

Statistika

Vanja Wagner, Ivan Biočić, Hrvoje Planinić
na temelju materijala B. Basraka i M. Marušića

0. Opis kolegija

Predavanja:

Hrvoje Planinić (predavanja srijedom 16-18h 003) — hrvoje.planinic@math.hr

Ivan Biočić (predavanja srijedom 16-18h 004) — ivan.biocic@math.hr

Vježbe:

Hrvoje Olić (vježbe četvrtkom 16-18h 004 i Pr2) — hrvoje.olic@math.hr,

Krunoslav Ivanović (vježbe četvrtkom 16-18h 006 i Pr3) — krunoslav.ivanovic@math.hr

Polaganje - redovito (120 bodova)

- Dva kolokvija (prvi u terminu izvanrednih jesenskih rokova, drugi na prvom zimskom roku) po 50 bodova (uz praktični dio).
- Aktivnost i dolazak na nastavu - 10 bodova vježbe, 10 bodova predavanja (za potpis potrebno 70% dolazaka).

Polaganje - rokovi

- Ispit od 90 bodova. Na ispitu se ne uzimaju u obzir bodovi s kolokvija (aktivnost da!), a ispit može biti pismeni i/ili usmeni (i praktični).

Glavna literatura:

- B. Basrak: Statistika s primjenama u R-u – [skripta-link](#)
- Michael C. Whitlock, Dolph Schluter, The Analysis of Biological Data (2020)

Glavni objekt proučavanja u statistici su *podaci*, a osnovne aktivnosti kojima ćemo se baviti su pitanja:

- prikupljanja podataka (dizajn pokusa, *engl. experimental design*),
- uređivanja, sažimanja i prikaza podataka (deskriptivna statistika, *engl. descriptive statistics*),
- analize i zaključivanja na osnovu prikupljenih podataka (inferencijalna statistika, *engl. statistical inference*).

1. Osnovni pojmovi i prikupljanje podataka

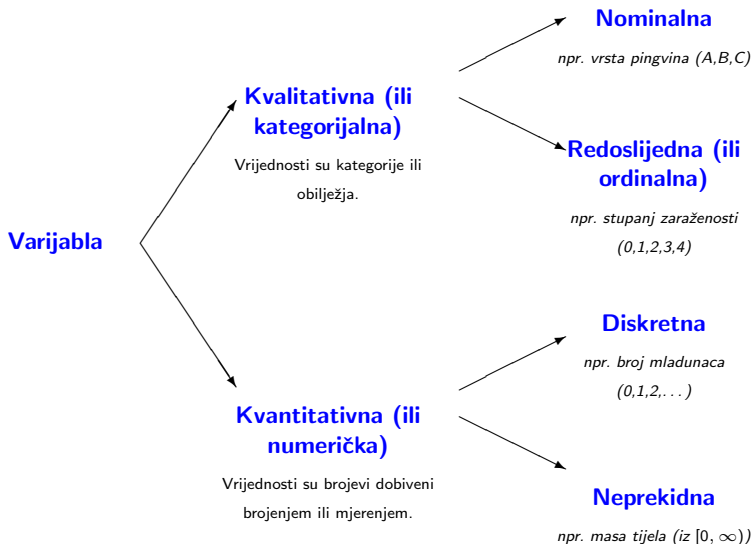
Osnovni pojmovi

- **Varijabla** je karakteristika (svojstvo, atribut, obilježje) koja može poprimiti različite vrijednosti. Primjeri varijable su: visina (osobe), težina (osobe), indeks tjelesne težine (BMI), broj sadnica u rasadniku, ocjena na ispitu, vrsta životinje, indikator je li osoba preživjela ili ne neku operaciju.
- **Podatak** je izmjerena vrijednost varijable. U statistici obično krećemo od skupa podataka (*engl.* data set). Npr. za danu varijablu *visina osobe* izmjerimo visine (u cm) 10 slučajno odabranih osoba i dobijemo skup podataka:

175, 166, 181, 159, 188, 176, 171, 183, 179, 162.

- **Populacija** je skup jedinki koje imaju neko zajedničko svojstvo, a koje su predmet proučavanja. Definicija populacije ovisi o istraživanju pa tako, na primjer, za varijablu *visina osobe* promatrana populacija može biti
 - svi ljudi na svijetu;
 - svi studenti Biološkog odsjeka PMF-a.
- **Uzorak** je (bilo koji) podskup populacije (grupa jedinki) na kojoj vršimo opažanja. Da bismo mogli valjano statistički obraditi uzorak, on mora **dobro reprezentirati** populaciju.

Klasifikacija varijabli i podataka



Primjeri varijabli

Klasificirajte sljedeće varijable:

- zemlja porijekla (populacija: natjecatelji na OI)
- dob (populacija: studenti na PMF-BO)
 - ako se mjeri precizno (npr. 19,3 godine)
 - ako se bilježi samo broj navršenih godina (19,20,21,...)
 - ako dob unaprijed grupiramo (dijete, adolescent, odrasla osoba)

Diskretna ili ordinalna varijabla? – ordinalna varijabla jest kategorijalna varijabla čije se kategorije mogu prirodno poredati, ali razlike između susjednih kategorija **nemaju jasno brojčano značenje**.

Loši podatci $\implies$ pogrešni zaključci

Kako dobro prikupiti podatke?

Neki principi prikupljanja i analize podataka

- Uzorak mora biti **reprezentativan**
- Kontrolna skupina mora biti odgovarajuća
- Hipoteze moraju biti jasne i unaprijed određene
- itd.

VAŽNO

Mnoge greške zapravo nastaju već u procesu dizajniranja pokusa, odnosno prikupljanja podataka zbog čega je daljnja statistička analiza (makar i valjano provedena) vrlo upitna.

Reprezentativnost uzorka

Reprezentativan uzorak jest podskup populacije koji *vjerno predstavlja cijelu populaciju*, tj. uzorak trebamo birati iz populacije na *slučajan i nepristran* način.

Primjer. Utjecaj obroka bogatog šećerima na rezultate studenata na ispitu

Populacija je *svi studenti prve godine PMF-MO*. Želimo istražiti kako *broj bodova na ispitu* (iz nekog kolegija) ovisi o *vrsti obroka prije ispita*. Planiramo uzorak veličine $n = 100$ studenata.

U uzorak smo uvrstili sve brucoše koji dolaze iz zagrebačkog MIOC-a, a ostatak smo odabrali slučajno. Je li to u redu?

Zašto često biramo kontrolnu skupinu?

Kod analize tretmana (npr. uzimanja lijeka), biramo **kontrolnu grupu** (u medicini tipično placebo grupu) koja je **po svemu usporediva s grupom pod tretmanom**, s ciljem da uklonimo druge brojne čimbenike koji su mogli utjecati na rezultat tretmana (vrijeme, drukčije okolnosti pri mjerenju varijable i sl.). U istraživanjima bez kontrolne skupine, puno je teže donositi ozbiljne zaključke.

Primjer. Utjecaj obroka bogatog šećerima na rezultate studenata na ispitu iz statistike (nastavak)

Recimo da smo studente izabrali na reprezentativan način, zamolili ih da dođu na test iz statistike ujutro natašte. Nakon toga studenti bivaju podjeljeni u dvije grupe – jedna dobiva obrok bogat šećerima, posebno glukozom (**testna skupina**), a druga klasični uravnoteženi doručak (**kontrolna skupina**).

Ima li smisla prvih 50 studenata koji dođu na testiranje staviti u testnu, a preostalih 50 u kontrolnu skupinu? → *randomized control trials*.

Ima li smisla da studenti znaju u kojoj su grupi? → *placebo tretman*.

Ima li smisla da ispravljači znaju u kojoj je student grupi? → *double blind*.

Nejasno postavljene hipoteze i postavljanje hipoteza *a posteriori*

Istaživači bi mogli obaviti pokus i mjerenja korektno te prikupiti veliki broj podataka, ali postaviti svoje hipoteze tek nakon inicijalne statističke analize. Tako bi npr. mogli otkriti da je obrok bogat šećerom imao pozitivan efekt na rezultate testa, no samo za studentice i to one koje dolaze iz kontinentalnih dijelova zemlje.

Ovakav pristup istraživanju može biti zanimljiv, no ponovljeni pokus će rezultate rijetko replicirati. Ovakav način analize je karakterističan u području *data mininga*. Može poslužiti za inicijalnu analizu i postavljanje hipoteza koje bi trebalo **potvrditi na novim nezavisno prikupljenim podacima**.

Važno!

Prikupljanju podataka se pristupa nakon postavljenih hipoteza, te nakon što je jasno kojom tehnikom će se podatci obraditi!

2. Deskriptivna statistika

Cilj

Vjerno, indikativno i jasno predočiti podatke koristeći brojeve, grafičke prikaze i druga sredstva!

Grafički prikaz kategorijalnih podataka

Primjer: Na zahtjev vlasnika farme, radnik je bilježio spol 397 zečeva. Sljedeći dan je rezultat proslijedio svome nadređenom:

M M Ž Ž M Ž M M Ž M Ž M Ž M Ž Ž M Ž Ž M Ž Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž
 M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M M M Ž M Ž M Ž M Ž M Ž Ž
 M Ž M Ž M Ž Ž Ž M Ž M Ž Ž M Ž M Ž M Ž M Ž M Ž M Ž Ž M Ž M Ž M Ž M Ž M Ž
 M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž Ž M Ž M Ž M Ž Ž M Ž M Ž
 M Ž Ž Ž Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž M Ž Ž M Ž Ž M Ž M Ž Ž M Ž
 M Ž M Ž Ž M Ž Ž M Ž Ž M Ž M Ž M Ž M Ž Ž M Ž M Ž Ž Ž M Ž Ž M Ž Ž M Ž M Ž Ž
 Ž M Ž M Ž M Ž M Ž Ž M Ž M Ž M Ž M Ž Ž M Ž Ž M Ž M Ž M Ž Ž M Ž M Ž M
 Ž M Ž Ž M Ž Ž M Ž M Ž Ž M Ž M Ž M Ž Ž M Ž M Ž M Ž Ž M Ž M Ž Ž M Ž M Ž Ž
 M Ž M Ž Ž M Ž M Ž Ž M Ž M Ž Ž M Ž M Ž Ž M Ž M Ž Ž M Ž M Ž Ž M Ž M Ž
 Ž M Ž M Ž Ž M Ž M Ž Ž M Ž M Ž Ž M Ž Ž M Ž Ž M Ž M Ž Ž M Ž M Ž M Ž Ž M Ž
 M Ž M Ž M Ž M Ž M Ž Ž M Ž M Ž Ž M Ž Ž M Ž M Ž Ž M Ž M Ž M Ž M Ž M Ž

Ni ovo nije puno bolje...

M
M
M
M
M
M
M
Ž
Ž
Ž
Ž
Ž
Ž
Ž
Ž Ž Ž Ž Ž Ž Ž

Tablica frekvencija

U nizu podataka M se pojavljuje 172 puta, a Ž 225 puta. Dovoljno je znati:

$$f_1 = 172, \quad f_2 = 225.$$

→ **Frekvencija** (apsolutna frekvencija) je broj pojavljivanja određene vrijednosti varijable (u skupu podataka).

Alternativno, podatke smo mogli reprezentirati i udjelima M, odnosno Ž podataka u nizu podataka - 0.4332 (43.32%) podataka je M, a 0.5668 (56.68%) podataka je Ž:

$$r_1 = 0.4332, \quad r_2 = 0.5668.$$

→ **Relativna frekvencija** je omjer broja pojavljivanja određene vrijednosti varijable (u skupu podataka) i ukupne veličine podataka (**opseg podataka**).

Ove podatke možemo jednostavno prikazati i tablično

Spol	frekvencija	relativna frekvencija
M	172	0.4332
Ž	225	0.5668
Σ	397	1

Frenkvencijska tablica - opći oblik:

Vrijednost varijable	frekvencija	rel. frekvencija	rel. kumulativne frekvencije
a_1	f_1	r_1	F_1
a_2	f_2	r_2	F_2
$\cdot$	$\cdot$	$\cdot$	$\cdot$
$\cdot$	$\cdot$	$\cdot$	$\cdot$
$\cdot$	$\cdot$	$\cdot$	$\cdot$
a_k	f_k	r_k	$F_k (= 1)$
Σ	n	1	

- $F_1 = r_1, F_2 = r_1 + r_2, \dots$
- kumulativne frekvencije imaju smisla ako su vrijednosti $a_1, a_2, \dots, a_k$ **uređene**, tj. radi se ordinalnoj ili diskretnoj varijabli

Primjer.

Promotrimo frekvencijsku tablicu kojom prikazujemo dob 20 ispitanika:

Dob	Frekvencija	Relativna frekvencija	Kumulativna frekvencija	Relativna kumulativna frekvencija
19	1	0.05	1	0.05
20	0	0.00	1	0.05
21	5	0.25	6	0.30
22	2	0.10	8	0.40
23	2	0.10	10	0.50
24	1	0.05	11	0.55
25	4	0.20	15	0.75
26	1	0.05	16	0.80
27	0	0.00	16	0.80
28	1	0.05	17	0.85
29	0	0.00	17	0.85
30	1	0.05	18	0.90
31	1	0.05	19	0.95
32	1	0.05	20	1.00
Σ	20	1	-	-

Primjer. Promotrimo bodove na kratkom testu u jednoj generaciji od 200 studenata.

Bodovi na testu	Frekvencija	Kumulativna frekvencija	Relativna frekvencija	Relativna kumulativna frekvencija (%)
0	25	25	0.125	12.5
1	17	42	0.085	21.0
2	15	57	0.075	28.5
3	23	80	0.115	40.0
4	14	94	0.070	47.0
5	15	109	0.075	54.5
6	33	142	0.165	71.0
7	25	167	0.125	83.5
8	12	179	0.060	89.5
9	9	188	0.045	94.0
10	12	200	0.060	100.0
Ukupno	200		1.00	

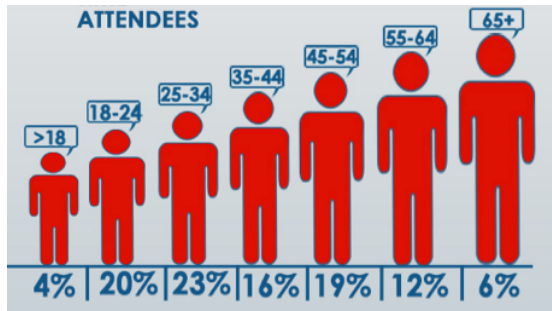
- ❶ Koliko je studenata dobilo 5 ili 6 bodova? $15 + 33 = 142 - 94 = 48$
- ❷ Za izlazak na kolokvij potrebno je ostvariti minimalno 5 bodova na kratkom testu.
Koliko je postotak studenata koji su ostvarili izlazak na kolokvij? $100\% - 47\% = 53\%$

Grafički prikaz podataka

Poželjna svojstva

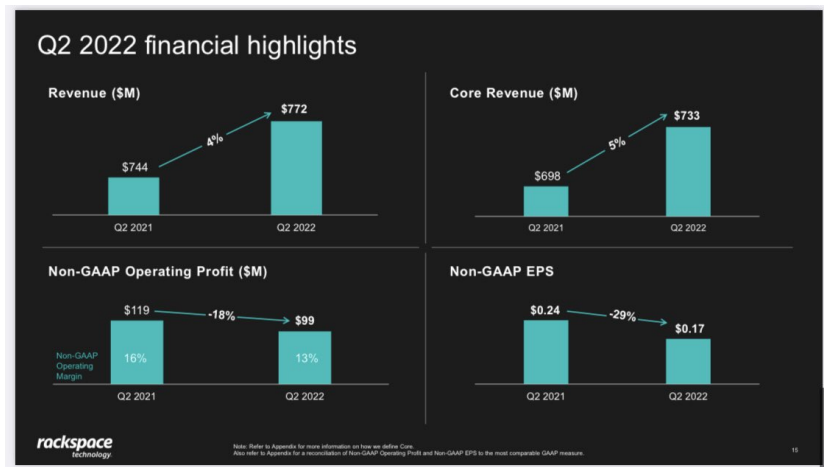
- jednostavnost
- preglednost
- indikativno prikazuju podatke

Primjer: Neprecizni i nepregledni grafički prikazi.



- veličina osoba **nije povezana** sa postocima ispod

Primjer: Neprecizni i nepregledni grafički prikazi.

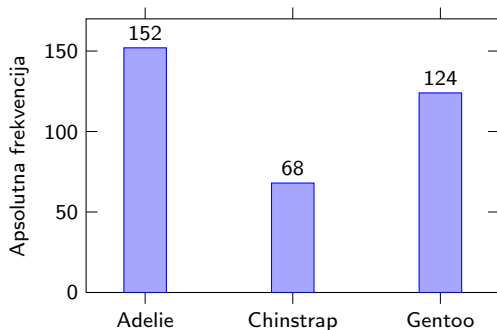


- rast od 4% izgleda značajniji nego pad od 18%

Stupčasti dijagram (*barplot*)

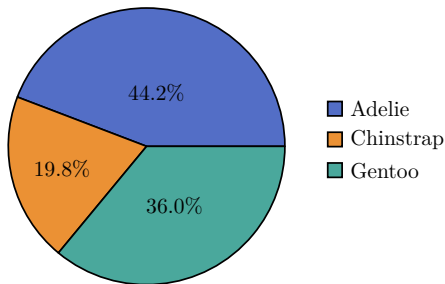
Za kategorijalne podatke **stupčasti dijagram** prikazuje apsolutne ili relativne frekvencije pojedinih kategorija.

Primjer: Podaci Palmer Penguins – 344 odrasla pingvina triju vrsta, opažena na Palmerovu arhipelagu na Antarktici.



Vrsta	<i>f</i>	<i>r</i>
Adelie	152	0.442
Chinstrap	68	0.198
Gentoo	124	0.360
Ukupno	344	1.000

Tortni dijagram (*pie-chart*)



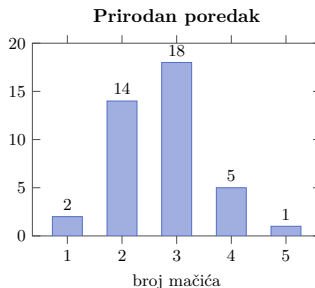
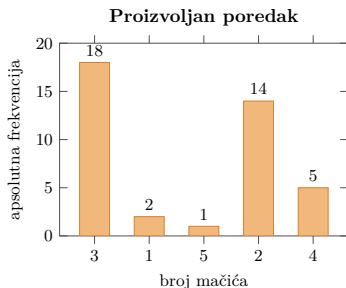
- (Prečesto) popularna alternativa stupčastom dijagramu.
- Usporedba kutova i površina nije jednostavna.
- Male razlike među kategorijama teško se uočavaju.
- Veći broj kategorija brzo čini prikaz nepreglednim.

Stupčasti dijagram za uređene podatke

Stupčasti dijagram možemo koristiti i za

- redosljedne kategorijalne podatke;
- diskretne numeričke podatke.

Primjer: broj mačića pri prvom okotu za uzorak od 40 mačaka. Riječ je o diskretnoj numeričkoj varijabli.



Kod uređenih podataka vrijednosti na x -osi prikazujemo u **prirodnom poretku**.

Od podataka do histograma

Promatramo **duljinu peraje** 342 pingvina, izraženu u milimetrima.

181, 186, 195, 193, 190, 181, 195, 193, 190, 186, ...

Duljina peraje je **neprekidna kvantitativna varijabla**, iako je izmjerena s preciznošću od jednog milimetra.

Zbog prevelikog broja različitih vrijednosti podatke grupiramo u **razrede**, npr.

[170, 180), [180, 190), [190, 200), ..., [230, 240).

Histogram dobijemo tako da iznad svakog razreda crtamo pravokutnik čija je **površina** jednaka **relativnoj frekvenciji** tog razreda, tj. udjelu podataka koji je upao u taj razred.

Drugim riječima, visina iznad svakog razreda je

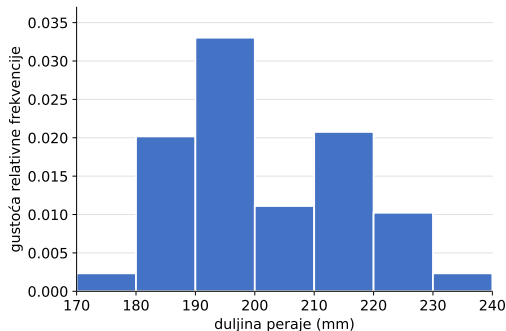
$$\text{visina} = \frac{\text{relativna frekvencija razreda}}{\text{širina razreda}}.$$

Histogram duljine peraje

Razred	f_j	r_j
[170, 180)	8	0,0234
[180, 190)	69	0,2018
[190, 200)	113	0,3304
[200, 210)	38	0,1111
[210, 220)	71	0,2076
[220, 230)	35	0,1023
[230, 240)	8	0,0234

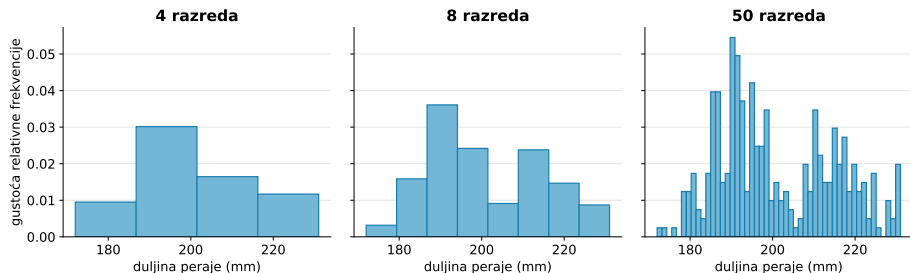
Visina stupca iznad j -tog razreda je

$$h_j = \frac{r_j}{10} = \frac{f_j}{342 \cdot 10}.$$



- za razliku od stupčastog dijagrama, ovdje su bitne i **širine** stupaca, tj. razreda – npr. ukupna površina histograma, tj. svih stupaca je uvijek 1.
- ako su svi razredi **jednake širine** (kao u gornjem primjeru), često se crtaju visine jednake apsolutnim frekvencijama – "histogram apolutnih frekvencija". Oblik histograma ostaje isti.

Koliko razreda odabrati?



- **Premalo razreda:** mogu se izgubiti važna obilježja podataka.
- **Previše razreda:** histogram naglašava slučajne varijacije.
- U praksi je ključan "dobar" izbor broja razreda, što nije uvijek lagan zadatak.

Usput: dva izraženija vrha (u srednjem histogramu) mogu se objasniti razlikama među vrstama. Prosječne duljine peraje iznose približno 190 mm za Adelie, 196 mm za Chinstrap i 217 mm za Gentoo pingvine. Uočite da se to ne vidi na prvom i trećem histogramu.

Mjere srednje vrijednosti

- Cilj nam je opisati naš (često jako) velik skup podataka pomoću **nekoliko brojeva** koji nam daju preglednu i jasnu informaciju o našim podacima.
- Za početak ćemo vidjeti nekoliko mjera koje nam govore gdje je "**sredina**" našeg uzorka:
 - aritmetička sredina (prosjek);
 - medijan;
 - mod.

Aritmetička sredina ili prosjek

Primjer. Prosječna ocjena: 3, 3, 4, 2, 5, 4.

$$\text{Prosječna ocjena} = \frac{3 + 3 + 4 + 2 + 5 + 4}{6} = \frac{21}{6} = 3.5.$$

Općenito, za niz numeričkih podataka $x_1, x_2, \dots, x_n$, **aritmetička sredina** ($\bar{x}$ ili $\bar{x}_n$) se definira kao

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

Σ - oznaka za sumiranje. Još se koristi i $\sum_i x_i$, $\sum x_i$, $\sum x$.

Računanje srednje vrijednosti iz frekvencija

Recimo da umjesto podataka $x_1, x_2, x_3, \dots, x_n$ imamo danu frekvencijsku tablicu:

vrijednost	a_1	a_2	a_3	$\dots$	a_k	$\sum$
frekvencija	f_1	f_2	f_3	$\dots$	f_k	n

gdje su

- $a_1, a_2, a_3, \dots, a_k$ različite vrijednosti promatranog obilježja (ukupno $k \in \mathbf{N}$ različitih vrijednosti);
- $f_1, f_2, f_3, \dots, f_k$ pripadne frekvencije (podsjetimo se da je ukupna duljina podataka $n = f_1 + f_2 + f_3 + \dots + f_k$).

Tada je

$$\bar{x} = \frac{\sum_{i=1}^k f_i x_i}{n}.$$

Računanje srednje vrijednosti iz relativnih frekvencija

$$\bar{x} = \left(\frac{1}{n} \sum_{i=1}^k f_i x_i = \sum_{i=1}^k \frac{f_i}{n} x_i \right) = \sum_{i=1}^k r_i x_i.$$

Primjer. Podaci o broju mladunaca lastavica dani su sljedećom frekvencijskom tablicom:

broj mladunaca u gnijezdu x_j	2	3	4	5	6	7	Σ
frekvencija f_j	1	10	7	26	11	2	57
relativna frekvencija r_j	0.018	0.175	0.123	0.456	0.193	0.035	1.000

Prosječni broj mladunaca po gnijezdu je

$$\bar{x} = \frac{2 \cdot 1 + 3 \cdot 10 + 4 \cdot 7 + 5 \cdot 26 + 6 \cdot 11 + 7 \cdot 2}{57} = \frac{270}{57} \approx 4.737.$$

ili preko relativnih frekvencija:

$$\begin{aligned}\bar{x} &= 2 \cdot 0.018 + 3 \cdot 0.175 + 4 \cdot 0.123 + 5 \cdot 0.456 + 6 \cdot 0.193 + 7 \cdot 0.035 \\ &= 4.736.\end{aligned}$$

Problem sa aritmetičkom sredinom

- Aritmetička sredina je vrlo osjetljiva na **ekstremne podatke** u uzorku!
- na primjer, u nekom poduzeću radi 6 radnika s mjesečnim plaćama: 800, 1000, 1000, 1000, 1000, 1200, te direktor s mjesečnom plaćom od 5000 EUR;
- prosječna plaća "običnih" radnika je 1000, ali ako gledamo prosječnu plaću svih zaposlenih ona je 1571.42 EUR $\rightsquigarrow$ ima li smisla taj broj shvatiti kao plaću *tipičnog/prosječnog* zaposlenika?

Poželjna je i mjera za sredinu podataka koja je manje osjetljiva na ekstremne vrijednosti u uzorku, tj. koja je *robustnija*, a mi ćemo obraditi **medijan i mod**.

Medijan

- numeričke podatke $x_1, \dots, x_n$ poredamo po veličini:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}.$$

- na primjer, $x_{(1)}$ je najmanja, a $x_{(n)}$ je najveći broj u uzorku;
- **medijan** se definira kao **vrijednost** m koja uzorak dijeli na dva jednaka dijela, tj. 50% uzorka je manje ili jednako od m , te 50% veće ili jednako od m .
- problem je što često postoje **više takvih brojeva**, pa medijan formalno definiramo na sljedeći način:

- Ako je broj podataka neparan - medijan je središnji član niza podataka

$$m = x_{(\frac{n+1}{2})}.$$

- Ako je broj podataka paran - medijan je aritmetička sredina dvaju središnjih članova niza podataka

$$m = \frac{1}{2} \left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right).$$

Primjer. Odredite medijan za podatke

7, 8, 17, 9, 4, 9, 10, 11, 11, 3, 6, 7, 11, 5, 9

Podatke poredamo po veličini:

3	4	5	6	7	7	8	9	9	9	10	11	11	11	17	podaci
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	redni broj

$m = 9$

Primjer. Odredite medijan za podatke

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

Podatke poredamo po veličini:

1	1	2	2	2	3	3	4	4	5	podaci
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	redni broj

$$m = \frac{2 + 3}{2} = 2.5$$

Osjetljivost na ekstremne vrijednosti

- u primjeru s plaćama, imali smo plaće:

800, 1000, 1000, 1000, 1000, 1200, 5000 ,

- prosječna plaća je 1571.42, a medijan je 1000;
- koliki bi bio medijan ako bi direktor imao plaću 10000 EUR?

Određivanje medijana iz distribucije frekvencija.

Dob	Frekvencija	Relativna frekvencija	Kumulativna frekvencija	Relativna kumulativna frekvencija
19	1	0.05	1	0.05
20	0	0.00	1	0.05
21	5	0.25	6	0.30
22	2	0.10	8	0.40
23	2	0.10	10	0.50
24	1	0.05	11	0.55
25	4	0.20	15	0.75
26	1	0.05	16	0.80
27	0	0.00	16	0.80
28	1	0.05	17	0.85
29	0	0.00	17	0.85
30	1	0.05	18	0.90
31	1	0.05	19	0.95
32	1	0.05	20	1.00

Dob	Frekvencija	Relativna frekvencija	Kumulativna frekvencija	Relativna kumulativna frekvencija
19	1	0.05	1	0.05
20	0	0.00	1	0.05
21	5	0.25	6	0.30
22	2	0.10	8	0.40
23	2	0.10	10	0.50
24	1	0.05	11	0.55
25	4	0.20	15	0.75
26	1	0.05	16	0.80
27	0	0.00	16	0.80
28	1	0.05	17	0.85
29	0	0.00	17	0.85
30	1	0.05	18	0.90
31	1	0.05	19	0.95
32	1	0.05	20	1.00

- 10 (50%) podataka je manje ili jednako od 23
- 10 (50%) podataka je veće ili jednako od 24

Znači, 23 i 24 su na središnjem mjestu u nizu podataka.

$$m = \frac{23 + 24}{2} = 23.5$$

Mod

- **Mod** je vrijednost varijable s najvećom frekvencijom, tj. najčešća vrijednost.

Primjer. Odredimo mod za podatke

ACCABACABCAABBCBAACCBAAABCCBCAA.

Podatke prvo presložimo tako da bismo lakše odredili frekvencije pojedinih vrijednosti:

AAAAAAAAAAAA BBBBBBBB CCCCCCCCCC

Mod je A jer se pojavljuje najviše puta (12).

- Može postojati i više modova, npr. u nizu

1 1 2 2 2 3 4 4 5 5 5 6 7 7

modovi su i 2 i 5.

- također, mod može biti i najmanja i najveća vrijednost u uzorku;
- Iz ovih razloga mod se i ne koristi jako često kao mjera sredine uzorka.