

Opisna statistika

Martina Gaćina

Prirodoslovno–matematički fakultet — Matematički odsjek

U Zagrebu, 1. lipnja 2020.

Sadržaj

1 Što je statistika?	1
2 Opisna statistika	2
2.1 Grafički i tabelarni prikaz podataka	3
3 Zadaci	4

1 Što je statistika?

Statistika je matematička disciplina koja obuhvaća sakupljanje, analizu, interpretaciju i prezentaciju podataka te izradu predviđanja koja se temelje na tim podacima. Veliku važnost u korištenju statistike imaju i planiranje i provođenje pokusa, odnosno sakupljanje podataka koji će se analizirati, ali i interpretacija dobivenih rezultata. O povijesti uporabe statističkih metoda i porijeklu naziva možete pročitati na [1]. Statistika se primjenjuje u mnogim strukama, kao i u svakodnevnom životu. U svakodnevnom životu ju koristimo za prikupljanje, registriranje i prikaz podataka o nekom fenomenu, npr. dnevne cijene na zagrebačkim tržnicama, broj postignutih golova u nogometnom kolu. . . . Zaključci izvedeni statističkom analizom su nesigurni jer se zasnivaju na nepotpunim podacima (npr. broj nezaposlenih u državi se procjenjuje ispitivanjem uzorka od nekoliko tisuća ljudi) ili na podacima koji u sebi sadrže slučajnu komponentu (npr. rast borova posijanih iz iste grupe sjemenki na istom tlu u istim vremenskim uvjetima). Statistika se prvenstveno bavi situacijama u kojima se pojavljivanje nekog događaja ne može predvidjeti sa sigurnošću.

Definicija 1. Sakupljeni podaci nazivaju se *uzorkom*. Familija svih podataka (potencijalnih promatranja) naziva se *populacijom*. (*Statistička*) *populacija* je potpun skup

mogućih mjerenja ili podataka o nekom kvalitativnom svojstvu koji odgovaraju cijeloj familiji jedinki o kojoj treba dati zaključak. Populacija predstavlja cilj istraživanja i svrha procesa sakupljanja podataka je izvođenje zaključka o populaciji. *Varijabla* je neko svojstvo svakog člana populacije. *Uzorak* iz statističke populacije je skup mjerenja (podataka) koja su sprovedena u toku istraživanja. \triangleleft

Ciljevi statistike su:

- Zaključivanje o populaciji iz podataka u uzorku.
- Ocjena nesigurnosti (neizvjesnosti) koje su obuhvaćene tim zaključivanjem.

Statistiku dijelimo na *deskriptivnu* i *induktivnu* te *matematičku* i *egzaktnu*.

Deskriptivna statistika (*descriptive statistics*) bavi se organizacijom sakupljenih podataka te njihovim sažetim opisom pomoću numeričkih i grafičkih prikaza.

Induktivna statistika (*inferential statistics*) bavi se izvođenjem zaključaka o populaciji na temelju svojstava uzorka.

Matematička statistika je proučavanje statistike s matematičke točke gledišta korištenjem teorije vjerojatnosti, matematičke analize i linearne algebre.

Egzaktna statistika je grana statistike koja daje točne rezultate za pripadne statističke testove.

2 Opisna statistika

Tijekom izvođenja eksperimenta ili istraživanja opaža se (mjeri) neko numeričko ili ne-numeričko *statističko obilježje* ili *varijabla* X . Rezultat mjerenja varijable X na jednoj *statističkoj jedinici* (elementu statističkog skupa) označit ćemo s x . Opažene vrijednosti varijable X na određenom statističkom skupu x_1, x_2, \dots, x_n čine skup podataka o obilježju X na tom skupu. Oznake:

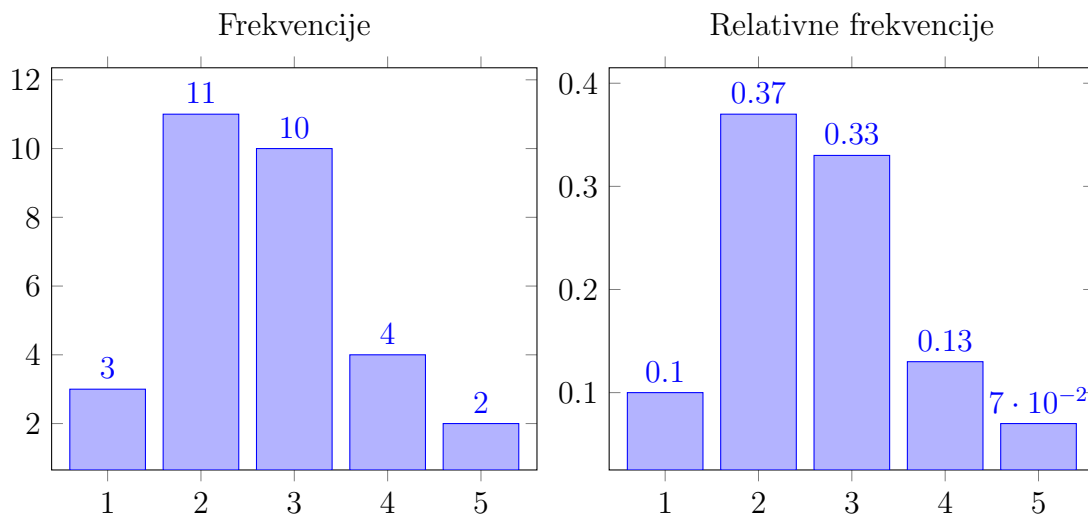
- statistički skup — Ω
- statistička jedinica — $\omega \in \Omega$
- varijabla — $X : \Omega \rightarrow K$, varijable označavamo velikim slovima: X, Y, U, V, W, \dots
- vrijednost varijable X na statističkoj jedinici ω — $x = X(\omega)$

2.1 Grafički i tabelarni prikaz podataka

Primjer 2. Uvidom u imenik jednog razreda od 30 učenika popisane su ocjene iz matematike na kraju školske godine:

1 4 2 3 1 1 2 4 3 4 5 3 2 2 3
2 5 3 2 3 3 4 2 3 2 3 3 2 2 2

Varijabla: X = „ocjena iz matematike na kraju školske godine”, slika od $X : \text{Im}X = \{1, 2, 3, 4, 5\}$. Razdioba varijable X na danom skupu podataka se opisuje *frekvencijama* ili *relativnim frekvencijama* vrijednosti od X organiziranim u *frekvencijske tablice*.



Slika 1: Stupčasti dijagram frekvencija i relativnih frekvencija

Neka je

$$f_i := \#\{\omega \in \Omega : X(\omega) = i\} \quad (1)$$

$$= \text{frekvencija vrijednosti } i \in \text{Im}X; \quad (2)$$

$$r_i := \frac{f_i}{n} \quad (3)$$

$$= \text{relativna frekvencija od } i \in \text{Im}X, \quad (4)$$

gdje je $n = \#\Omega$ *duljina* skupa podataka ($n = 30 =$ broj učenika u razredu).

Izračunajmo frekvencijsku tablicu (tablica 1). U prvi stupac zapisujemo vrijednosti ocjena iz matematike (1–5), u drugi stupac pišemo frekvenciju vrijednosti (ocjene) te u treći relativnu frekvenciju ocjene izračunate kao u (3). Pomoću frekvencijske tablice nacrtajmo *stupčasti dijagram* frekvencija i/ili relativnih frekvencija (slika 1). Vidimo da stupčasti dijagrami zapravo izgledaju isto. Dakle, svejedno nam je hoćemo li kod crtanja dijagrama koristiti frekvencije ili relativne frekvencije. Mi ćemo uglavnom crtati dijagrame pomoću relativnih frekvencija. ♡

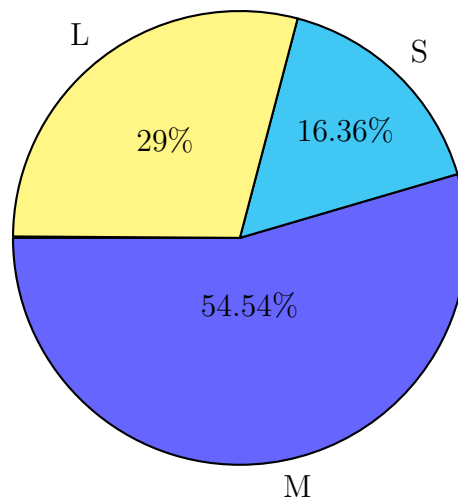
i	f_i	r_i
1	3	$\frac{3}{30} = 0.10$
2	11	$\frac{11}{30} \approx 0.37$
3	10	$\frac{10}{30} \approx 0.33$
4	4	$\frac{4}{30} \approx 0.13$
5	2	$\frac{2}{30} \approx 0.07$
Σ	30	$1(\approx 1.00)$

Tablica 1: Frekvencijska tablica

Primjer 3. Rezultati prodaje ljetnih majica tijekom tri ljetna mjeseca prikazani su frekvencijskom tablicom:

veličina	frekvencija	relativna frekvencija
S	9	0.16
M	30	0.55
L	16	0.29
Σ	55	1.00

X = „veličina majice”, $\text{Im}X = \{S, M, L\}$. Prikažimo podatke o veličinama majica pomoću *strukturnog dijagrama* (slika 2). ♡



Slika 2: Strukturni dijagram

3 Zadaci

Osnovne pojmove potrebne za rješavanje i razumijevanje zadataka možete pogledati u [2].

Zadatak 4. Kocku smo bacali 20 puta i zabilježili smo sljedeće rezultate:

$$\begin{array}{cccccccccc} 6 & 3 & 3 & 6 & 3 & 5 & 6 & 1 & 4 & 6 \\ 3 & 5 & 5 & 2 & 2 & 2 & 2 & 3 & 2 & 3 \end{array}$$

- (a) Nacrtajte histogram relativnih frekvencija.
- (b) Odredite aritmetičku sredinu, mod i medijan uzorka.
- (c) Odredite varijancu i standardnu devijaciju uzorka.
- (d) Odredite raspon uzorka.
- (e) Odredite donji i gornji kvartil te interkvartil uzorka.
- (f) Nacrtajte dijagram pravokutnika (box and whisker plot).

Rješenje. (a) Frekvencijska tablica: tablica 2, histogram relativnih frekvencija: slika 3.

i	y_i	f_i	$r_i = f_i/20$
1	1	1	0.05
2	2	5	0.25
3	3	6	0.3
4	4	1	0.05
5	5	3	0.15
6	6	4	0.2
Σ		20	1.00

Tablica 2: Frekvencijska tablica

(b) Aritmetička sredina iznosi

$$\bar{x} = \frac{f_1 y_1 + \dots + f_6 y_6}{f_1 + \dots + f_6} = \frac{1 \cdot 1 + 5 \cdot 2 + 6 \cdot 3 + 1 \cdot 4 + 3 \cdot 5 + 4 \cdot 6}{1 + 5 + 6 + 1 + 3 + 4} = 3.6 \quad (5)$$

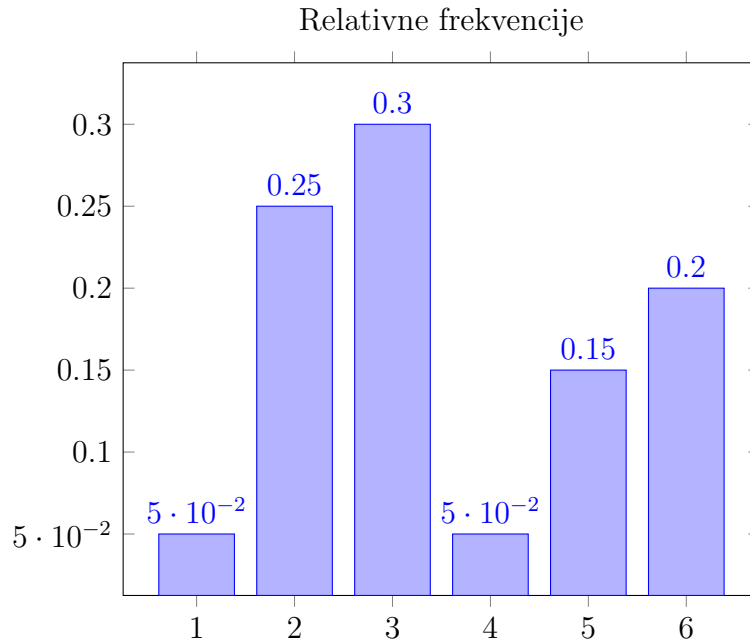
Mod uzorka je 3.

Podatke x_1, \dots, x_n poredane po veličini označavamo s $x_{(1)}, \dots, x_{(n)}$. Pri tome za $s = k + r$, gdje je $k \in \mathbb{Z}, r \in [0, 1)$, vrijedi formula

$$x_{(s)} = x_{(k)} + r [x_{(k+1)} - x_{(k)}]. \quad (6)$$

Medijan uzorka je

$$m = x_{(\frac{n+1}{2})} = x_{(\frac{20+1}{2})} = x_{(10+\frac{1}{2})} = x_{(10)} + \frac{1}{2} [x_{(11)} - x_{(10)}] = 3 + \frac{1}{2}(3 - 3) = 3. \quad (7)$$



Slika 3: Histogram relativnih frekvencija

(c) Varijanca uzorka iznosi

$$s^2 = \frac{1 \cdot 1^2 + 5 \cdot 2^2 + 6 \cdot 3^2 + 1 \cdot 4^2 + 3 \cdot 5^2 + 4 \cdot 6^2 - 20 \cdot 3.6^2}{1 + 5 + 6 + 1 + 3 + 4 - 1} = 2.673684 \quad (8)$$

pa je standardna devijacija $s = 1.64$.

(d) Raspon uzorka je $R = x_{(20)} - x_{(1)} = 6 - 1 = 5$.

(e) Donji kvartil je

$$q_L = x_{(\frac{n+1}{4})} = x_{(\frac{20+1}{4})} = x_{(5+\frac{1}{4})} = x_{(5)} + \frac{1}{4} [x_{(6)} - x_{(5)}] = 2 + \frac{1}{4}(2 - 2) = 2. \quad (9)$$

Gornji kvartil je

$$q_U = x_{(\frac{3(n+1)}{4})} = x_{(\frac{3(20+1)}{4})} = x_{(15+\frac{3}{4})} = x_{(15)} + \frac{3}{4} [x_{(16)} - x_{(15)}] = 5 + \frac{3}{4}(5 - 5) = 5. \quad (10)$$

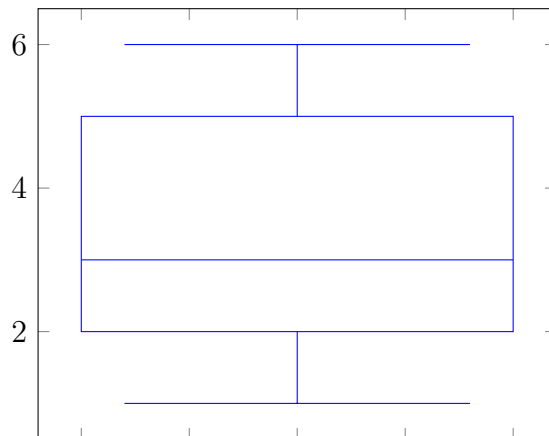
Interkvartil iznosi $d_q = q_U - q_L = 5 - 2 = 3$

(f) Dijagram pravokutnika: slika 4.

||

Napomena 5. $(x_{(1)}, q_L, m, q_U, x_{(n)})$ se zove *karakteristična petorka uzorka*.

Napomena 6. Pri formiranju dijagrama pravokutnika *outlieri* su sve vrijednosti koje su od donjeg ili gornjeg kvartila udaljene za više od $\frac{3}{2}d_q$. *Brkovi* su najmanja i najveća vrijednost koje nisu *outlieri*. *Outlieri* se posebno naznačavaju na dijagramu pravokutnika



Slika 4: Dijagram pravokutnika

Primijetite da u prethodnom zadatku nema *outliera*.

Zadatak 7. Na nekom fakultetu je odabran uzorak od 40 studenata i izmjerene su im visine:

140	188	175	176	177	168	162	181
183	187	187	162	184	161	188	169
195	171	170	199	181	169	189	191
172	182	183	178	180	165	185	205
183	187	188	182	163	179	178	188

(a) Odredite karakterističnu petorku uzorka.

(b) Nacrtajte dijagram pravokutnika.

Rješenje. (a) Gledamo uređene podatke (tablica 3). Tada je

140	161	162	162	163	165	168	169
169	170	171	172	175	176	177	178
178	179	180	181	181	182	182	183
183	183	184	185	187	187	187	188
188	188	188	189	191	195	199	205

Tablica 3: Uređeni podaci

$$n = 40, \quad (11)$$

$$x_{(1)} = 140, \quad (12)$$

$$\begin{aligned} q_L &= x_{\left(\frac{40+1}{4}\right)} = x_{\left(10+\frac{1}{4}\right)} = x_{(10)} + \frac{1}{4} [x_{(11)} - x_{(10)}] \\ &= 170 + \frac{1}{4}(171 - 170) = 170.25, \end{aligned} \quad (13)$$

$$\begin{aligned} m &= x_{\left(\frac{40+1}{2}\right)} = x_{\left(20+\frac{1}{2}\right)} = x_{(20)} + \frac{1}{2} [x_{(21)} - x_{(20)}] \\ &= 180 + \frac{1}{2}(181 - 180) = 180.5, \end{aligned} \quad (14)$$

$$\begin{aligned} q_U &= x_{\left(\frac{3(40+1)}{4}\right)} = x_{\left(30+\frac{3}{4}\right)} = x_{(30)} + \frac{3}{4} [x_{(31)} - x_{(30)}] \\ &= 187 + \frac{3}{4}(187 - 187) = 187, \end{aligned} \quad (15)$$

$$x_{(40)} = 205. \quad (16)$$

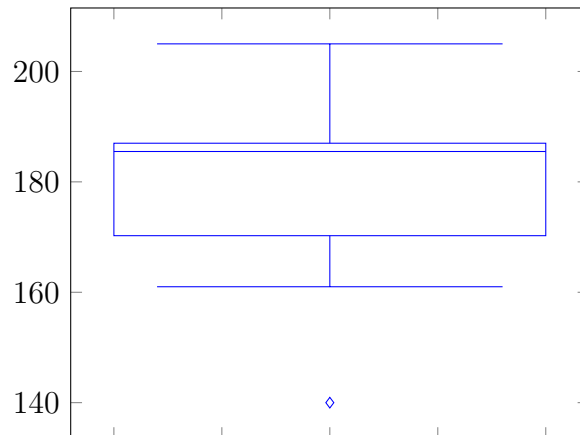
Karakteristična petorka je $(140, 170.25, 180.5, 187, 250)$.

(b) Interkvantil iznosi $d_q = q_U - q_L = 187 - 170.25 = 16.75$ pa je

$$q_L - \frac{3}{2}d_q = 145.125, \quad (17)$$

$$q_U + \frac{3}{2}d_q = 212.125. \quad (18)$$

Dijagram pravokutnika: slika 5.



Slika 5: Dijagram pravokutnika

||

Literatura

- [1] URL: https://www.pmf.unizg.hr/_download/repository/PREDAVANJE7.pdf.

- [2] Miljenko Huzak. „Predavanja iz statistike”.
- [3] Ante Mimica i Marina Ninčević. *Statistika*. 2010.

Popis slika

1	Stupčasti dijagram frekvencija i relativnih frekvencija	3
2	Strukturni dijagram	4
3	Histogram relativnih frekvencija	6
4	Dijagram pravokutnika	7
5	Dijagram pravokutnika	8

Popis tablica

1	Frekvencijska tablica	4
2	Frekvencijska tablica	5
3	Uređeni podaci	7