Grupirani podatci,
II. dio

Sanja Varošanec, Zagreb

U prošlom smo članku opisali kako grupirati dane podatke te ih grafički prikazati. U ovom ćemo broju opisati kako grupiranim podatcima odrediti neke numeričke pokazatelje kao što su aritmetička sredina, medijan i oba kvartila.

Primjer 1. U tvornici žarulja ispituje se vrijeme trajanja proizvedenih žarulja. Ispitano je $10\,000$ žarulja i dobiveni su podatci, tj. vremena ispravnog rada žarulja iskazana u satima, grupirani u 9 razreda jednake širine. Frekvencije su dane u ovoj tablici.

broj sati trajanja žarulje	broj žarulja (f_i)
$0 - 296$	291
$297 - 593$	637
$594 - 890$	1256
$891 - 1187$	2020
$1188 - 1484$	2425
$1485 - 1781$	2210
$1782 - 2078$	743
$2079 - 2375$	418

a) Izračunajmo prosječno vrijeme trajanja žarulja.

b) Nacrtajmo graf kumulativnih relativnih frekvencija i odredimo medijan, donji i gornji kvartil.

Rješenje. a) Kako bismo tražili prosječno vrijeme trajanja žarulja da raspolažemo primarnim podatcima? Izračunali bismo aritmetičku sredinu, tj. zbrojili bismo sva vremena trajanja žarulja i podijelili s $N = 10\,000$. Ali u ovom slučaju primarni su podatci izgubljeni, imamo samo grupirane podatke. Tako, primjerice, znamo da je 1256 žarulja imalo vrijeme trajanja od 594 sata do 890 sati, ali ne znamo koliko je sati točno trajala svaka od tih 1256 žarulja. Stoga smatramo da su te žarulje trajale u prosjeku $\dfrac{594+890}{2} = 742$ sata. To je sredina trećeg razreda, tj. svaka će od tih 1256 žarulja iz trećeg razreda u izračunu aritmetičke sredine biti uračunata sa 742 sata trajanja.

Drugim riječima, prosječno vrijeme trajanja svih žarulja je težinska aritmetička sredina svih sredina razreda, tj.

$$\overline{x} = \frac{f_1\overline{x_1} + \cdots + f_n\overline{x_n}}{N}$$

gdje su $\overline{x_1}, \ldots, \overline{x_n}$ sredine razreda, $f_1, \ldots, f_n$ frekvencije tih razreda, a N je ukupan broj podataka, tj. $N = f_1 + f_2 + \cdots + f_n$.

Dakle, dopunimo tablicu sa stupcem sredina razreda i izračunajmo aritmetičku sredinu.

i	broj sati trajanja žarulje	broj žarulja (f_i)	sredine razreda $(\overline{x_i})$
1	$0-296$	291	148
2	$297-593$	637	445
3	$594-890$	1256	742
4	$891-1187$	2020	1039
5	$1188-1484$	2425	1336
6	$1485-1781$	2210	1633
7	$1782-2078$	743	1930
8	$2079-2375$	418	2227

$$\begin{aligned}\overline{x} &= \Big(291 \cdot 148 + 637 \cdot 445 + 1256 \cdot 742 + \cdots \\ &\qquad + 743 \cdot 1930 + 418 \cdot 2227\Big) : 10\,000 \\ &= \frac{12\,570\,871}{10\,000} = 1257.1\end{aligned}$$

Prosječno vrijeme trajanja žarulja je 1257.1 h.

Komentar. U prethodnom računu pojavilo se množenje troznamenkastih i četveroznamenkastih, odnosno četveroznamenkastih i četveroznamenkastih brojeva što ne predstavlja problem ako se račun u cijelosti radi na računalu. Međutim, ako se račun djelomično radi ručno, tada postoji velika vjerojatnost pojavljivanja greške u pisanju i prepisivanju velikih brojeva. Stoga su u prošlosti, kad računala nisu bila široko rasprostranjena, matematičari razvili metodu kako račun za aritmetičku sredinu pojednostavniti.

Budući da su širine razreda jednake veličine $(\check{s} = 297)$, sredine razreda čine aritmetički niz. Za prvi član niza odaberemo po volji jednu od sredina razreda. Obično se uzima neka srednja sredina, primjerice $\overline{x_0} = 1336$. Uvedemo novu varijablu d_i ovako

$$d_i = \frac{\overline{x_i} - \overline{x_0}}{\check{s}}.$$

Tada je u petom razredu $d_5 = 0$, u sljedećem, 6. razredu $d_6 = 1$, u prethodnom, 4. razredu je $d_4 = -1$ itd. Laganim se računom pokazuje da je aritmetička sredina uz ovu novu varijablu jednaka

$$\overline{x} = \frac{f_1 d_1 + f_2 d_2 + \cdots + f_n d_n}{N} \cdot \check{s} + \overline{x_0}.$$

Stoga se tablica nadopunjava sa stupcima d_i i $f_i d_i$.

i	broj sati trajanja žarulje	broj žarulja (f_i)	sredine razreda $(\overline{x_i})$	d_i	$f_i d_i$
1	$0-296$	291	148	-4	-1164
2	$297-593$	637	445	-3	-1911
3	$594-890$	1256	742	-2	-2512
4	$891-1187$	2020	1039	-1	-2020
5	$1188-1484$	2425	1336	0	0
6	$1485-1781$	2210	1633	1	2210
7	$1782-2078$	743	1930	2	1486
8	$2079-2375$	418	2227	3	1254

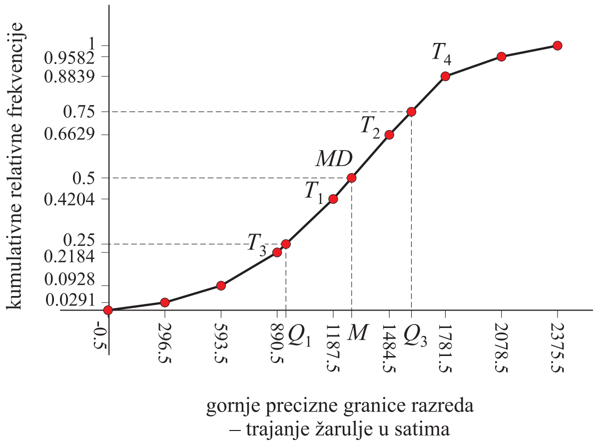
Zbroje se sve vrijednosti u stupcu $f_i d_i$, podijele s $N = 10\,000$, pomnože sa širinom razreda 297 i doda se vrijednost $\overline{x_0} = 1336$ te se opet dobiva da je $\overline{x} = 1257.1$. Sada su se u računu pojavili nešto manji brojevi te se račun brzo može izvesti i bez uporabe računala.

b) Medijan i kvartile odredit ćemo koristeći graf kumulativnih relativnih frekvencija. Kumulativna frekvencija razreda je zbroj svih frekvencija prethodnih razreda uključujući i promatrani razred. Analogno, kumulativna relativna frekvencija razreda je zbroj svih relativnih frekvencija prethodnih razreda i njega samog. Ustvari, kumulativna frekvencija nekog podatka nam kazuje koliko je podataka manje ili jednako promatranom podatku. Jasno je zašto su kumulativne relacije frekvencije pogodne za izračunavanje medijana. Prisjetimo se da je medijan podatak za koji je $50\ \%$ svih podataka manje ili jednako njemu. Koristeći terminologiju kumulativnih frekvencija, možemo reći da je medijan podatak čija je kumulativna relativna frekvencija 0.5.

Upotpunimo tablicu dvama stupcima: stupcem preciznih granica razreda i stupcem kumulativnih relativnih frekvencija. Primjerice, kumulativna relativna frekvencija trećeg razreda jednaka je zbroju relativnih frekvencija prvog, drugog i trećeg razreda, tj. jednaka je

$$\frac{291+637+1256}{10\,000}=\frac{2184}{10\,000}=0.2184.$$

i	broj sati trajanja žarulje	broj žarulja (f_i)	sredine razreda $(\overline{x_i})$	precizne granice razreda	kumulativne rel. frekvencije
1	$0-296$	291	148	$-0.5-296.5$	0.0291
2	$297-593$	637	445	$296.5-593.5$	0.0928
3	$594-890$	1256	742	$593.5-890.5$	0.2184
4	$891-1187$	2020	1039	$890.5-1187.5$	0.4204
5	$1188-1484$	2425	1336	$1187.5-1484.5$	0.6629
6	$1485-1781$	2210	1633	$1484.5-1781.5$	0.8839
7	$1782-2078$	743	1930	$1781.5-2078.5$	0.9582
8	$2079-2375$	418	2227	$2078.5-2375.5$	1



Slika 1.

Graf je poligonalna linija koja spaja točke (čvorove) oblika

(gornja precizna granica razreda,

kumulativna relativna frekvencija razreda).

Osim toga, prva se točka gore opisanog oblika dodatno spaja s točkom na x-osi čija je apscisa donja precizna granica prvog razreda. Dakle, čvorovi u ovom primjeru su $(-0.5,0)$, $(296.5,0.0291)$, $(593.5,0.0928)\dots(2375.5,1)$. Zamijetimo da je ovo graf jedne rastuće funkcije koja poprima vrijednosti od 0 do 1.

Medijan M je onaj podatak kojemu je kumulativna relativna frekvencija jednaka 0.5. Dakle, tražimo apscisu točke $MD(M,0.5)$. Nacrtamo li tu točku MD na grafu kumulativnih relativnih frekvencija, uočavamo da se nalazi na dužini koja spaja točke $T_1(1187.5,0.4204)$ i $T_2(1484.5,0.6629)$, tj. medijan će se nalaziti u 5. razredu. Taj se razred stoga naziva **medijalnim razredom**. Izračunajmo jednadžbu pravca kroz točke T_1 i T_2.

$$y-y_1=\frac{y_2-y_1}{x_2-x_1}(x-x_1)$$

$$y-0.4204=\frac{0.6629-0.4204}{1484.5-1187.5}(x-1187.5)$$

$$y=8.165\cdot10^{-4}x-0.54919.$$

Za $y=0.5$ dobivamo da je odgovarajuća apscisa $M=1285$ i to je medijan danih podataka.

Donji i gornji kvartil određuju se analogno. Donji je kvartil točka čija je ordinata 0.25, jer kao što smo rekli u članku [2], donji kvartil je podatak za koji je $25\ \%$ podataka manje ili jednako njemu. Ta se točka nalazi na grafu kumulativnih relativnih frekvencija između točaka $T_3(890.5,0.2184)$ i $T_1(1187.5,0.4204)$. Opet odredimo jednadžbu pravca kroz te točke

$$T_3T_1\dots\ y=6.8\cdot10^{-4}x-0.38726,$$

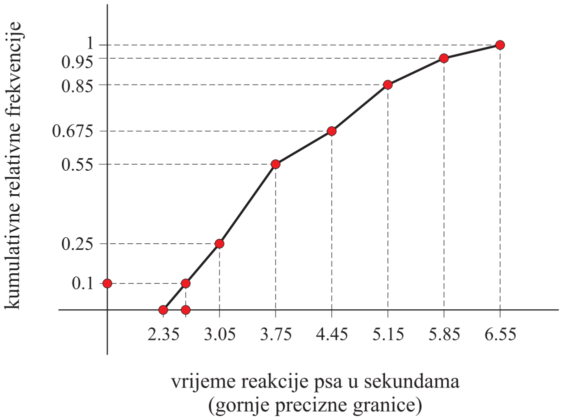
a točka čija je ordinata jednaka 0.25 ima apscisu $Q_1=937.1$ i to je donji kvartil.

Gornji kvartil je točka grafa čija je ordinata jednaka 0.75 i nalazi se na dužini koja spaja točke $T_2(1484.5,0.6629)$ i $T_4(1781.5,0.8839)$. Jednadžba pravca T_2T_4 glasi

$$y=7.44\cdot10^{-4}x-0.441728,$$

a za $y=0.75$ dobivamo $Q_3=1601.8$ i to je gornji kvartil.

Primjer 2. Prilikom istraživanja brzine reakcije 80 vojnih pasa na određeni podražaj dobiveni su podatci o vremenu reakcije izraženi u sekundama i desetinkama sekunda. Podatci su grupirani u 6 razreda jednakih širina i prikazani su u ovom grafu kumulativnih relativnih frekvencija. Odredimo granice razreda, frekvencije razreda, te formirajmo tablicu frekvencija. Potom odredimo prosječno vrijeme reakcije te koliki je interval vremena reakcije $10\ \%$ najbržih pasa.



Slika 2.

Iz podataka na vodoravnoj osi očitavamo da je donja precizna granica prvog razreda 2.35, a gornja granica istog razreda je 3.05. Dakle, precizne granice prvog razreda su: $2.35 - 3.05$. Tada su granice prvog razreda $2.4 - 3.0$. Precizne granice drugog razreda su: $3.05 - 3.75$ i tako dalje. Kumulativne relativne frekvencije su redom: $0.25, 0.55, 0.675, 0.85, 0.95, 1$, a množenjem s 80 (to je ukupan broj promatranih pasa) dobivamo kumulativne frekvencije: $20, 44, 54, 68, 76, 80$. Dakle, frekvencija prvog razreda je 20, frekvencija drugog razreda je $44 - 20 = 24$, frekvencija trećeg razreda je $68 - 54 = 14$ itd. Tablica frekvencija izgleda ovako:

razredi (sekunde)	precizne granice razreda	frekvencije	kumulativne frekvencije	kumulativne rel. frekvencije
$2.4 - 3.0$	$2.35 - 3.05$	20	20	0.25
$3.1 - 3.7$	$3.05 - 3.75$	24	44	0.55
$3.8 - 4.4$	$3.75 - 4.45$	10	54	0.675
$4.5 - 5.1$	$4.45 - 5.15$	14	68	0.85
$5.2 - 5.8$	$5.15 - 5.85$	8	76	0.95
$5.9 - 6.5$	$5.85 - 6.55$	4	80	1

Za izračunavanje prosječnog vremena reakcije gornju ćemo tablicu dopuniti stupcem sredina razreda koje redom glase ovako: $2.7, 3.4, 4.1, 4.8, 5.5, 6.2$. Težinska aritmetička sredina jednaka je

$$\bar{x} = \Big(20 \cdot 2.7 + 24 \cdot 3.4 + 10 \cdot 4.1 + 14 \cdot 4.8 + \\ + 8 \cdot 5.5 + 4 \cdot 6.2\Big) : 80 \\ = 3.91.$$

Dakle, prosječno vrijeme reakcije je 3.91 sekunda.

Na grafu odredimo točku čija je ordinata jednaka 0.1. Ta se točka nalazi na dužini određenoj točkama $(2.35, 0)$ i $(3.05, 0.25)$. Jednadžba pravca kroz te dvije točke glasi $y = \frac{5}{14}x - \frac{47}{56}$, a uvrštavanjem ordinate $y = 0.1$ dobivamo da je odgovarajući $x = 2.63$. Dakle, $10\ \%$ najbržih pasa ima vrijeme reakcije u intervalu $[2.35, 2.63]$. Broj 2.63 naziva se i prvi **decil** ili 10. **centil** ovog skupa podataka.

LITERATURA

1/ S. Varošanec: *Peteljka-list dijagram. Medijan*, Matematika i škola, 71 (2013), 10–13.

2/ S. Varošanec: *Karakteristična petorka i brkata kutija*, Matematika i škola, 72 (2013), 56–59.

3/ S. Varošanec: *Grupirani podatci I*, Matematika i škola, 73 (2014), 100–104.