* → ta strana je iskosena
- Duzi whisker → rep je duži na toj strani
- Sve jednako → simetrična distribucija

Zašto je oblik distribucije važan?

Oblik nosi informaciju

Ukupan oblik distribucije je bitna informacija o varijabli — određuje koje statističke metode su primjerene.

Primjeri iz prakse

- **Visina odraslih** \approx simetrična (normalna)
- **Prihodi** — jako iskošeni udesno (mного umjerenih, rijetko izuzetno visoki)
- **Cijene dionica** — desno iskošene (ne mogu biti negativne)
- **Laki testovi** — lijevo iskošeni (nagomilani pri maksimumu)

Praktična posljedica u ekonomiji

Distribucija prihoda je jako iskosena udesno:

- Aritmetička sredina je znatno viša od medijane
- Politika bazirana na “prosječnim prihodima” može zavaravati
- Medijana bolje opisuje “tipičnog” građanina

Pr.: prosječna plata u BiH može biti visoka zbog rijetko visokih prihoda, dok medijalna plata ostaje niska.

Zahtjev

Uvijek kombinujte histogram (ili boxplot) s numeričkim mjerama.

Definicija

Medijana je srednja tačka distribucije: polovina opservacija je manja, polovina veća.

Algoritam izračuna

- 1 Poredati opservacije od najmanje do najveće
- 2 Ako je n **neparan**: M je na poziciji $\frac{n+1}{2}$
- 3 Ako je n **paran**: M je prosjek vrijednosti na pozicijama $\frac{n}{2}$ i $\frac{n}{2} + 1$

Numeričke ilustracije

Niz: **3, 5, 7, 9, 12, 15, 21** ($n = 7$)

Pozicija 4 $\Rightarrow M = 9$

Niz: **3, 5, 7, 9, 12, 15** ($n = 6$)

Prosjek 3. i 4.: $M = \frac{7+9}{2} = 8$

Ključna prednost medijane

Medijana je **robustna mjera** — na nju ne utječu ekstremne vrijednosti (outlieri). Zato je primjerenija za iskošene distribucije (prihodi, cijene nekretnina).

Definicija kvartila

Q1 = medijana donje polovine podataka (ispod M)

Q3 = medijana gornje polovine podataka (iznad M)

Ukupna medijana M nije uključena ni u jednu od polovina.

Pet-brojni sažetak

Min	Q1	M	Q3	Max
x_{\min}	25. pct	50. pct	75. pct	x_{\max}

IQR = $Q3 - Q1$ = raspon srednje 50 % podataka.

Percentili

p -ti percentil = vrijednost ispod koje se nalazi $p\%$ opservacija.

- $M = 50.$ percentil
- $Q1 = 25.$ percentil
- $Q3 = 75.$ percentil

50 % vrijednosti je jednako ili manje od medijane.

Aritmetička sredina \bar{x}

Definicija

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

Osjetljivost na outliere

Aritmetička sredina je “ravnotezna tačka” — outlieri je jako pomjeraju.

Plate (u 1000 KM): 2, 3, 3, 4, **48**

$$\bar{x} = \frac{2+3+3+4+48}{5} = 12,0$$

$$M = 3,0$$

Medijana od 3.000 KM opisuje “tipičnog” radnika daleko bolje.

Sredina vs. medijana: kada sto?

Situacija	Mjera	Zasto
Simetrična, bez outliera	\bar{x}	Koristi sve podatke
Iskosena ili s outlierima	M	Robusna
Prihodi, cijene	M	Desni rep
Visina, IQ	\bar{x}	Simetričnost

Definicija

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (\text{varijansa}) \quad s = \sqrt{s^2} \quad (\text{standardna devijacija})$$

Postupak izračuna

- 1 Izračunaj \bar{x}
- 2 Za svaku opservaciju: $(x_i - \bar{x})^2$
- 3 Zbir kvadrata podijeli s $n - 1 \rightarrow s^2$
- 4 Kvadratni korijen $\rightarrow s$

Dijelimo s $n - 1$ za nepristrasnu procjenu populacijske varijanse.

Numeričke vrijednosti

Podaci: 16, 25, 24, 19, 33 $\bar{x} = 23,4$

$x_i - \bar{x}$	$(x_i - \bar{x})^2$
-7,4	54,76
1,6	2,56
0,6	0,36
-4,4	19,36
9,6	92,16
Suma	169,2

$$s^2 = 169,2/4 = 42,3; \quad s = 6,50$$

Outlieri i njihov uticaj

Outlier (izdvojenica)

Opservacija koja numericki daleko pada od ostatka skupa podataka.

Uticaj na mjere

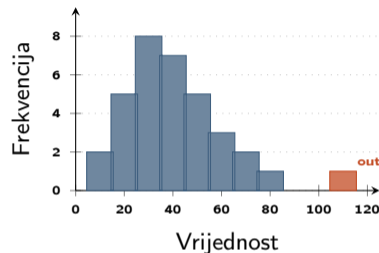
- \bar{x} i s — jako osjetljivi
- M i IQR — robusni, malo se mijenjaju

Detekcija: $1,5 \times \text{IQR}$ pravilo

Outlier je vrijednost:

ispod $Q1 - 1,5 \cdot \text{IQR}$

ili iznad $Q3 + 1,5 \cdot \text{IQR}$



Sta napraviti?

- Provjeri gresku unosa
- Ne brisati automatski
- Analizuj sa i bez njega

Uticaj linearnih transformacija na \bar{x} i s

Linearna transformacija $y = a + b \cdot x$

Transformacija	Uticaj na \bar{x}	Uticaj na s
Dodavanje konstante $+a$	$\bar{x} + a$	nepromijenjeno
Množenje konstantom $\times b$	$b \cdot \bar{x}$	$b \cdot s$
Kombinacija $y = a + b \cdot x$	$a + b \cdot \bar{x}$	$b \cdot s$

Primjer: dnevni unos vode

Student pije 2,5 mL po kg tjelesne težine, plus 750 mL.

Polazne vrijednosti: $\bar{x} = 74,4$ kg, $s = 16,47$ kg

$\bar{y} = 2,5 \times 74,4 + 750 = 936$ mL (i pomak i množenje utjecu na sredinu)

$s_y = 2,5 \times 16,47 = 41,18$ mL (samo množenje utjece na disperziju)

Kada koristiti koji sažetak?

Opste pravilo

Grafikon uvijek daje najcjelovitiju sliku distribucije. Numeričke mjere je nadopunjuju — ali ih ne mogu zamijeniti.

Distribucija	Preporučeni sažetak	Razlog
Simetrična, bez outliera	\bar{x} i s	Koriste sve podatke
Iskosena ili s outlierima	Pet-brojčani (M , $Q1$, $Q3$)	Robusnost
Vise modova (pikova)	Histogram ili boxplot	Numeričke mjere ne otkrivaju pikove
Usporedba distribucija	Uspored. boxploti	Laka vizuelna usporedba

Ceste greske

- Koristiti samo \bar{x} za iskošene distribucije (prosječna vs. medijalna plata)
- Zaboraviti mjeru disperzije uz mjeru centra
- Donositi zaključke bez prethodnog grafikona

