# Brojčane karakteristike numeričkih nizova

**Numerički niz**

$$x_1, x_2, \ldots, x_N,$$

- $N = \sum_{i=1}^{k} f_i$ – opseg statističkog skupa
- $f_i$ je apsolutna frekvencija modaliteta $x_i$
- $f_{r_i} = \frac{f_i}{N}$ je relativna frekvencija $i$-tog razreda, $\sum_{i=1}^{k} f_{r_i} = 1$
- korigirane frekvencije (u slučaju razreda), $f_{c_i} = \frac{f_i \cdot l_{i_0}}{l_i}$, gdje je $l_i$ veličina $i$-tog razreda, a $l_{i_0}$ najčešća veličina razreda

**Kumulativna funkcija apsolutnih frekvencija**

$$K(x) = \begin{cases} 0, & x < x_{min} \\ \sum\limits_{x_i \le x} f_i, & x_{min} \le x \le x_{max} \\ \mathrm{N}, & x > x_{max} \end{cases}$$

**Kumulativna funkcija relativnih frekvencija**

$$K(x) = \begin{cases} 0, & x < x_{min} \\ \sum\limits_{x_i \le x} f_{r_i}, & x_{min} \le x \le x_{max} \\ \mathrm{N}, & x > x_{max} \end{cases}$$

**Aritmetička sredina**

$$\overline{x} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{x_1 + x_2 + \cdots + x_N}{N} \qquad \text{ili} \qquad \overline{x} = \frac{\sum_{i=1}^{k} x_i f_i}{N}$$

**Geometrijska sredina**

$$x_G = \sqrt[N]{x_1 \cdot x_2 \cdots x_N} \qquad \text{ili} \qquad x_G = \sqrt[N]{x_1^{f_1} \cdot x_2^{f_2} \cdots x_k^{f_k}}$$

**Harmonijska sredina**

$$x_H = \frac{N}{\sum\limits_{i=1}^{N} \frac{1}{x_i}} \qquad \text{ili} \qquad x_H = \frac{N}{\sum\limits_{i=1}^{k} \frac{f_i}{x_i}}$$

**Mod (u slučaju intervala)**

$$Mod = L + \frac{f_j - f_{j-1}}{(f_j - f_{j-1}) + (f_j - f_{j+1})} \cdot d_j,$$

- $L$ donja granica modalnog razreda koji ima frekvenciju $f_j$ (ako su razredi nejednakih veličina, modalni razred je razred s najvećom *korigiranom* frekvencijom, inače apsolutnom)
- $f_{j-1}$ i $f_{j+1}$ su (korigirane) frekvencije razreda ispred i iza modalnog razreda
- $d_j$ veličina modalnog razreda

**Medijan**

$$Med = \begin{cases} x_j, & \text{N neparan } (j = [\frac{N}{2}] + 1) \\ \frac{x_j + x_{j+1}}{2}, & \text{N paran } (j = \frac{N}{2}) \end{cases}$$

U slučaju intervala koristi se formula

$$Med = L + \frac{d_i}{f_{med}} \left( \frac{N}{2} - \sum_{i=1}^{m} f_i \right)$$

- $L$ donja granica medijalnog razreda
- $f_{med}$ je originalna frekvencija medijalnog razreda
- $\sum_{i=1}^{m} f_i$ zbroj frekvencija do medijalnog razreda ne uključujući frekvenciju medijalnog razreda
- $d_i$ veličina medijalnog razreda.

**Raspon varijacije**

$$R_x = x_{max} - x_{min}.$$

U slučaju razreda, raspon varijacije računa se kao razlika gornje granice posljednjeg razreda i donje granice prvog razreda ili kao razlika između razrednih sredina posljednjeg i prvog razreda.

**Varijanca**

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2 \quad \text{za negrupirani niz,} \qquad \sigma^2 = \frac{1}{N}\sum_{i=1}^{k}(x_i - \overline{x})^2 f_i \quad \text{za grupirani niz}$$

odnosno

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \overline{x}^2 \qquad \text{i} \qquad \sigma^2 = \frac{1}{N}\sum_{i=1}^{k} x_i^2 f_i - \overline{x}^2$$

**Standarna devijacija**

$$\sigma = \sqrt{\sigma^2}$$

**Koeficijent varijacije** (u postotcima)

$$KV = \frac{\sigma}{\overline{x}} \cdot 100(\%)$$

**Centralni moment $r$-tog reda**

$$\mu_r = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^r \qquad \text{ili} \qquad \mu_r = \frac{1}{N}\sum_{i=1}^{k}(x_i - \overline{x})^r f_i, \quad \sum_{i=1}^{k} f_i = N$$

**Koeficijent asimetrije** $\gamma_3 = \dfrac{\mu_3}{\sigma^3}$

**Koeficijent spljoštenosti (eksces)** $\varepsilon = \dfrac{\mu_4}{\sigma^4} - 3$

## Uzorak

**Aritmetička sredina uzorka**

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{ili} \quad \overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i f_i$$

gdje je $n$ opseg uzorka (uzorak sadrži $n$ elemenata osnovnog skupa)

**Varijanca uzorka**

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \overline{x})^2 f_i \qquad \text{odnosno} \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{k} x_i^2 f_i - \overline{x}^2$$

U slučaju intervalnih razreda umjesto $x_i$ stavljamo $\overline{x_i}$. Standardna devijacija uzorka $s$ je drugi korijen iz varijance.

**Koeficijent korelacije u uzorku**

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \cdot \sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

**Koeficijent korelacije za grupirane vrijednosti obilježja**

| $x_i \backslash y_j$ | $y_1$ | $y_2$ | $\cdots$ | $y_l$ | $\Sigma$ |
|---|---|---|---|---|---|
| $x_1$ | $f_{11}$ | $f_{12}$ | $\cdots$ | $f_{1l}$ | $f_1^{(x)}$ |
| $x_2$ | $f_{21}$ | $f_{22}$ | $\cdots$ | $f_{2l}$ | $f_2^{(x)}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\cdots$ | $\vdots$ | $\vdots$ |
| $x_k$ | $f_{k1}$ | $f_{k2}$ | $\cdots$ | $f_{kl}$ | $f_k^{(x)}$ |
| $\Sigma$ | $f_1^{(y)}$ | $f_2^{(y)}$ | $\cdots$ | $f_l^{(y)}$ | |

$$r=\frac{n\cdot\sum_{i=1}^{k}\sum_{j=1}^{l}x_iy_jf_{ij}-\sum_{i=1}^{k}x_if_i^{(x)}\cdot\sum_{j=1}^{l}y_jf_j^{(y)}}{\sqrt{\left[n\sum_{i=1}^{k}x_i^2f_i^{(x)}-\left(\sum_{i=1}^{k}x_if_i^{(x)}\right)^2\right]\cdot\left[n\sum_{j=1}^{l}y_j^2f_j^{(y)}-\left(\sum_{j=1}^{l}y_jf_j^{(y)}\right)^2\right]}}$$

**Pravac regresije**

$$y=a+bx,$$

$$a=\frac{\sum_i y_i\cdot\sum_i x_i^2-\sum_i x_i\cdot\sum_i x_iy_i}{n\cdot\sum_i x_i^2-\left(\sum_i x_i\right)^2},\quad b=\frac{n\cdot\sum_i x_iy_i-\sum_i x_i\cdot\sum_i y_i}{n\cdot\sum_i x_i^2-\left(\sum_i x_i\right)^2}$$

$\chi^2$–**test** ($\alpha$ – nivo značajnosti, $(1-\alpha)$ – pouzdanost testa )

$$\chi_0^2=\sum_{i=1}^{k}\frac{(f_i-f_i^*)^2}{f_i^*}$$

$f_i$ empirijske apsolutne frekvencije

$f_i^*$ teorijske apsolutne frekvencije ($f_i^*<5\Rightarrow$ razredi se spajaju)

$$f_i^*=n\cdot P(X=x_i)=n\cdot p_i \quad \text{za diskretnu razdiobu}$$

$$f_i^*=n\cdot P(x_i\le X\le x_{i+1})=n\cdot[F(x_{i+1})-F(x_i)] \quad \text{za neprekidnu razdiobu}$$

c konstanta koja se određuje iz $\chi^2$-razdiobe (tablice) iz uvjeta $F(c)=1-\alpha$, uz $k-r-1$ stupnjeva slobode: $k$ je broj razreda empirijske razdiobe, $r$ broj parametara pretpostavljene razdiobe

Ako je $\chi_0^2\le c$ hipoteza $H_0$ se ne odbacuje, ako je $\chi_0^2>c$ hipoteza $H_0$ se odbacuje.

**Intervalne procjene parametara populacije**

aritmetička sredina populacije: $\mu\in[\overline{x}-c,\overline{x}+c]$ s pouzdanošću $1-\alpha$

- $\overline{x}$ je aritmetička sredina uzorka veličine $n$
- $c$ se računa iz jednakosti $D(\frac{c\sqrt{n}}{\sigma})=1-\alpha$ (tablica standarne norm. razd.)

varijanca populacije: $\sigma^2\in\left[\frac{(n-1)s^2}{c_{(\frac{\alpha}{2},n-1)}},\frac{(n-1)s^2}{c_{(1-\frac{\alpha}{2},n-1)}}\right]$ s pouzdanošću $1-\alpha$

- $s^2$ je varijanca uzorka veličine $n$
- $c_{(\alpha,n)}$ očitava se iz tablice $\chi^2$ razdiobe kao $F(c)=1-\alpha$ za $n$ stupnjeva slobode