

ĆWICZENIE 9

KORELACJA

Cel

Wykonanie obliczeń współczynników korelacji Pearsona i Spearmana oraz analiza otrzymanych wyników.

Wprowadzenie teoretyczne

Współczynnik korelacji Pearsona zwany jest również współczynnikiem korelacji z próby. Określa on zależność liniową między zmiennymi losowymi. Niech X i Y będą zmiennymi losowymi. Niech x_i i y_i będą elementami prób losowych, gdzie $i = 1, 2, \dots, n$. Współczynnikiem korelacji Pearsona jest statystyka dana wzorem

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Analogiczną miarą zależności zmiennych losowych jest **współczynnik korelacji rang Spearmana**. Jest to jedna z nieparametrycznych miar zależności statystycznej między zmiennymi. Współczynnik ten dany jest wzorem

$$r_{xy} = 1 - \frac{6 \sum_{i=1}^n (R_i - S_i)^2}{n(n^2 - 1)},$$

gdzie R_i jest rangą elementu x_i z próby x oraz S_i jest rangą elementu y_i z próby y .

Zadanie do wykonania

1. Obliczyć współczynnik korelacji Pearsona oraz współczynnik korelacji Spearmana dla ciśnienia skurczowego i rozkurczowego z pliku *cisnienie.txt*. Zinterpretować i porównać wyniki. Przedstawić zależność zmiennych od siebie na wykresie.
2. Wykonać analizę korelacji pomiędzy zmiennymi z pliku *dane.xls* (dane o zbiorach pszenicy). Wybrane zależności przedstawić na wykresie. Znaleźć zależność pomiędzy wartością współczynnika korelacji a wyglądem wykresu.

Źródła:

Magiera R. „Modele i metody statystyki matematycznej” GiS, Wrocław 2002