

STATYSTYKA OPISOWA - wzory
dr inż Krzysztof Bryś

Szeregi statystyczne

1) **Szereg wyliczający uporządkowany:** (x_1, x_2, \dots, x_n)

przy czym $x_1 \leq x_2 \leq \dots \leq x_n$.

2) **Szereg rozdzielczy punktowy:** $(x_1, x_2, \dots, x_k), (n_1, n_2, \dots, n_k)$,

gdzie $x_1 < x_2 < \dots < x_k$

oraz dla każdego $i = 1, 2, \dots, k$:

n_i -liczba realizacji (obserwacji) wartości x_i ($\sum_{i=1}^k n_i = n$).

3) **Szereg rozdzielczy przedziałowy:** $(y_0; y_1 >, (y_1; y_2 >, \dots, (y_{k-1}; y_k), (n_1, n_2, \dots, n_k)$,

gdzie $y_0 < y_1 < y_2 < \dots < y_{k-1} < y_k$

oraz dla każdego $i = 1, 2, \dots, k$:

n_i -liczba realizacji (obserwacji) wartości należącej do przedziału $(y_{i-1}; y_i)$ ($\sum_{i=1}^k n_i = n$).

Wszystkie wartości należące do przedziału $(y_{i-1}; y_i >, i = 1, 2, \dots, k$ utożsamia się z jego środkiem x_i .

Reguły wyznaczania liczby przedziałów (klas): $k \approx \sqrt{n}, k \leq 5 \log n$.

Parametry empiryczne

Miary położenia rozkładu

1) **Średnia z próby \bar{x}**

- dla szeregu wyliczającego:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

- dla szeregu rozdzielczego:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i \cdot x_i$$

2) **Dominanta (moda, wartość modalna) D** = punkt, w którym funkcja prawdopodobieństwa osiąga największą wartość

- dla szeregu wyliczającego: najczęściej występująca wartość,

- dla szeregu rozdzielczego punktowego: punkt, dla którego liczebność (częstość) osiąga największą wartość, - dla szeregu rozdzielczego przedziałowego (wzór interpolacyjny):

$$D = x_{0d} + \frac{n_d - n_{d-1}}{(n_d - n_{d-1}) + (n_d - n_{d+1})} \cdot h_d,$$

gdzie

x_{0d} - początek przedziału zawierającego dominantę (przedziału o największej liczebności),

h_d - szerokość przedziału zawierającego dominantę (przedziału o największej liczebności),

n_d - liczebność przedziału zawierającego dominantę (największa liczebność),

n_{d-1} - liczebność przedziału poprzedzającego przedział zawierający dominantę,

n_{d+1} - liczebność przedziału następnego po przedziale zawierającym dominantę.

3) **Dystrybuanta empiryczna (częstość skumulowana $F_n(x)$)**

- dla szeregu wyliczającego:

$$F_n(x) = \frac{1}{n} |\{i : x_i < x, i = 1, \dots, n\}|$$

- dla szeregu rozdzielczego:

$$F_n(x) = \sum_{i: x_i < x} \frac{n_i}{n}$$

4) **Kwantyl empiryczny rzędu p $x_{p,n}$:**

(punkt w którym dystrybuanta empiryczna po raz pierwszy osiąga wartość nie mniejszą niż p)

- dla szeregu wyliczającego:

$$x_{p,n} = x_{[np]}$$

- dla szeregu rozdzielczego punktowego:

$$x_{p,n} = x_q \text{ gdzie } q = \min\{r : p \leq \sum_{i=1}^r \frac{n_i}{n}\}$$

- dla szeregu rozdzielczego przedziałowego (wzór interpolacyjny):

$$x_{p,n} = x_{0p} + (np - \sum_{x_i < x_{0p}} n_i) \cdot \frac{h_p}{n_p},$$

gdzie

x_{0p} - początek przedziału zawierającego $x_{p,n}$ (przedziału w którym dystrybuanta empiryczna po raz pierwszy osiąga wartość nie mniejszą niż p),

h_p - szerokość przedziału zawierającego $x_{p,n}$,

n_p - liczebność przedziału zawierającego $x_{p,n}$,

$\sum_{x_i < x_{0p}} n_i$ - liczebność skumulowana dla przedziału poprzedzającego przedział zawierający $x_{p,n}$ (suma liczebności przedziałów poprzedzających)

Mediana: $Me =$ kwantyl rzędu $\frac{1}{2}$

Kwantyl dolny: $Q_1 =$ kwantyl rzędu $\frac{1}{4}$

Kwantyl górny: $Q_3 =$ kwantyl rzędu $\frac{3}{4}$.

Miary rozproszenia rozkładu

5) Wariancja z próby s^2

- dla szeregu wyliczającego:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

- dla szeregu rozdzielczego:

$$s^2 = \frac{1}{n} \sum_{i=1}^k n_i \cdot (x_i - \bar{x})^2$$

6) Odchylenie standardowe z próby $s = \sqrt{s^2}$.

7) Współczynnik zmienności $V = \frac{s}{\bar{x}} \cdot 100\%$.

8) Rozstęp $R =$ różnica między największą i najmniejszą wartością w próbie.

9) Współczynnik asymetrii A_s :

- dla szeregu wyliczającego:

$$A_s = \frac{1}{s^3} \cdot \left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 \right)$$

- dla szeregu rozdzielczego:

$$A_s = \frac{1}{s^3} \cdot \left(\frac{1}{n} \sum_{i=1}^k n_i \cdot (x_i - \bar{x})^3 \right)$$

10) Kurtoza (współczynnik skupienia) K :

- dla szeregu wyliczającego:

$$K = \frac{1}{s^4} \cdot \left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 \right)$$

- dla szeregu rozdzielczego:

$$K = \frac{1}{s^4} \cdot \left(\frac{1}{n} \sum_{i=1}^k n_i \cdot (x_i - \bar{x})^4 \right)$$

11) Współczynnik skośności A_1 :

$$A_1 = \frac{\bar{x} - D}{s}$$