

Roman Brilej, Boro Nikič

# alfa

## **Obdelava podatkov**

Zbirka nalog za matematiko v  
srednjem strokovnem izobraževanju

Ljubljana 2020

# Kazalo

<b>Obdelava podatkov</b>	<b>5</b>
1. Osnovni statistični pojmi . . . . .	6
2. Urejanje podatkov . . . . .	12
3. Prikazovanje podatkov . . . . .	32
4. Merila za osrednjost podatkov . . . . .	55
5. Merila za razpršenost številskih podatkov . . . . .	76
6. Povezanost podatkov . . . . .	95
<b>Rešitve</b>	<b>115</b>

## 4. Merila za osrednjost podatkov

### Aritmetična sredina

Naј bodo  $x_1, x_2, \dots, x_n$  številski podatki, ki nas pri neki obdelavi podatkov zanimajo. **Aritmetično sredino** teh podatkov definiramo tako:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Torej aritmetično sredino podatkov izračunamo tako, da vse podatke seštejemo in dobljeno vsoto delimo s številom teh podatkov. Aritmetični sredini pogostokrat rečemo kar **povprečje**.

Aritmetično sredino  $n$  podatkov, kjer podatek  $x_1$  nastopa s frekvenco  $f_1$ , podatek  $x_2$  s frekvenco  $f_2 \dots$  in podatek  $x_k$  s frekvenco  $f_k$ , pri čemer je  $f_1 + f_2 + \dots + f_k = n$ , lahko izračunamo tudi tako:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_k f_k}{f_1 + f_2 + \dots + f_k}$$

Tako aritmetično sredino imenujemo tudi **tehtana aritmetična sredina (tehtano povprečje)**.

Če številskih podatkov ne poznamo, poznamo pa njihovo frekvenčno porazdelitev po razredih (razredi so intervali), lahko aritmetično sredino podatkov ocenimo. Naј bodo števila  $x_1, x_2, \dots, x_k$  sredine  $k$  razredov, v katere so podatki razvrščeni. Aritmetično sredino takih podatkov izračunamo po naslednji formuli:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_k f_k}{f_1 + f_2 + \dots + f_k}$$

### Mediana

**Mediana** ali **središčnica (Me) številskih podatkov** je število, za katero je vsaj 50 % podatkov manjših ali enakih in vsaj 50 % podatkov večjih ali enakih od tega števila.

Za naše namene bo mediana število, ki je določeno z naslednjo metodo:

1. Podatke razvrstimo v vrsto od najmanjšega do največjega.
2. Mediana lihega števila podatkov je srednji podatek v tej vrsti, mediana sodega števila podatkov pa je aritmetična sredina srednjih dveh podatkov.

Primeri:

- a) 1, 3, 7, 10, 20;  $Me = 7$   
b) 1, 3, 7, 9, 10, 20;  $Me = \frac{7+9}{2} = 8$

### Modus

**Modus** ali **gostiščnica (Mo)** podatkov je tisti podatek, ki se pojavi najpogosteje (podatek z največjo frekvenco). Modusov je lahko več. Če se vsi podatki pojavijo enako mnogokrat, modus ne obstaja. Modus je edino merilo za osrednjost, ki je primerno tudi za opisne podatke.

## Zgledi

1. Izračunaj povprečje podatkov 12, 8, -5, 9.

*Rešitev:* Povprečje ali aritmetično sredino  $\bar{x}$  podatkov  $x_1, x_2, \dots, x_n$  izračunamo kot vsoto teh podatkov, deljeno s številom podatkov:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Tako je v našem primeru povprečje teh podatkov enako:

$$\frac{12 + 8 + (-5) + 9}{4} = 6$$

2. Preglednica prikazuje najvišje in najnižje temperature za posamezna leta v obdobju 2008–2017 za nekatera ameriška mesta.

	New York		Los Angeles		Chicago		Miami		Phoenix	
	maks. temp. (°C)	min. temp. (°C)	maks. temp. (°C)	min. temp. (°C)	maks. temp. (°C)	min. temp. (°C)	maks. temp. (°C)	min. temp. (°C)	maks. temp. (°C)	min. temp. (°C)
2017	34	-13	40	4	35	-19	37	8	48	2
2016	36	-18	40	4	34	-21	35	8	48	1
2015	36	-17	38	2	35	-22	36	6	47	-1
2014	33	-16	39	4	34	-26	35	8	47	2
2013	37	-12	36	1	36	-19	34	7	48	-2
2012	38	-11	39	4	41	-14	34	6	47	3
2011	40	-14	37	3	38	-21	36	7	48	-1
2010	39	-11	45	3	35	-18	36	2	46	0
2009	33	-14	38	4	34	-28	37	3	46	2
2008	36	-12	38	4	35	-21	36	4	46	1

- a) Izračunaj povprečne najnižje temperature v tem obdobju za posamezna mesta in vrednosti nato uredi od najnižje do najvišje.

*Rešitev:* Izračunajmo povprečne najnižje temperature. Številke v desnem stolpcu pri posameznem mestu seštejemo in delimo z njihovim številom, torej z 10:

$$\text{New York: } \frac{-13 + (-18) + (-17) + \dots + (-12)}{10} = -13,8^\circ\text{C}$$

$$\text{Los Angeles: } \frac{4 + 4 + 2 + 4 + 1 + 4 + 3 + 3 + 4 + 4}{10} = 3,3^\circ\text{C}$$

$$\text{Chicago: } \frac{(-19) + (-21) + (-22) + \dots + (-21)}{10} = -20,9^\circ\text{C}$$

$$\text{Miami: } \frac{8 + 8 + 6 + 8 + 7 + 6 + 7 + 2 + 3 + 4}{10} = 5,9^\circ\text{C}$$

$$\text{Phoenix: } \frac{2 + 1 + -1 + 2 + -2 + 3 + -1 + 0 + 2 + 1}{10} = 0,7^\circ\text{C}$$

Podatke prikažemo v preglednici, urejene, kot zahteva naloga:

	Chicago	New York	Phoenix	Los Angeles	Miami
<b>Povprečna najnižja temperatura (°C)</b>	-20·9	-13·8	0·7	3·3	5·9

- b) V katerem mestu in katerega leta je bil razpon med najvišjo in najnižjo temperaturo največji?

*Rešitev:* Izračunajmo temperaturne razpore. Od najvišje odštejemo najnižjo temperaturo.

	New York	Los Angeles	Chicago	Miami	Phoenix
2017	47 °C	36 °C	54 °C	29 °C	46 °C
2016	54 °C	36 °C	55 °C	27 °C	47 °C
2015	53 °C	36 °C	57 °C	30 °C	48 °C
2014	49 °C	35 °C	60 °C	27 °C	45 °C
2013	49 °C	35 °C	55 °C	27 °C	50 °C
2012	49 °C	35 °C	55 °C	28 °C	44 °C
2011	54 °C	34 °C	59 °C	29 °C	49 °C
2010	50 °C	42 °C	53 °C	34 °C	46 °C
2009	47 °C	34 °C	62 °C	34 °C	44 °C
2008	48 °C	34 °C	56 °C	32 °C	45 °C

Po pregledu vseh številok opazimo, da je bil največji temperaturni razpon 62 °C leta 2009 v Chicagu.

- c) V katerem mestu ležijo najnižje temperature za posamezna leta v danem obdobju v najožjem intervalu?

*Rešitev:* Tokrat pogledimo za posamezno mesto največjo in najmanjšo številko v stolpcu najnižjih temperatur. Tako lahko ugotovimo, da ležijo za New York te temperature na intervalu od  $-18^{\circ}\text{C}$  do  $-11^{\circ}\text{C}$ . Širina tega intervala je  $7^{\circ}\text{C}$ . Nadalje ugotovimo, da so najnižje temperature za Los Angeles na intervalu  $[1, 4]$ , za Chicago na  $[-28, -14]$ , za Miami na  $[2, 8]$  in za Phoenix na  $[-2, 3]$ . Najožji interval je pri Los Angelesu, in sicer znaša njegova širina samo  $3^{\circ}\text{C}$ .

### 3. Določi mediano podatkov:

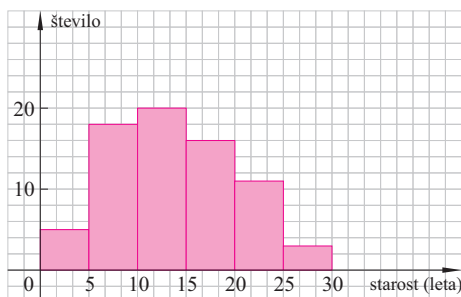
- a) 2, 5, 1, 13, 7

*Rešitev:* Da bomo lahko določili mediano, podatke najprej uredimo po velikosti:

$$1, 2, 5, 7, 13$$

Ker je število podatkov liho, je mediana podatek, ki je na sredini, torej 5.

15. Oцени povprečno vrednost podatkov, predstavljenih s histogramom.



*Rešitev:* S histograma lahko razberemo, da so podatki starosti, ki so razdeljene v posamezne razrede. Povprečno starost bomo ocenili podobno, kot smo vajeni, kadar delamo z razredi. Sredine razredov moramo pomnožiti z ustreznimi frekvencami in nato dobljeno vsoto deliti s številom podatkov, torej z vsoto vseh frekvenc. Frekvenco za posamezni razred preberemo s histograma kot višino stolpca. Tako lahko po vrsti preberemo naslednje frekvence: 5, 18, 20, 16, 11 in 3. Sredine posameznih razredov določimo kar na pamet, računsko pa je npr. za prvi razred:

$$\frac{0 + 5}{2} = 2{,}5$$

Ocenimo povprečno starost:

$$\begin{aligned} \bar{x} &\doteq \frac{5 \cdot 2{,}5 + 18 \cdot 7{,}5 + 20 \cdot 12{,}5 + 16 \cdot 17{,}5 + 11 \cdot 22{,}5 + 3 \cdot 27{,}5}{5 + 18 + 20 + 16 + 11 + 3} = \\ &= \frac{1\,007{,}5}{73} \doteq 13{,}8 \text{ leta} \end{aligned}$$

## Naloge

67. Izračunaj povprečje podatkov:

a) 3, 5, 7

b) -1, 1

c) 100

d) -2, 6, -5, 3, 10, -9, -3

e) 4, 4, 4, 4, 4, 4, 4, 4, 4

f)  $\frac{1}{2}, \frac{3}{4}, \frac{5}{4}, \frac{5}{6}, 1, \frac{2}{3}$

g) 2·3, -4·7, 5·5, 7·6, -12·4, 6·1, -3·3

h)  $5, \frac{3}{4}, 2{,}7, 0\bar{4}, 0$

68. Izračunaj povprečje podatkov in rezultat izrazi z največjo enoto, ki se pojavlja v podatkih:

a) 16 EUR, 12 EUR, 730 centov

b) 1 kg, 400 g, 70 dag, 3·3 kg, 160 dag

c) 6200 dm, 79000 cm, 70 m, 0·01 km

d) 300 min, 1·2 ure, 1·5 dni

e) 3 l, 22 dcl

69. Izračunaj povprečno dolžino imen dijakov svojega razreda. Za kateri spol je značilna daljša povprečna dolžina imen?

70. V preglednici so podatki o desetih filmih, ki so prinesli največ zaslужka v letu 2007. Izračunaj povprečni zaslужek omenjenih filmov.

	Film	Zaslужek (\$)
1	Spider-Man 3	336 530 303
2	Shrek Tretji	321 012 359
3	Transformerji	319 071 806
4	Pirati s Karibov: Na robu sveta	309 420 425
5	Harry Potter in Feniksov red	292 004 738
6	Bournov ultimat	227 471 070
7	300	210 614 939
8	Ratatouille	206 445 654
9	Simpsonovi	183 132 370
10	Divjaka	168 273 550

Vir: <http://www.kolosej.si/>

71. V preglednicah so podatki o najvišjih in najnižjih dnevnih izmerjenih temperaturah za prvih deset junijskih dni leta 2008 za mesta Ljubljana, Maribor, Novo mesto in Portorož.

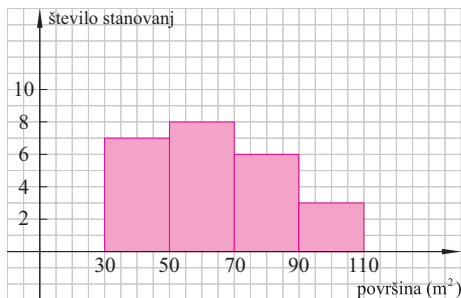
Datum/mesto	Najvišja izmerjena temperatura			
	Ljubljana	Maribor	Novo mesto	Portorož
01. 06. 2008	28°3	29°5	29	26°1
02. 06. 2008	25°8	27°4	25°8	28°6
03. 06. 2008	22°6	25°6	25°6	26°3
04. 06. 2008	23°4	23°8	24°4	22°4
05. 06. 2008	18°9	18°3	21	22°5
06. 06. 2008	20	21°5	18°7	22°2
07. 06. 2008	23°7	20°1	22°2	23°5
08. 06. 2008	22°7	20°4	21°8	26°7
09. 06. 2008	23°2	24	23	26°6
10. 06. 2008	26°2	26°5	26°1	26°9

Datum/mesto	Najnižja izmerjena temperatura			
	Ljubljana	Maribor	Novo mesto	Portorož
01. 06. 2008	14°8	15°7	14°9	15°6
02. 06. 2008	17°2	17°8	15°2	15°4
03. 06. 2008	15°8	15°8	16°1	17°1
04. 06. 2008	15°8	15°2	16°1	17°3
05. 06. 2008	16°4	16°6	16°5	16°9
06. 06. 2008	15°4	14°7	15°2	16
07. 06. 2008	14°7	15°1	14°7	14
08. 06. 2008	14°9	15°6	15°2	14°5
09. 06. 2008	14°1	14°7	15°1	14°2
10. 06. 2008	14°5	13°7	12°4	14

Vir: <http://www.arso.gov.si/> (letopis 2008)

51. a)

Površina (m <sup>2</sup> )	$f_k$
od 30 do pod 50	7
od 50 do pod 70	8
od 70 do pod 90	6
od 90 do pod 110	3



b) V razredu od 50 m<sup>2</sup> do pod 70 m<sup>2</sup>.

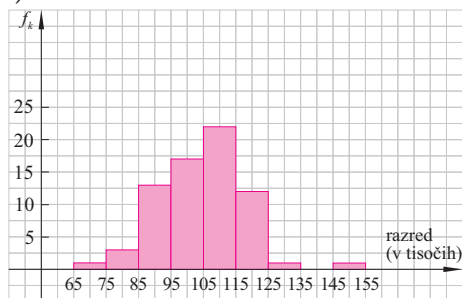
53. a) stolpčni diagram    b) histogram    c) histogram    d) stolpčni diagram  
e) stolpčni diagram

54. a) diskretni

b)

Razred (v tisočih)	$f_k$
od 65 do pod 75	1
od 75 do pod 85	3
od 85 do pod 95	13
od 95 do pod 105	17
od 105 do pod 115	22
od 115 do pod 125	12
od 125 do pod 135	1
od 135 do pod 145	0
od 145 do pod 155	1

c)



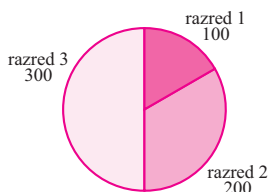
55. a) Najmanj priljubljena sta izdelka B in C. Najbolj priljubljen izdelek E je izbralo 360 oseb.

b) Za izdelek A je glasovalo 25 % anketiranih oseb.

c)

Razred	$f_k$	$f_k^0$
A	270	0,250
B	135	0,125
C	135	0,125
D	180	0,167
E	360	0,333

56. a)  $\alpha_1 = 60^\circ$ ,  $\alpha_2 = 120^\circ$ ,  $\alpha_3 = 180^\circ$



b)  $\alpha_1 = 90^\circ$ ,  $\alpha_2 = 90^\circ$ ,  $\alpha_3 = 120^\circ$ ,  
 $\alpha_4 = 60^\circ$

