

# Grupirani podatci,

## I. dio



Sanja Varošaneć, Zagreb

U prethodnim smo člancima razmatrali tabelarno i grafičko prikazivanje primarnih podataka te smo upoznali i neke numeričke pokazatelje vezane uz dani skup podataka kao što su medijan, kvantili, aritmetička sredina itd. Ovdje ćemo proučiti situaciju kad su frekvencije danih podataka male te je takve podatke pogodno **grupirati** u grupe, tzv. razrede. Pokažimo na jednom primjeru grupiranje podataka i njihovo grafičko prikazivanje.

**Primjer 1.** Učenicima 1.a razreda izmjerena je visina i dobiveni su ovi podatci (iskazani u centimetrima):

182 153 164 173 184 175 180 155  
201 177 180 183 186 188 182 178  
169 168 173 159 152 162 163 185

Visina učenika je tzv. **numeričko kontinuirano obilježje**. Prikažimo ove podatke u tablici. Budući da je frekvencija svake vrijednosti mala (1 ili 2), podatke ćemo grupirati. Te se grupe nazivaju **razredi**. Obično se stvara od 6 do 20 razreda jednake širine.

Popis razreda nalazi se u prvom stupcu, a frekvencija, tj. broj učenika čija visina se nalazi u pojedinom razredu, napisane su u drugom stupcu.

Prvo ćemo napraviti grupiranje u 6 razreda jednake širine. Najmanja visina je 152, a najveća 201, dakle, imamo

$$201 - 152 + 1 = 50$$

mogućih vrijednosti od 152 do 201. Taj broj podijelimo s brojem razreda i dobiveni broj zaokružimo na više. U ovom slučaju imamo račun

$$50 : 6 = 8.33,$$

i broj 8.33 zaokružimo na 9. Dakle, **širina razreda** je 9. To znači da se u prvom razredu nalaze vrijednosti: 152, 153, 154, 155, 156, 157, 158, 159 i 160. U drugom razredu se nalaze vrijednosti 161, 162,

163, 164, 165, 166, 167, 169. I tako dalje za svaki razred.

visina $x$ učenika (u cm)	frekvencija
$152 \leq x \leq 160$	4
$161 \leq x \leq 169$	5
$170 \leq x \leq 178$	5
$179 \leq x \leq 187$	8
$188 \leq x \leq 196$	1
$197 \leq x \leq 205$	1

U ovakvom slučaju, kad se radi o numeričkim podacima, korisno je znati i tzv. precizne granice razreda. Naime, razredi se moraju nastavljati jedan na drugoga. Zato se kao precizne granice prvoga razreda stavljaju 151.5 kao donja granica i 160.5 kao gornja. Precizne granice drugoga razreda su 160.5 i 169.5 itd. Ovo uvođenje preciznih granica razreda ima i svoje fizikalno objašnjenje. Naime, kod mjerenja smo svaki rezultat koji je bio veći od 160.5 registrirali kao 161, a svaki rezultat koji je bio manji od 160.5 smo registrirali kao 160. Unesimo sada u gornju tablicu i precizne granice razreda. Uz to ćemo i u prvom stupcu umjesto  $152 \leq x \leq 160$  skraćeno pisati  $152 - 160$ .

visina učenika (u cm)	precizne granice razreda	frekvencija
152 – 160	151.5 – 160.5	4
161 – 169	160.5 – 169.5	5
170 – 178	169.5 – 178.5	5
179 – 187	178.5 – 187.5	8
188 – 196	187.5 – 196.5	1
197 – 205	196.5 – 205.5	1

Kao što vidimo, kad podatke grupiramo, gubimo dio informacija. U tom slučaju, ako nemamo primarne podatke, nego samo tablicu frekvencija i gledajući prvi redak možemo samo zaključiti da se u intervalu  $\langle 151.5, 160.5 \rangle$  nalaze 4 podatka, ali ne znamo koji su to podatci.

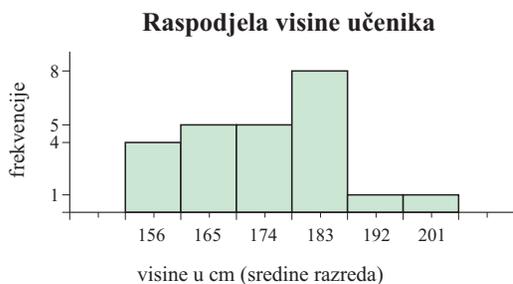
Pri grupiranju podataka postoji nekoliko momenata u kojima donosimo odluke koje nisu precizno determinirane. Prvo pitanje s kojim se suočava istraživač je grupirati podatke u razrede jednake širine (veličine) ili različite. To u mnogome ovisi o danom kontekstu istraživanja, ali i o budućoj upotrebi rezultata. Na kojem mjestu počinje prvi razred? U prethodnom smo primjeru grupiranja odabrali upravo najmanji podatak za donju granicu prvog razreda. Ali, mogli smo, primjerice, uzeti za prvi razred  $151 - 159$  ili  $150 - 158$ . Pri odluci o granicama treba paziti da se razredi ne preklapaju, da se nadovezuju te da se svi podatci nalaze u točno jednom razredu. I obično se prvi razred bira tako da ima bar jedan element. Broj razreda je također odluka istraživača. Uobičajeno je rabiti 6–20 razreda iako postoje situacije kad se i pri tome rade iznimke.

U ovom smo primjeru precizne granice razreda odredili tako da smo izračunali aritmetičku sredinu najveće granice donjeg razreda i najmanje granice gornjeg razreda. Tako je precizna granica između razreda  $170 - 178$  i  $179 - 187$  broj  $\frac{178 + 179}{2} = 178.5$ . No, u literaturi ćemo naići i na ovakve razrede:  $152 - 161$ ,  $161 - 170$ ,  $170 - 179 \dots$  U tom slučaju mora biti jasno naznačeno kojem razredu pripada rubni podatak. Primjerice, pripada li podatak 170 razredu  $161 - 170$  ili  $170 - 179$ ? U takvom slučaju obično se popratnom rečenicom opiše izgled razreda. Primjerice, navede se da je razred oblika  $[161, 170)$  ili  $\langle 161, 170]$ , tj. opiše se pripada li ili ne desni rub razreda tom razredu. Da bismo izbjegli ovakve dvosmislenosti, koristit ćemo precizne granice razreda koje jasno definiraju pripadaju li dani podatci u odgovarajući razred ili ne.

## Grafički prikaz grupiranih podataka

Postoji nekoliko načina grafičkog prikazivanja grupiranih podataka. Ovdje ćemo ih prikazati s pomoću stupčastih dijagrama, histograma i poligona frekvencija.

Na donjoj slici je **stupčasti dijagram frekvencija** podataka iz prethodnog primjera 1.



Slika 1.

Stupci su pravokutnici čija se jedna stranica proteže od donje precizne granice razreda do gornje precizne granice razreda, a visina stupca je frekvencija tog razreda. Dakle, u gornjem grafu, prvi se pravokutnik crta nad intervalom [151.5, 160.5] i ima visinu 4, drugi nad [160.5, 169.5] i ima visinu 5 itd. Kako se na apscisnoj osi ne bi pojavila velika količina brojaka, tj. kako bi graf bio što čitljiviji, na apscisnoj se osi ne označavaju precizne granice razreda nego sredine razreda.

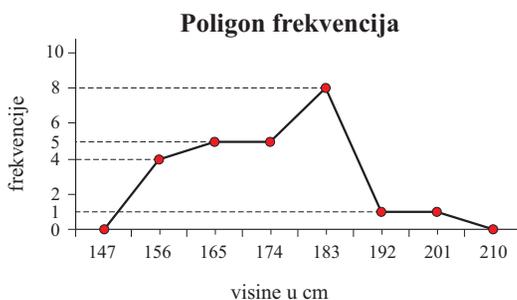
**Sredina razreda** se računa tako da se izračuna aritmetička sredina donje i gornje precizne granice razreda. Za prvi razred sredina je

$$\frac{151.5 + 160.5}{2} = 156.$$

Dopunimo tablicu frekvencija stupcem sa sredinama razreda.

visina učenika (u cm)	precizne granice razreda	frekvencija	sredina razreda
152 – 160	151.5 – 160.5	4	156
161 – 169	160.5 – 169.5	5	165
170 – 178	169.5 – 178.5	5	174
179 – 187	178.5 – 187.5	8	183
188 – 196	187.5 – 196.5	1	192
197 – 205	196.5 – 205.5	1	201

Nacrtajmo sada **poligon frekvencija** koji prikazuje ove podatke.



Slika 2.

Čvorovi poligonalne linije su točke oblika (sredina razreda, frekvencija razreda) koje se spajaju dužinama. Tradicionalno, poligonalna linija počinje i završava na  $x$ -osi, pa se osim čvorova koji proizlaze iz postojećih razreda dodaju još dvije točke, obje na  $x$ -osi. Prvoj je apscisa jednaka sredini prvog razreda umanjenoj za širinu, a drugoj je apscisa jednaka sredini posljednjeg razreda uvećanoj za širinu razreda. U našem primjeru to su točke (147, 0) i (210, 0). Dakle, kao da smo popis razreda dopunili "nultim" razredom i još jednim razredom nakon posljednjeg.

Osim frekvencija na  $y$ -os se mogu nanositi relativne frekvencije, čime dobivamo stupčasti dijagram relativnih frekvencija i poligon relativnih frekvencija. Prikaz s pomoću poligona frekvencija je izrazito pogodan kad na jednom grafu želimo prikazati nekoliko nizova podataka.

Iz stupčastog dijagrama lako se prelazi u novi grafički prikaz: **histogram**. Histogram se također sastoji od (spojenih) pravokutnika koji se crtaju nad intervalima određenim preciznim granicama razreda, ali visina stupca jednaka je **gustoći** frekvencije, tj. **korigiranoj relativnoj frekvenciji**. Gustoća frekvencije je kvocijent relativne frekvencije i širine razreda.

U našem primjeru, gustoća frekvencije za prvi razred je

$$f_{c1} = \frac{\frac{4}{24}}{9} = \frac{4}{216}.$$

Dopunimo tablicu našeg prvog primjera dvama stupcima – stupcem relativnih frekvencija i stupcem gustoće frekvencija.

visina učenika (u cm)	precizne granice razreda	frekv. $f_i$	sredina razreda	relativna frekvencija $r_i$	gustoća $f_{c1}$
152 – 160	151.5 – 160.5	4	156	$\frac{4}{24}$	$\frac{4}{216}$
161 – 169	160.5 – 169.5	5	165	$\frac{5}{24}$	$\frac{5}{216}$
170 – 178	169.5 – 178.5	5	174	$\frac{5}{24}$	$\frac{5}{216}$
179 – 187	178.5 – 187.5	8	183	$\frac{8}{24}$	$\frac{8}{216}$
188 – 196	187.5 – 196.5	1	192	$\frac{1}{24}$	$\frac{1}{216}$
197 – 205	196.5 – 205.5	1	201	$\frac{1}{24}$	$\frac{1}{216}$

Kad su podatci grupirani u razrede različitih širina, tada stupčasti dijagram i histogram nemaju isti oblik.

**Primjer 2.** Dani su podatci o stanovništvu prema navršenim godinama starosti. Nacrtajmo histogram.

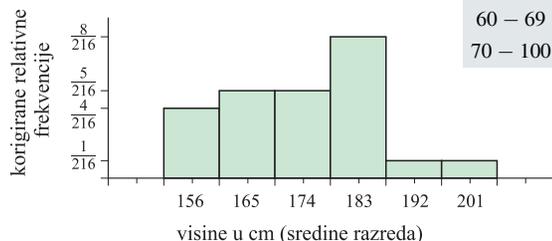
Navršene godine starosti	Broj stanovnika u tisućama
0 – 14	755
15 – 24	604
25 – 39	907
40 – 49	667
50 – 59	530
60 – 69	515
70 – 100	486

Odredit ćemo precizne granice razreda, sredine i gustoće frekvencija. Zbroj svih frekvencija je

$$755 + 604 + 907 + 667 + 530 + 515 + 486 = 4464.$$

Podatci u posljednja dva stupca namjerno nisu skraćeni kako bi se uočila sličnost tih stupaca s trećim stupcem. Zbog te sličnosti histogram podataka grupiranih u razrede jednakih širina ima isti oblik kao i stupčasti dijagram. Histogram je dan na donjoj slici:

### Raspodjela visine učenika

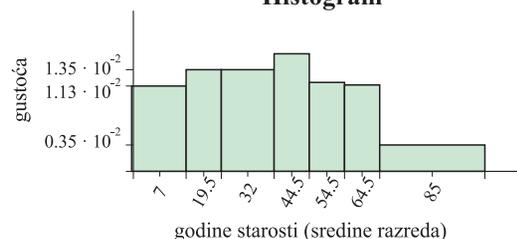


Slika 3.

Karakteristika histograma je da je zbroj površina svih stupaca jednak 1. Brzim računom lako provjeravamo da i gornji prikaz ima to svojstvo.

Navršene godine starosti	Broj stanovnika u tisućama	Precizne granice razreda	Sredine razreda	Širine razreda	Gustoća frekvencije
0 – 14	755	-0.5 – 14.5	7	15	$\frac{755}{4464 \cdot 15} = 1.13 \cdot 10^{-2}$
15 – 24	604	14.5 – 24.5	19.5	10	$1.35 \cdot 10^{-2}$
25 – 39	907	24.5 – 39.5	32	15	$1.35 \cdot 10^{-2}$
40 – 49	667	39.5 – 49.5	44.5	10	$1.49 \cdot 10^{-2}$
50 – 59	530	49.5 – 59.5	54.5	10	$1.19 \cdot 10^{-2}$
60 – 69	515	59.5 – 69.5	64.5	10	$1.15 \cdot 10^{-2}$
70 – 100	486	69.5 – 100.5	85	31	$0.35 \cdot 10^{-2}$

### Histogram



Slika 4.

(Nastavak u idućem broju.)