

KORELACJA

Niech X i Y będą zmiennymi losowymi, są one **niezależne** gdy dla każdych liczb rzeczywistych a, b zachodzi równość

$$P(X \leq a)P(Y \leq b) = P(X \leq a \wedge Y \leq b).$$

Natomiast zmienne te są statystycznie zależne, gdy zachodzi nierówność j.w.

Dodatnia zależność monotoniczna – dla dowolnych x, y :

$$P(X < x | Y > y) \leq P(X < x).$$

Ujemna zależność monotoniczna – dla dowolnych x, y :

$$P(X < x | Y > y) \geq P(X < x).$$

UWAGA: korelacja nie musi oznaczać wynikania!

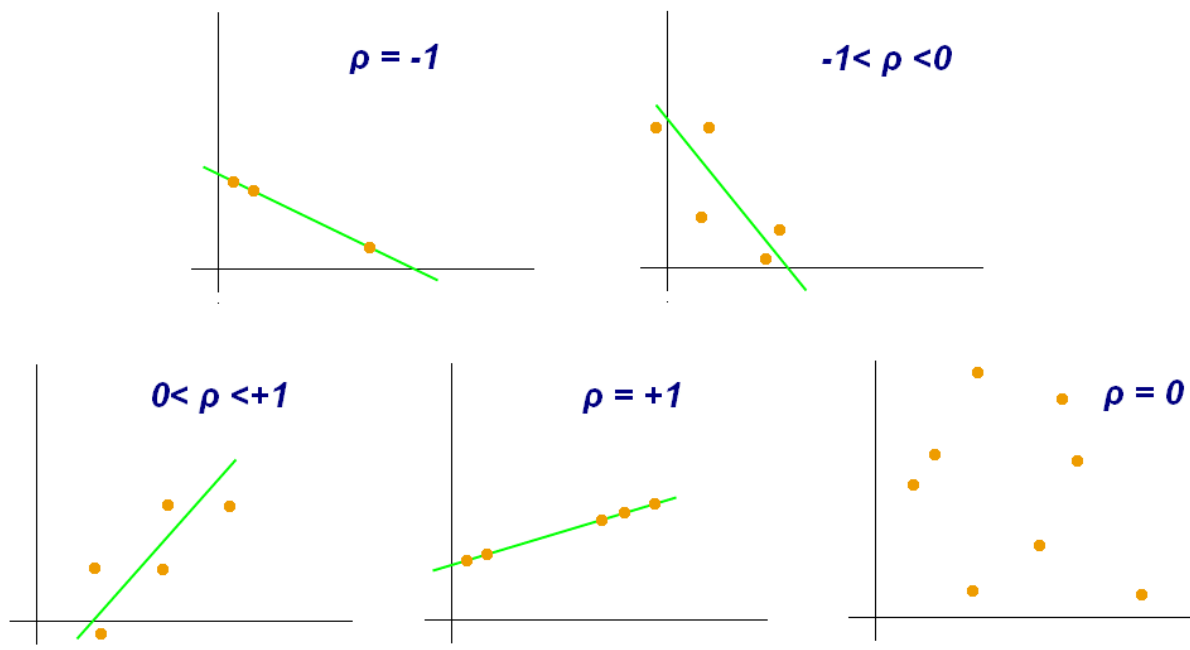
Zależność liniowa między zmiennymi losowymi?

Korelacja Pearsona

$$\text{corr}(X, Y) = \frac{E \left((X - m_x)(Y - m_y) \right)}{\sigma_x \sigma_y}$$

Dla próbki:

$$p_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$



Metoda najmniejszych kwadratów błędów

Zakładamy model zależności Y od X za pomocą funkcji $Y = f(X, \mathbf{a})$, gdzie $\mathbf{a}$ to wektor $a_0, a_1, \dots, a_n$ parametrów modelu, które należy estymować na podstawie próbki złożonej z n par liczb x_i, y_i .

Różnica (błąd) między zmierzonymi wartościami y_i a modelem:

$$r_i = y_i - f(x_i, \mathbf{a})$$

Metoda najmniejszych kwadratów polega na minimalizacji sumy kwadratów błędów $S^2 = \sum_{i=1}^n r_i^2$

W tym celu należy policzyć pochodne S^2 po wektorze $\mathbf{a}$ i rozwiązać układ równań, co da estymator wektora $\mathbf{a}$.

UWAGA: jest tu milczące założenie o odchyłce zmiennej Y !

Metoda najmniejszych kwadratów błędów

Przykład (bardzo prosty) – zależność liniowa $y = ax$

Szukamy minimum $S^2 = \sum_{i=1}^n (y_i - ax_i)^2 = \sum_{i=1}^n (y_i^2 - 2ax_iy_i + a^2x_i^2)$

$$\frac{\partial S^2}{\partial a} = -2 \sum_{i=1}^n x_i y_i + 2a \sum_{i=1}^n x_i^2 = 0$$

Stąd estymator a równy jest

$$\hat{a} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

natomiast estymator odchylenia standardowego a :

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \frac{\sum_{i=1}^n (ax_i - y_i)^2}{\sum_{i=1}^n x_i^2}}$$

Zadanie: wyprowadzić wzory na estymatory dla funkcji $y = a_1x + a_0$

Zaimplementować wzory w arkuszu kalkulacyjnym i sprawdzić, czy one działają.

