

MATEMATIKA

1. letnik – splošna gimnazija

Jan Kastelic

Gimnazija Antona Aškerca,
Šolski center Ljubljana

23. februar 2025

1 Statistika

Section 1

Statistika

1 Statistika

- Osnovni pojmi statistike
- Urejanje in grupiranje podatkov
- Mere osredinjenosti
- Mere razpršenosti
- Grafično prikazovanje podatkov

Osnovni pojmi statistike

Populacija je množica, ki jo statistično proučujemo. Element populacije imenujemo **statistična enota**.

Vzorec je podmnožica populacije, katere elementi predstavljajo največjo možno mero značilnosti celotne množice. Vzorec izberemo, kadar je celotna populacija prevelika množica, da bi analizirali vse njene elemente.

- **Reprezentativen vzorec** je vzorec, ki je izbran tako, da predstavlja značilnosti celotne populacije.
- **Slučajni vzorec** je vzorec, ki je izbran naključno – vsi elementi populacije imajo enako možnost, da bodo izbrani.

Numerus je število elementov vzorca. Oznaka N .

Statistična spremenljivka/podatek/znak je vrednost ali lastnost, ki jo proučujemo.

Vrste statističnih spremenljivk

- **opisne/kvalitativne** statistične spremenljivke
- **vrstne/ordinalne** statistične spremenljivke
- **številске/kvantitativne** statistične spremenljivke

Številске statistične spremenljivke

- **diskretne** številске spremenljivke – zavzamejo lahko posamezne vrednosti
- **zvezne** številске spremenljivke – zavzamejo lahko vsako vrednost z nekega intervala

Naloga

Zapišite, kaj je v danem primeru populacija, vzorec, statistična enota, spremenljivka in ugotovite ali je spremenljivka opisna ali numerična in, če je numerična, ugotovite, ali je zvezna ali diskretna.

Naloga

Zapišite, kaj je v danem primeru populacija, vzorec, statistična enota, spremenljivka in ugotovite ali je spremenljivka opisna ali numerična in, če je numerična, ugotovite, ali je zvezna ali diskretna.

- Na spletni strani je anketa z vprašanjem "Ali imate doma pomivalni stroj?". Nanjo je odgovorilo 254 ljudi.
- V televizijski oddaji gledalci glasujejo za dva kandidata.
- Razrednik svojih 28 dijakov vpraša, kolikšna je oddaljenost njihovega doma do šole.
- Maturant piše seminarsko nalogo z naslovom "Uporaba TikTok-a med srednješolci". Pridobil je odgovore 369 srednješolcev, ki so odgovarjali na vprašanje "Ali uporabljaš TikTok?"
- Znanstveniki pri raziskavi spremljajo, koliko jajc znesejo kokoši na slovenskih farmah na mesec.

Naloga

Slovenija ima več kot 6000 naselij. Statistični urad Republike Slovenije je januarja 2024 naredil analizo naselij glede na število prebivalcev. Rezultati so prikazani v tabeli.

Naloga

Slovenija ima več kot 6000 naselij. Statistični urad Republike Slovenije je januarja 2024 naredil analizo naselij glede na število prebivalcev. Rezultati so prikazani v tabeli.

velikostni razred naselja	število naselij
0	57
1 – 24	719
25 – 49	851
50 – 99	1256
100 – 199	1444
200 – 499	1109
500 – 999	359
1000 – 4999	199
5000 – 9999	23
10000 – 49999	16
50000+	2

Naloga

Slovenija ima več kot 6000 naselij. Statistični urad Republike Slovenije je januarja 2024 naredil analizo naselij glede na število prebivalcev. Rezultati so prikazani v tabeli.

Zapišite, kaj je v danem primeru populacija, statistična enota, spremenljivka in ugotovite ali je spremenljivka opisna ali numerična in, če je numerična, ugotovite ali je zvezna ali diskretna.

velikostni razred naselja	število naselij
0	57
1 – 24	719
25 – 49	851
50 – 99	1256
100 – 199	1444
200 – 499	1109
500 – 999	359
1000 – 4999	199
5000 – 9999	23
10000 – 49999	16
50000+	2

Urejanje in grupiranje podatkov

Podatke, pridobljene v posamezni raziskavi, moramo najprej urediti.

Če podatkov ni veliko, jih uredimo po velikosti v **ranžirno vrsto**, sicer jih združujemo v skupine, **frekvenčne razrede**.

Podatek z največjo vrednostjo označimo z x_{max} , podatek z najnižjo vrednostjo pa x_{min} .

Frekvenca f statističnega znaka je posamezno število diskretnih statističnih enot iste vrednosti.

Frekvenčni razred je skupina podatkov iz vzorca. Frekvenčni razredi so navadno enako široki, in skupaj zajamejo celoten razpon podatkov. Za zvezen nabor podatkov za frekvenčne razrede izberemo intervale (navadno oblike $[a, b)$).

Širina frekvenčnega razreda d_k je razlika med zgornjo (z_k) in spodnjo (s_k) mejo frekvenčnega razreda:

$$d_k = z_k - s_k.$$

Če so razredi enako široki, določimo njihovo širino kot kvocient med celotnim razponom podatkov $x_{max} - x_{min}$ in številom razredov.

Sredina frekvenčnega razreda x_k je aritmetična sredina zgornje in spodnje meje razreda:

$$x_k = \frac{z_k + s_k}{2}.$$

Grupirane podatke predstavimo s **frekvenčno preglednico/porazdelitvijo**.

Za podatke v frekvenčnih preglednicah računamo:

- **(absolutno) frekvenco** f_k – število podatkov z vrednostmi v danem frekvenčnem razredu;
- **relativno frekvenco** f'_k – delež celote, ki ga predstavlja število podatkov v danem frekvenčnem razredu;
- **(absolutno) kumulativno frekvenco** F_k – število podatkov, katerih vrednosti zavzemajo manjšo vrednost od zgornje meje danega frekvenčnega razreda;
- **relativno kumulativno frekvenco** F'_k – delež celote, ki ga predstavlja število podatkov v danem in vseh manjših frekvenčnih razredih.

Naloga

Na šoli analizirajo količino prevzetih obrokov v jedilnici. Rezultati so zbrani v tabeli.

Naloga

Na šoli analizirajo količino prevzetih obrokov v jedilnici. Rezultati so zbrani v tabeli.

Oddelek	Število prevzetih obrokov
1.a	12
1.b	14
1.c	20
2.a	17
2.b	16
2.c	9
3.a	13
3.b	16
3.c	14
4.a	21
4.b	8
4.c	12

Naloga

Na šoli analizirajo količino prevzetih obrokov v jedilnici. Rezultati so zbrani v tabeli. Analizirajte podatke s frekvenčno preglednico. Podatke razdelite v razrede 5 – 9, 10 – 14, 15 – 19, 20 in več.

Oddelek	Število prevzetih obrokov
1.a	12
1.b	14
1.c	20
2.a	17
2.b	16
2.c	9
3.a	13
3.b	16
3.c	14
4.a	21
4.b	8
4.c	12

Naloga

Dijaki 3. a oddelka so zapisovali svoje pribljubljene barve.

Zapisali so jih: modra, rdeča, rdeča, zelena, rumena, rdeča, modra, zelena, modra, modra, rumena, rdeča, zelena, modra, rumena, rumena, zelena, rdeča.

Analizirajte rezultate s frekvenčno preglednico.

Naloga

Dijaki 3. a oddelka so zapisovali svoje pribljubljene barve.

Zapisali so jih: modra, rdeča, rdeča, zelena, rumena, rdeča, modra, zelena, modra, modra, rumena, rdeča, zelena, modra, rumena, rumena, zelena, rdeča.

Analizirajte rezultate s frekvenčno preglednico.

Naloga

Lokostrelec si beleži rezultate treningov.

Vrednosti so bile: 10.3, 10.4, 9.9, 9.7, 10.2, 8.9, 9.4, 10.1, 9.0, 10.3, 9.5, 10.6.

Analizirajte rezultate s frekvenčno preglednico.

Naloga

V frekvenčni preglednici so zbrani podatki o številu sorojencev dijakov 2. b oddelka. Dopolnite preglednico.

Število sorojencev	f_k	f'_k	F_k	F'_k
0	5			
1		25 %		
2				
3		10 %		
skupaj	20	100 %	/	/

Mere osredinjenosti

Aritmetična sredina

Aritmetična sredina ali **povprečje** je količnik vsote vseh vrednosti statistične spremenljivke in števila teh vrednosti.

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

Če se vrednosti statistične spremenljivke ponavljajo (k_i vrednosti x_i), je formula sledeča:

$$\bar{x} = \frac{k_1 x_1 + k_2 x_2 + \cdots + k_m x_m}{k_1 + k_2 + \cdots + k_m} = \frac{\sum_{i=1}^m k_i x_i}{\sum_{i=1}^m k_i}; \quad \sum_{i=1}^m k_i = N$$

Pri grupiranih podatkih za vrednosti vzamemo sredine frekvenčnih razredov.

Modus

Modus ali **gostiščnica** M_o je vrednost statistične spremenljivke, ki se v množici vseh vrednosti najpogosteje ponavlja.

Če se v neki množici dve vrednosti pojavita enako mnogokrat najpogosteje, rečemo, da je porazdelitev vrednosti **bimodalna**.

Za grupirane podatke določamo **modalni razred**, to je tisti razred, ki ima največjo frekvenčno gostoto.

Mediana

Mediana ali **središčnica** Me je tista vrednost statistične spremenljivke, pri kateri je polovica vrednosti večjih ali enakih, druga polovica vrednosti pa manjših ali enakih od te vrednosti.

Če imamo liho število vrednosti statistične spremenljivke, za mediano vzamemo vrednost, ki stoji na mestu $\frac{n+1}{2}$ po velikosti urejenih podatkov.

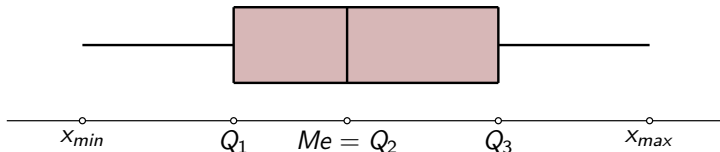
Če je število vrednosti sodo, za vrednost mediane vzamemo aritmetično sredino srednjih dveh podatkov.

Mediana razdeli podatke na dve polovici. Ti dve polovici lahko spet razdelimo na dve polovici in dobimo štiri enako močne množice podatkov. Meje teh skupin imenujemo **kvartili**.

Kvartili

Prvi kvartil Q_1 je mediana prve (spodnje) polovice podatkov, **drugi kvartil** Q_2 je mediana Me vseh podatkov, **tretji kvartil** Q_3 pa je mediana druge (zgornje) polovice podatkov.

Vrednosti kvartilov, minimalno vrednost in maksimalno vrednost množice podatkov grafično predstavimo z **diagramom kvartilov** oziroma **šktalo z brki**.



Naloga

Izračunajte aritmetično sredino količin.

- 1.5 s, 3.5 s, 1 s
- 4 km, 2000 m, 3 km
- 4 £, 2 £, 3 £, 1 £, 5 £

Naloga

Izračunajte aritmetično sredino količin.

- 1.5 s, 3.5 s, 1 s
- 4 km, 2000 m, 3 km
- 4 £, 2 £, 3 £, 1 £, 5 £

Naloga

Izračunajte aritmetično sredino danim podatkom.

- 2, 3, 1, 8, 19, 2, 7
- 13, 39, 12
- 0.3, 0.4, 0.5, 0.7, 0.6

Naloga

Določite modus danim številskim podatkom.

- 1, 4, 2, 4, 1, 6, 3, 4, 1, 4, 6, 4, 4, 8

- 3, 25, 10, 3, 5, 7, 5, 7, 9, 4, 49

- $\frac{1}{3}, \frac{3}{4}, \frac{1}{2}, \frac{6}{8}, \frac{2}{9}$

- $\frac{1}{2}, \frac{2}{4}, \frac{1}{4}, \frac{5}{10}, \frac{8}{9}$

Naloga

Določite modus danim številskim podatkom.

- 1, 4, 2, 4, 1, 6, 3, 4, 1, 4, 6, 4, 4, 8
- 3, 25, 10, 3, 5, 7, 5, 7, 9, 4, 49
- $\frac{1}{3}, \frac{3}{4}, \frac{1}{2}, \frac{6}{8}, \frac{2}{9}$
- $\frac{1}{2}, \frac{2}{4}, \frac{1}{4}, \frac{5}{10}, \frac{8}{9}$

Naloga

V porodnišnici so izmerili dolžine dojenčkov, ki so se rodili v enem dnevu.

50, 51, 51, 44, 47, 48, 53, 49, 52, 55, 46, 50, 50, 49, 47, 47

Določite mediano podatkov.

Naloga

Otroci v vrtcu so metali žogo na koš in si zapisovali dosežke. Podatki so prikazani v preglednici.

Otrok	Jaka	Jure	Miha	Polona	Valerija	Tina	Mojca	Cene	Darja
Št. košev	5	7	10	8	5	6	9	9	4

Izračunajte, koliko košev je otrok zadel v povprečju. Podatke uredite po vrsti in določite Mo , Me ter narišite škatlo z brki.

Mere razpršenosti

Informacijo o **porazdelitvi** oziroma **razpršenosti** podatkov lahko izračunamo s pomočjo: variacijskega razmika, interkvartilnega ranga, variance in standardnega odklona.

Variacijski razmik

Variacijski razmik R je razlika med maksimalno in minimalno vrednostjo statistične spremenljivke:

$$R = x_{\max} - x_{\min}.$$

Variacijski razmik je zelo odvisen od ekstremnih vrednosti, posebno osamelcev, zato ga uporabljamo le v kombinaciji z drugimi merami razpršenosti.

Interkvartilni rang

Interkvartilni rang oziroma **medčetrtnski razmik** IR je razlika med vrednostjo prvega in tretjega kvartila:

$$IR = Q_3 - Q_1.$$

Osamelec je podatek, katerega vrednost je za več kot 3-kratnik interkvartilnega ranga IR nad tretjim kvartilom Q_3 ali pod prvim kvartilom Q_1 .

Podatek je “pogojno osamelec”, če je njegova vrednost za več kot 1.5-kratnik interkvartilnega ranga IR nad tretjim kvartilom Q_3 ali pod prvim kvartilom Q_1 .

Interkvartilni rang je mera razpršenosti, ki ni občutljiva na osamelce.

Varianca

Varianca σ^2 predstavlja aritmetično sredino kvadratov odklikov vrednosti statistične spremenljivke od aritmetične sredine:

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{N} = \frac{1}{N} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Standardni odklon

Standardni odklon σ izračunamo kot koren variance:

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{N}} = \sqrt{\frac{1}{N} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Predstavlja povprečje odklikov vrednosti statistične spremenljivke od aritmetične sredine.

Naloga

V preglednici so predstavljene cene treh izdelkov v trgovini po posameznih mesecih leta 2019.

Izračunajte povprečno ceno in standardni odklon cene vsakega izdelka.

Izdelek	Jan	Feb	Mar	Apr	Maj	Jun	Jul	Avg	Sep	Okt	Nov
Kruh	3.35	3.29	3.34	3.38	3.38	3.37	3.38	3.55	3.53	3.54	3.49
Jagode	8.73	7.18	5.52	4.48	5.72	5.64	6.49	6.58	7.15	7.58	8.34
Cvetača	2.04	2.17	1.58	1.75	2.13	1.85	1.93	1.87	1.81	1.99	1.80

Naloga

V preglednici je prikazano število rojstev v Sloveniji po letih. Izračunajte povprečno število rojstev in standardni odklon.

Leto	2013	2014	2015	2016	2017	2018	2019	2020	2021
Število	21111	21165	20641	20345	20241	19585	19328	18767	18989

Naloga

V preglednici je prikazano število rojstev v Sloveniji po letih. Izračunajte povprečno število rojstev in standardni odklon.

Leto	2013	2014	2015	2016	2017	2018	2019	2020	2021
Število	21111	21165	20641	20345	20241	19585	19328	18767	18989

Naloga

Pridobili smo podatke (urejene po velikosti): 1, 13, 14, 15, 15, 15, 17, 18, 18, 19, 19, 19, 19, 20 in 40.

- Opišite razpršenost podatkov R , IR , Q_1 , Q_3 , σ , \bar{x} .
- Največjo in najmanjšo vrednost (v tem primeru sta to osamelca) odstranimo. Kako se spremeni razpršenost podatkov?

Grafično prikazovanje podatkov

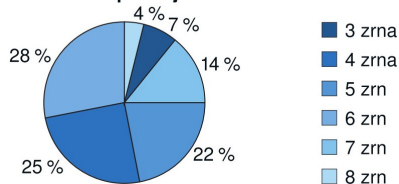
Strukturni krog

Strukturni krog ali krožni diagram

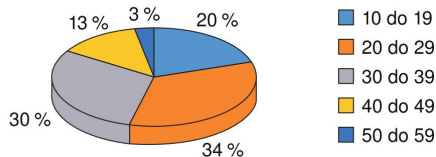
uporabljamo, kadar so podatki razvrščeni v malo frekvenčnih razredov ali ne dosežejo veliko različnih diskretnih vrednosti.

Celoto predstavlja 360° , za ostale deleže središčne kote izračunamo s sklepnim računom.

Grah s prve njive



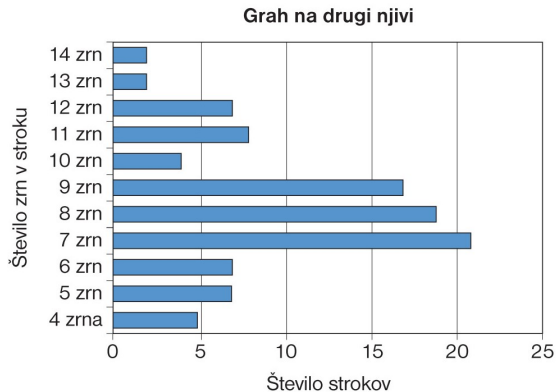
Frekvenčni razredi avtomobilov mimo vrtca

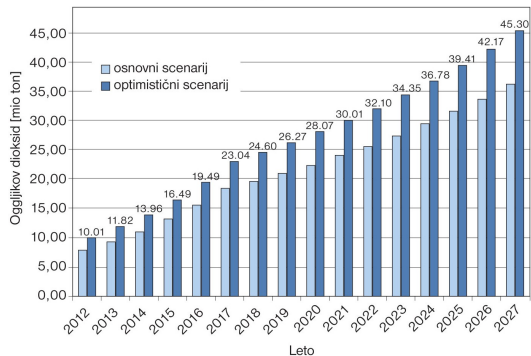


Stolpčni diagram

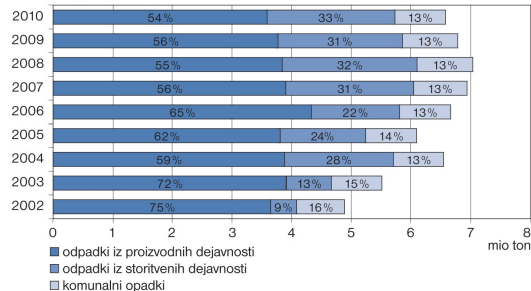
Stolpčni diagram uporabljamo, ko so podatki razvrščeni v veliko frekvenčnih razredov ali lahko dosežejo veliko diskretnih vrednosti.

Stolpčni diagrami so lahko **pokončni** ali **ležeči**. Če želimo prikazati več podatkov naenkrat, uporabimo **sestavljeni** ali **strukturni** stolpčni diagram.

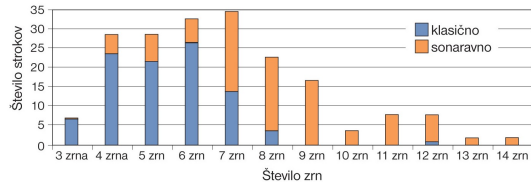


Napoved zmanjšanja izpustov CO₂ po osnovnem in optimističnem scenariju

Struktura komunalnih in nekomunalnih odpadkov v Sloveniji od 2002 do 2009 (Vir: SURS)



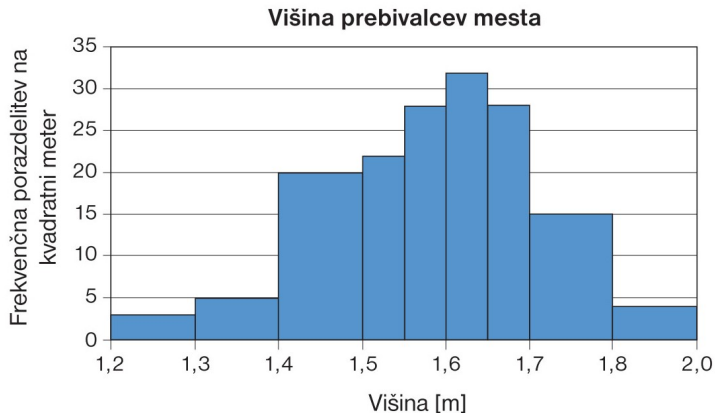
Primerjava rodnosti klasično in sonaravno gojenega graha



Histogram

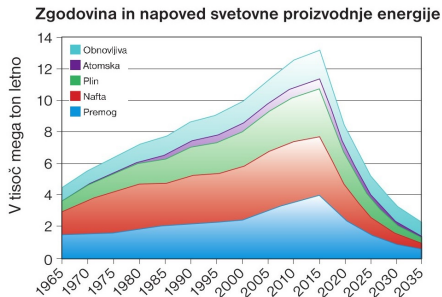
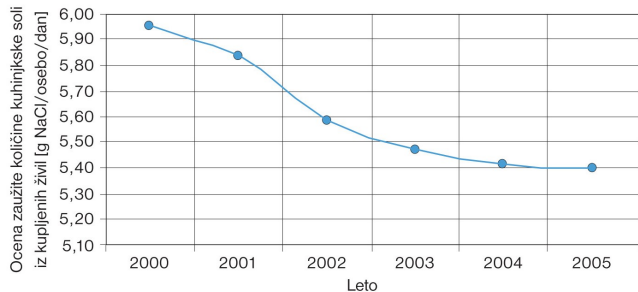
Histogram uporabljamo za prikaz grupiranih podatkov.

Širine frekvenčnih razredov niso nujno enake. Meje razredov narišemo na vodoravni osi, frekvence posameznih razredov pa na navpični osi.



Linijski diagram

Linijski diagram/poligon uporabljamo, ko želimo prikazati postopno spreminjanje vrednosti nekega podatka skozi daljše časovno obdobje. Frekvenčne porazdelitve ponazorimo s **frekvenčnim poligonom**, podatki so lahko zvezni ali grupirani.



Naloga

Na matematičnem testu je bilo mogoče doseči 50 točk. Dosežki so bili:
35, 22, 41, 47, 36, 30, 27, 19, 31, 43, 48, 44, 23, 26, 36, 10, 33, 14, 9.

Razdelite jih v pet enako velikih razredov ter predstavite s histogramom.

Naloga

Na matematičnem testu je bilo mogoče doseči 50 točk. Dosežki so bili: 35, 22, 41, 47, 36, 30, 27, 19, 31, 43, 48, 44, 23, 26, 36, 10, 33, 14, 9.

Razdelite jih v pet enako velikih razredov ter predstavite s histogramom.

Naloga

Otroci v vrtcu so metali žogo na koš in si zapisovali dosežke. Podatki so prikazani v preglednici.

Izračunajte, koliko košev je otrok zadel v povprečju. Podatke uredite po vrsti in določite Mo , Me ter narišite škatlo z brki.

Otrok	Jaka	Jure	Miha	Polona	Valerija	Tina	Mojca	Cene	Darja
Št. košev	5	7	10	8	5	6	9	9	4

Naloga

Bojana beleži, koliko časa potrebuje za pot do šole. Podatke je zapisala v preglednico. S stolpčnim diagramom predstavite, kako pogosto v šolo potuje 8 minut, 9 minut ...

Dan	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
Čas [min]	9	11	10	8	11	10	9	12	9	11

Naloga

Bojana beleži, koliko časa potrebuje za pot do šole. Podatke je zapisala v preglednico. S stolpčnim diagramom predstavite, kako pogosto v šolo potuje 8 minut, 9 minut ...

Dan	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
Čas [min]	9	11	10	8	11	10	9	12	9	11

Naloga

V domu ostarelih občanov je 500 oskrbovancev. Od 50 do 60 let jih je 15 %, med 60 in 70 leti je 160 oskrbovancev, med 70 in 80 leti pa 200 starostnikov. Drugi so stari med 80 in 90 let.

- Iz grupiranih podatkov izračunajte povprečno starost oskrbovancev tega doma.
- Grafično ponazorite starost oskrbovancev.