

9. MJERE DISPERZIJE

Pod pojmom **disperzije** podrazumijevamo raspršenost vrijednosti numeričkoga obilježja. Mjere disperzije služe za ocjenjivanje reprezentativnosti srednje vrijednosti obilježja.

Najčešće u uporabi su ove mjere disperzije:

(a) **Apsolutne mjere disperzije:**

raspon varijacije obilježja,
prosječno apsolutno odstupanje
varijanca i standardna devijacija,
interkvartil.

(b) **Relativne mjere disperzije:**

koeficijent varijacije,
koeficijent kvartilne devijacije.

9.1. RASPON VARIJACIJE OBILJEŽJA

Razlika između najveće i najmanje vrijednosti numeričkoga obilježja naziva se raspon varijacije obilježja.

Ta mjera predstavlja grubu mjeru disperzije obilježja.

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

Npr. Ako je maksimalna težina 110 kg, a minimalna težina 50 kg, onda je raspon varijacije obilježja 60 kg.

9.2. PROSJEČNO APSOLUTNO ODSUPANJE

Prosječno apsolutno odstupanje ("Mean Absolute Deviation") dobiva se kao aritmetička sredina apsolutnih vrijednosti odstupanja od aritmetičke sredine vrijednosti obilježja:

$$MAD = \frac{\sum_{i=1}^N |X_i - \bar{X}|}{N}$$

Primjer 9.1. Visina petoro slučajno odabranih studenata EF-a

Visina u cm	$ X_i - \bar{X} $	$(X_i - \bar{X})^2$
178	1.2	1.44
168	11.2	125.44
176	3.2	10.24
172	7.2	51.84
202	22.8	519.84
896	45.6	708.80

$$MAD = \frac{\sum_{i=1}^N |X_i - \bar{X}|}{N} = \frac{45.6}{5} = 9.12 \text{ cm}$$

9.3. VARIJANCA I STANDARDNA DEVIJACIJA

Varijanca je srednje kvadratno odstupanje numeričkih vrijednosti obilježja od aritmetičke sredine. **Standardna devijacija** je pozitivni korijen iz varijance i predstavlja apsolutnu mjeru disperzije u prvome stupnju.

Varijanca za negrupirane podatke računa se preko sljedećega izraza:

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N} = \frac{708.8}{5} = 141.76 \text{ cm}^2 \quad \text{ili} \quad \sigma^2 = \frac{\sum_{i=1}^N x_i^2}{N} - (\bar{x})^2$$

Standardna devijacija

Standardna devijacija je pozitivni korijen iz varijance:

$$\sigma = +\sqrt{\sigma^2} = +\sqrt{141.76} = 11.9 \text{ cm}$$

Primjer 9.2. Domaćinstva prema broju djece

Broj djece	Broj domaćinstava f_i	Sredina razreda X_i	$f_i \cdot (X_i - \bar{X})^2$
0	100	0	484
1	200	1	288
2	400	2	16
3	150	3	96
4 – (6)	150	5	1176
Σ	1000	-	2060

Varijanca i standardna devijacija za grupirane nizove:

$$\sigma^2 = \frac{\sum_{i=1}^k f_i \cdot (x_i - \bar{x})^2}{\sum_{i=1}^k f_i} = \frac{2060}{1000} = 2.06 \text{ djece}^2$$

$$\sigma = +\sqrt{2.06} = 1.43 \text{ djece}$$

Alternativna formula:

$$\sigma^2 = \frac{\sum_{i=1}^k f_i \cdot x_i^2}{\sum_{i=1}^k f_i} - (\bar{x}^2) = \frac{6900}{1000} - 2.2^2 = 2.06 \text{ djece}^2$$

9.4. INTERKVARTIL

Interkvartil predstavlja apsolutnu mjeru disperzije srednjih 50% jedinica u statističkome skupu. Time interkvartil izbjegava ekstremne vrijednosti obilježja i isključuje 25% najmanjih i 25% najvećih vrijednosti.

$$I_q = Q_3 - Q_1$$

U gornjoj formuli Q_3 predstavlja gornji kvartil koji distribuciju dijeli u omjeru 3:1, tj. 75% jedinica u statističkome skupu ima vrijednost jednaku ili manju od gornjega kvartila, dok 25% jedinica u statističkome skupu ima vrijednost obilježja jednaku ili veću od gornjega kvartila.

Donji kvartil Q_1 dijeli distribuciju u omjeru 1:3.
To znači da 25% jedinica u statističkome skupu ima vrijednost obilježja manju ili jednaku donjemu kvartilu, dok 75% ima vrijednost obilježja veću ili jednaku donjemu kvartilu.

Npr. ako je donji kvartil vijeka trajanja proizvoda 1100 sati, a gornji kvartil 1300 sati, to znači da je raspon varijacije trajanja srednjih 50% jedinica u skupu (po položaju u nizu) jednak 200 sati.

Kvartili zajedno s medijanom dijele statistički skup na četiri jednaka dijela.

Osim kvartila rabe se se i **decili i percentili** s tumačenjem koje je slično kvartilima. U tome slučaju imamo i odgovarajuće mjere disperzije.

Ako je npr. deveti decil jednak 1150 sati onda to znači da 90% jedinica u statističkome skupu (proizvoda) traje 1150 sati ili manje a 10% 1150 sati ili više.

9.5. KOEFICIJENT VARIJACIJE

Koeficijent varijacije relativna je mjera disperzije. Služi za usporedbu disperzije u distribucijama koje mogu imati iste ili različite mjere obilježja. Pomoću standardne devijacije ne može se uspoređivati intenzitet disperzije u različitim distribucijama, naročito ako su jedinice mjere različite.

$$V = \frac{\sigma}{x} \cdot (100) = \frac{11.9}{179.2} \cdot (100) = 6.64\%$$

Njegova vrijednost može iznositi i više od 100% (jer to nije relativni broj strukture!).

9.6. KOEFICIJENT KVARTILNE DEVIJACIJE

Koeficijent kvartilne devijacije relativna je mjera disperzije srednjih 50% jedinica u statističkome nizu.

$$V_q = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{1300 - 1100}{1300 + 1100} = 0.0833$$

Njegova vrijednost kreće se između 0 i 1.
