

1 Statystyka opisowa

Niech $\mathbf{x} = (x_1, \dots, x_n)'$ będzie próbką, tzn. x_1, \dots, x_n są obserwacjami zmiennej (cechy) X .

Zadaniem **statystyki opisowej** jest prezentacja rozkładu cechy X w próbce (rozkładu empirycznego), przy pomocy tabeli lub wykresu. Często wystarczające jest jedynie podanie kilku liczb charakteryzujących ten rozkład.

Metody opisu rozkładu empirycznego:

1. tabelaryczny,
2. graficzny,
3. Statystyki opisowe

- klasyczne - bazujące na uśrednianiu obserwowanych wartości w próbce, np. moment zwykły rzędu r :

$$m_r = \frac{1}{n} \sum_{k=1}^n x_k^r$$

- pozycyjne - bazujące na posortowanych rosnąco wartościach w próbce, np. dolny kwartyl:

$$Q_1 = \frac{1}{2}(x_{(i)} + x_{(j)}),$$

gdzie

$$i = \lceil \frac{n+1}{4} \rceil, \quad j = \lceil \frac{n}{4} \rceil$$

lub górny kwartyl:

$$Q_3 = \frac{1}{2}(x_{(i)} + x_{(j)}),$$

gdzie

$$i = \lceil \frac{3(n+1)}{4} \rceil, \quad j = \lceil \frac{3n}{4} \rceil.$$

Charakterystyki tendencji centralnej rozkładu empirycznego:

1. średnia,

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$$

2. mediana,

$$Me = \begin{cases} x_{(\frac{n+1}{2})}, & n \text{ -- nieparzyste,} \\ \frac{1}{2}[x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}], & n \text{ -- parzyste.} \end{cases}$$

Charakterystyki rozrzutu rozkładu empirycznego:

1. odchylenie standardowe,

$$s = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2}$$

2. współczynnik zmienności

$$v = \frac{s}{\bar{x}} 100$$

Funkcje związane ze statystyką opisową:

1. **table** - szereg rozdzielczy (liczebności),
2. **prop.table** - szereg rozdzielczy (proporcje, częstości),
3. **cut** - dla cechy ilościowej ciągłej podział na przedziały klasowe,
4. **barplot** - wykres słupkowy (cecha jakościowa lub ilościowa dyskretna),
5. **pie** - wykres kołowy (cecha jakościowa lub ilościowa dyskretna),
6. **hist** - histogram (cecha ilościowa ciągła).
7. **mean** - średnia z próby,
8. **median** - mediana z próby,
9. **sd** - odchylenie standardowe z próby.