

Statystyka - wykład I

Pomoce do samodzielnej pracy

ANALIZA DANYCH

Kolejno zbierane informacje:

$$X_1, X_2, \dots, X_n$$

Uporządkowany ciąg danych:

$$X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$$

MIERNIKI POŁOŻENIA

grupa charakterystyk opisująca poziom obserwowanej cechy

- średnia
- mediana
- dolny kwartył
- górny kwartył
- dominanta

ŚREDNIA

liczba charakteryzująca „środek ciężkości”
danych

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

MEDIANA

„geometryczny środek” danych

$$Me = X_{[n/2]:n}$$

DOLNY KWARTYL

rozdziela dane w proporcji 1 : 3, tzn. poniżej dolnego kwartyła znajduje się czwarta część danych, zaś powyżej reszta

$$Q_1 = X_{[n/4]:n}$$

GÓRNY KWARTYL

rozdziela dane w proporcji 3 : 1

$$Q_3 = X_{[3n/4]:n}$$

DOMINANTA

najczęściej występującą wartość

MIERNIKI ROZPROSZENIA

grupa charakterystyk opisująca zróżnicowanie cechy

- wariancja
- odchylenie standardowe
- współczynnik zmienności
- odchylenie przeciętne
- rozstęp
- odchylenie ćwiartkowe

WARIANCJA

charakteryzuje rozrzut danych wokół ich „środka ciężkości”

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

ODCHYLENIE STANDARDOWE

pierwiastek z wariancji

Zaletą odchylenia standardowego jest to, że wyrażone jest ono w tych samych jednostkach co oryginalne dane (jednostką wariacji jest kwadrat jednostek pomiarowych).

WSPÓŁCZYNNIK ZMIENNOŚCI

względne różnicowanie danych, tzn. udział odchylenia standardowego w wartości średniej

$$V = \frac{S}{\bar{x}} 100\%$$

ODCHYLENIE PRZECIĘTNE

podobnie jak odchylenie standardowe, charakteryzuje rozrzut danych wokół ich średniej (jednak każda z danych ma taki sam udział)

$$d = \frac{1}{n} \sum_{i=1}^n |X_i - \bar{x}|$$

ROZSTĘP

różnicą między największą a najmniejszą daną

ODCHYLENIE ĆWIARTKOWE

zakres zmienności środkowych 50% danych

$$Q = \frac{Q_3 - Q_1}{2}$$