

Statystyka Opisowa

Wzory

Szereg rozdzielczy:

x_i	...	...	...	...
n_i	...	...	...	...

x_i - wartości cechy

n_i - liczebności wartości cechy

$N = \sum n_i$ - liczebność całej zbiorowości

Wskaźnik natężenia przy rysowaniu wykresu szeregu rozdzielczego przedziałowego o nierównych przedziałach:

$$\text{wskaźnik natężenia} = \frac{\text{liczebność klasy} \cdot \text{rozpiętość największej (najszerzej) klasy}}{\text{rozpiętość klasy}}$$

Liczba klas w szeregu rozdzielczym: $k \leq 5 \log N$

Rozpiętość klas w szeregu rozdzielczym: $i = \frac{x_{\max} - x_{\min}}{k}$

I. Miary średnie

Średnia arytmetyczna: $\frac{\sum x_i n_i}{N}$

Średnia harmoniczna: $\frac{N}{\sum \frac{1}{x_i} n_i}$

Średnia geometryczna: $\sqrt[N]{\prod x_i^{n_i}}$

Dominanta: D = Wartość cechy x_i , która ma największą liczebność n_i .

W przypadku szeregu przedziałowego: $D = x_D + \frac{n_D - n_{D-1}}{(n_D - n_{D-1}) + (n_D - n_{D+1})} i_D$,

gdzie:

x_D - dolna granica przedziału, w którym znajduje się dominanta

n_D - liczebność przedziału, w którym znajduje się dominanta

n_{D-1} - liczebność przedziału, poprzedzającego ten z dominantą

n_{D+1} - liczebność przedziału, następującego po tym z dominantą

i_D - rozpiętość przedziału, w którym znajduje się dominanta

Mediana: Mediana dzieli zbiorowość na dwie równe połowy.

Gdy N jest nieparzyste: $Me = \frac{x_{N+1}}{2}$ (wartość cechy, która jest „pośrodku” uporządkowanych rosnąco wartości cech)

Gdy N jest parzyste: $Me = \frac{1}{2} \left(x_{\frac{N}{2}} + x_{\frac{N}{2}+1} \right)$ (średnia wartości cechy, będących „pośrodku”)

Kwartyle $Q_1, Q_2 = Me, Q_3$: Dzielą zbiorowość podobnie jak mediana, tylko że w proporcjach odpowiednio 25% : 75%; 50% : 50%; 75% : 25%.

W przypadku szeregu przedziałowego:

$$Q_1 = x_{Q_1} + \frac{\frac{N}{4} - \sum_{i=1}^{k-1} n_i}{n_{Q_1}} i_{Q_1}$$

$$Q_2 = Me = x_{Me} + \frac{\frac{N}{2} - \sum_{i=1}^{k-1} n_i}{n_{Me}} i_{Me}$$

$$Q_3 = x_{Q_3} + \frac{\frac{3N}{4} - \sum_{i=1}^{k-1} n_i}{n_{Q_3}} i_{Q_3}$$

x_{Q_1}, x_{Me}, x_{Q_3} - dolne granice przedziałów, w którym znajduje się kwartył

$\sum_{i=1}^{k-1} n_i$ - suma liczebności przedziałów POPRZEDZAJĄCYCH ten, w którym znajduje się kwartył

n_{Q_1}, n_{Me}, n_{Q_3} - liczebności przedziałów, w którym znajduje się kwartył

i_{Q_1}, i_{Me}, i_{Q_3} - rozpiętości przedziałów, w którym znajduje się kwartył

Kwantyle: Dzielą zbiorowość podobnie jak mediana i kwartyle, ale w zupełnie dowolnych proporcjach. Na przykład kwantyl rzędu 0,4, czyli $x_{0,4}$ dzieli zbiorowość w proporcji: 40% : 60%.

II. Miary zmienności (dyspersji)

Empiryczny obszar zbieżności: $R = x_{\max} - x_{\min}$

Odchylenie przeciętne: $d = \frac{\sum |x_i - \bar{X}| n_i}{N}$

Odchylenie ćwiartkowe: $Q = \frac{Q_3 - Q_1}{2}$

Typowy obszar zmienności przy pomocy miar pozycyjnych: $Me - Q < x_{typ} < Me + Q$

Wariancja: $S^2 = \frac{\sum (x_i - \bar{X})^2 n_i}{N}$

Odchylenie standardowe: $S = \sqrt{S^2}$

Typowy obszar zmienności: $\bar{X} - S < x_{typ} < \bar{X} + S$

Współczynnik zmienności: $V_s = \frac{S}{\bar{X}} \cdot 100\%$, $V_Q = \frac{Q}{Me} \cdot 100\%$

III. Miary asymetrii

$D > Me > \bar{X}$ - asymetria lewostronna

$D < Me < \bar{X}$ - asymetria prawostronna

Wskaźnik asymetrii: $W_s = \bar{X} - D$, lub $W_s = (Q_3 - Q_2) - (Q_2 - Q_1)$

$W_s > 0$ - asymetria prawostronna

$W_s = 0$ - brak asymetrii, rozkład symetryczny

$W_s < 0$ - asymetria lewostronna

Współczynnik asymetrii (skośności): $A_s = \frac{W_s}{S}$, lub $A_s = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{2Q}$

$$-1 \leq A_s \leq 1$$

Oprócz kierunku asymetrii A_s wskazuje na siłę asymetrii.

Im bliżej 0, tym asymetria jest słabsza. Im bliżej ± 1 , tym asymetria jest silniejsza.

Moment centralny trzeciego rzędu: $m_3 = \frac{\sum (x_i - \bar{X})^3 \cdot n_i}{N}$

Interpretacja ujemności/dodatności taka sama jak W_s .

Moment standaryzowany trzeciego rzędu: $As = \frac{m_3}{S^3}$, $-1 \leq As \leq 1$

Interpretacja siły asymetrii taka sama jak w A_s .

IV. Miary koncentracji

Koncentracja w sensie podziału funduszu cechy:

Współczynnik koncentracji Lorenza: $k = \frac{a}{5000}$, gdzie a to pole pomiędzy linią

równomiernego podziału, a krzywą koncentracji Lorenza.

$k = 0$ oznacza zupełny brak koncentracji (każda jednostka ma taką samą część funduszu cechy)

$k = 1$ oznacza zupełną koncentrację (jedna jednostka posiada cały fundusz cechy)

Koncentracja w sensie koncentracji wokół średniej:

Moment centralny czwartego rzędu: $m_4 = \frac{\sum (x_i - \bar{X})^4 \cdot n_i}{N}$

Moment standaryzowany czwartego rzędu: $a_4 = \frac{m_4}{S^4}$

Punktem odniesienia jest rozkład normalny.

$a_4 > 3$ - wartości cechy są bardziej skoncentrowane wokół średniej, niż w rozkładzie normalnym (wykres wysmukły)

$a_4 = 3$ - wartości cechy tak samo skoncentrowane wokół średniej, jak w rozkładzie normalnym (wykres identyczny)

$a_4 < 3$ - wartości cechy są mniej skoncentrowane wokół średniej, niż w rozkładzie normalnym (wykres spłaszczony)