Metody Probabilistyczne i Statystyka - wykład 9.

Charakterystyki liczbowe dwuwymiarowych zmiennych losowych
Analiza korelacji

6 maja 2025

Wartość oczekiwana

Założenie:

Niech $g : \mathbb{R}^2 \to \mathbb{R}$ będzie funkcją taką, że $g(X, Y)$ jest zmienną losową jednowymiarową.

Wartość oczekiwana

Twierdzenie

Jeśli (X, Y) *ma rozkład dyskretny, to*

$$E\left(g(X,Y)\right) = \sum_{(x,y)\in S_{XY}} g(x,y)\cdot P(X=x, Y=y).$$

Wartość oczekiwana

Twierdzenie

Jeśli (X, Y) *ma rozkład dyskretny, to*

$$E\left(g(X, Y)\right) = \sum_{(x,y)\in S_{XY}} g(x, y) \cdot P(X = x, Y = y).$$

W szczególności:

$$EX = \sum_{(x,y)\in S_{XY}} x \cdot P(X = x, Y = y)$$

Wartość oczekiwana

Twierdzenie

Jeśli (X, Y) *ma rozkład dyskretny, to*

$$E\left(g(X,Y)\right) = \sum_{(x,y)\in S_{XY}} g(x,y)\cdot P(X=x, Y=y).$$

W szczególności:

$$EX = \sum_{(x,y)\in S_{XY}} x\cdot P(X=x, Y=y)$$

$$EY = \sum_{(x,y)\in S_{XY}} y\cdot P(X=x, Y=y)$$

Wartość oczekiwana

Twierdzenie

Jeśli (X, Y) *ma rozkład dyskretny, to*

$$E\left(g(X,Y)\right) = \sum_{(x,y)\in S_{XY}} g(x,y)\cdot P(X=x, Y=y).$$

W szczególności:

$$EX = \sum_{(x,y)\in S_{XY}} x\cdot P(X=x, Y=y)$$

$$EY = \sum_{(x,y)\in S_{XY}} y\cdot P(X=x, Y=y)$$

$$E(XY) = \sum_{(x,y)\in S_{XY}} x\cdot y\cdot P(X=x, Y=y).$$

Wartość oczekiwana

Przykład 1.

Wektor (X, Y) ma rozkład dyskretny o funkcji prawdopodobieństwa

$X \setminus Y$	-1	0	1	2
1	0.1	0.2	0	0.1
2	0.1	0	0.3	0.2

Obliczyć EX, EY, $E(XY)$.

Wartość oczekiwana

Twierdzenie

Jeśli (X, Y) *ma rozkład ciągły, to*

$$E\left(g(X,Y)\right) = \iint_{\mathbb{R}^2} g(x,y) f_{XY}(x,y)\,dx\,dy.$$

Wartość oczekiwana

Twierdzenie

Jeśli (X, Y) *ma rozkład ciągły, to*

$$E\left(g(X,Y)\right) = \iint_{\mathbb{R}^2} g(x,y) f_{XY}(x,y)\,dx\,dy.$$

W szczególności:

$$EX = \iint_{\mathbb{R}^2} x \cdot f_{XY}(x,y)\,dx\,dy$$

Wartość oczekiwana

Twierdzenie

Jeśli (X, Y) *ma rozkład ciągły, to*

$$E\left(g(X, Y)\right) = \iint_{\mathbb{R}^2} g(x, y) f_{XY}(x, y)\, dx\, dy.$$

W szczególności:

$$EX = \iint_{\mathbb{R}^2} x \cdot f_{XY}(x, y)\, dx\, dy$$

$$EY = \iint_{\mathbb{R}^2} y \cdot f_{XY}(x, y)\, dx\, dy$$

Wartość oczekiwana

Twierdzenie

Jeśli (X, Y) *ma rozkład ciągły, to*

$$E\left(g(X, Y)\right) = \iint_{\mathbb{R}^2} g(x, y) f_{XY}(x, y)\, dx\, dy.$$

W szczególności:

$$EX = \iint_{\mathbb{R}^2} x \cdot f_{XY}(x, y)\, dx\, dy$$

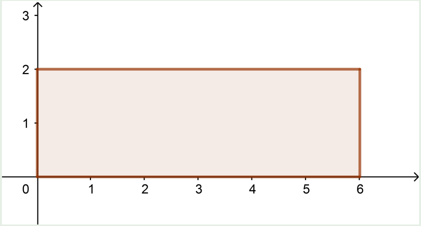
$$EY = \iint_{\mathbb{R}^2} y \cdot f_{XY}(x, y)\, dx\, dy$$

$$E(XY) = \iint_{\mathbb{R}^2} x \cdot y \cdot f_{XY}(x, y)\, dx\, dy.$$

Wartość oczekiwana

Przykład 2.

Wektor (X, Y) ma rozkład jednostajny w prostokącie $D = [0; 6] \times [0; 2]$. Wyznaczyć EX, EY, $E(XY)$.



Wartość oczekiwana

Uwagi:

- Punkt o współrzędnych (EX, EY) jest środkiem ciężkości nośnika rozkładu łącznego zmiennej losowej (X, Y).

Uwagi:

- Punkt o współrzędnych (EX, EY) jest środkiem ciężkości nośnika rozkładu łącznego zmiennej losowej (X, Y).
- Jeśli X i Y są niezależnymi zmiennymi losowymi, to

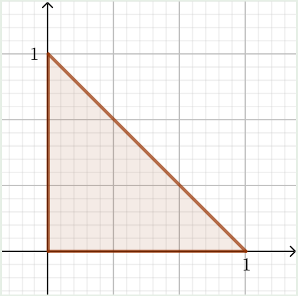
$$E(X \cdot Y) = EX \cdot EY.$$

Wartość oczekiwana

Przykład 3.

Wektor (X, Y) ma rozkład jednostajny w trójkącie D o wierzchołkach $(0,0)$, $(1,0)$, $(0,1)$.

$$f_{XY}(x,y) = 2 \cdot 1_D(x,y).$$



Sprawdzić, czy $E(XY) = EX \cdot EY$.

Analiza korelacji

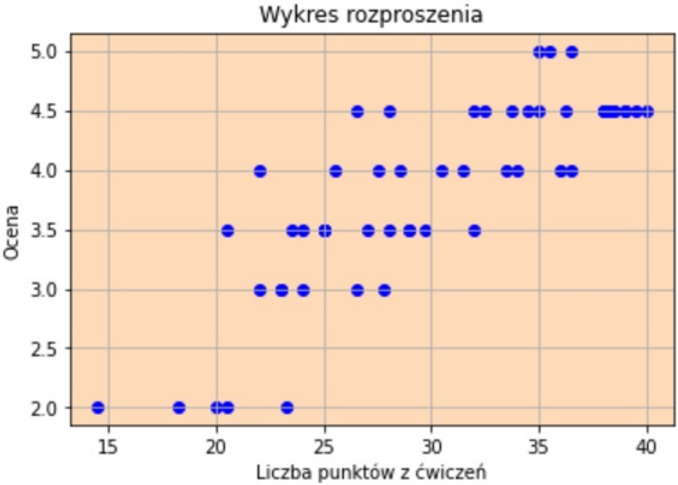
Analiza korelacji

Analiza korelacji -

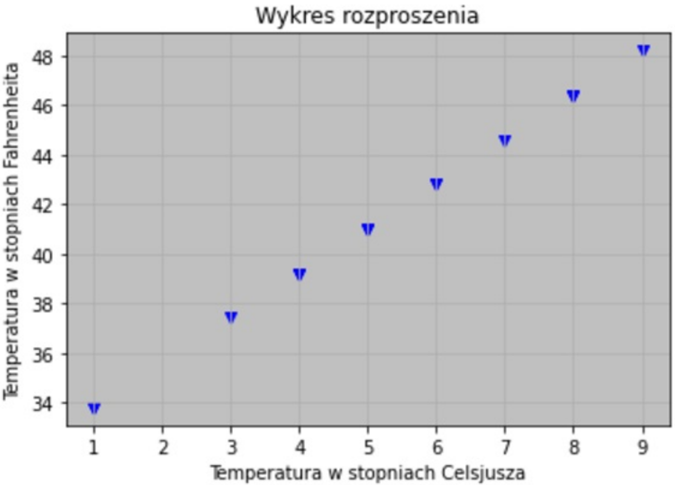
Analiza korelacji

Analiza korelacji - badanie istnienia i siły liniowego związku między cechami populacji (zmiennymi losowymi).

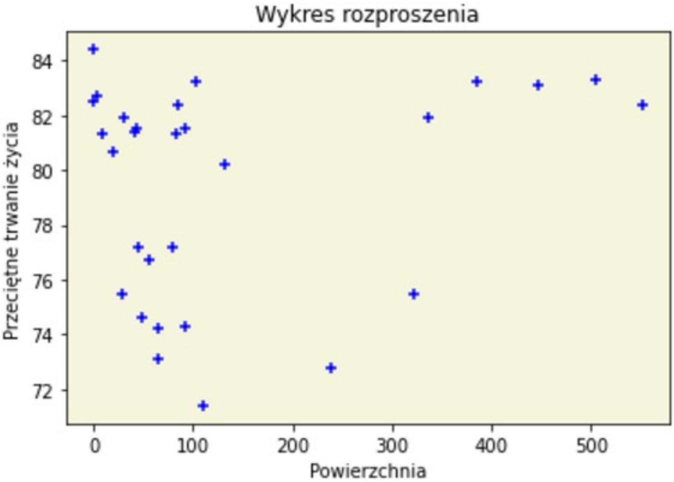
Analiza korelacji



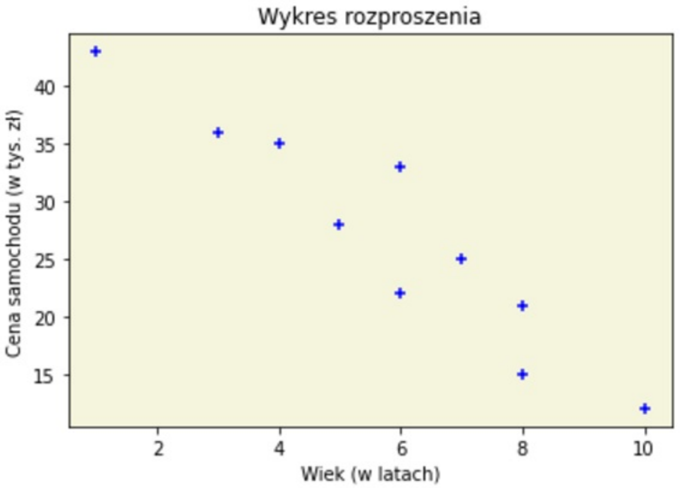
Analiza korelacji



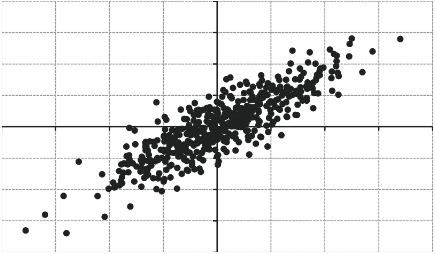
Analiza korelacji



Analiza korelacji

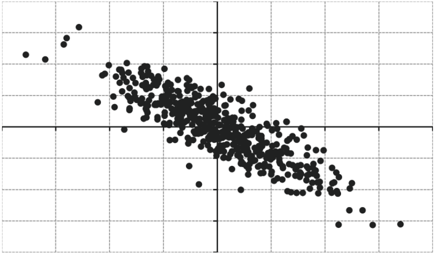


Korelacja



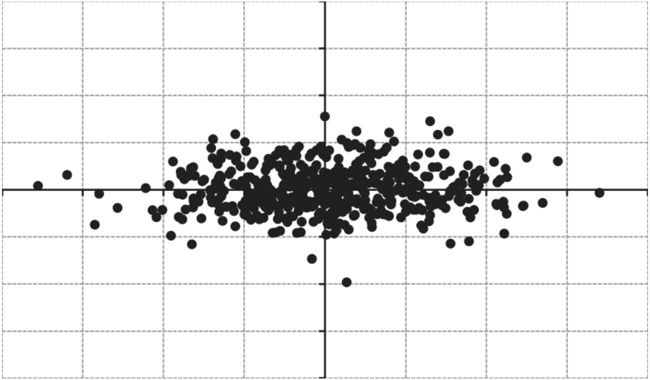
- **Korelacja dodatnia** - wzrostowi (spadkowi) wartości jednej cechy towarzyszy wzrost (spadek) wartości drugiej cechy

Korelacja



- **Korelacja ujemna** - wzrostowi (spadkowi) wartości jednej cechy towarzyszy spadek (wzrost) wartości drugiej cechy

Korelacja



- **Korelacja zerowa** - brak korelacji (cechy są **nieskorelowane**).

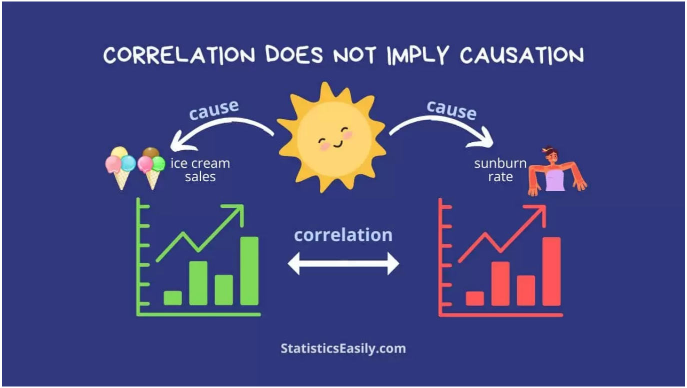
Korelacja nie oznacza przyczynowości:

Korelacja

Korelacja nie oznacza przyczynowości: fakt, że dwie cechy są ze sobą skorelowane, nie musi oznaczać, że jedna jest przyczyną drugiej.

Korelacja

Korelacja nie oznacza przyczynowości: fakt, że dwie cechy są ze sobą skorelowane, nie musi oznaczać, że jedna jest przyczyną drugiej.



Kowariancja i współczynnik korelacji

Definicja

Kowariancją *zmiennych losowych* X *i* Y, *dla których istnieją* EX, EY, $E(XY)$, *nazywamy liczbę*

$$cov(X,Y) = E[(X - EX)(Y - EY)].$$

Kowariancja i współczynnik korelacji

Definicja

Kowariancją *zmiennych losowych* X *i* Y, *dla których istnieją* EX, EY, $E(XY)$, *nazywamy liczbę*

$$cov(X, Y) = E[(X - EX)(Y - EY)].$$

Uwaga

Miernikiem liniowej zależności między X i Y jest współczynnik korelacji.

Kowariancja i współczynnik korelacji

Definicja

Kowariancją *zmiennych losowych* X *i* Y*, dla których istnieją* EX, EY, $E(XY)$*, nazywamy liczbę*

$$cov(X,Y) = E[(X - EX)(Y - EY)].$$

Uwaga

Miernikiem liniowej zależności między X i Y jest współczynnik korelacji.

Definicja

Niech X *i* Y *będą zmiennymi losowymi takimi, że* $VX > 0$ *i* $VY > 0$*.*

Kowariancja i współczynnik korelacji

Definicja

Kowariancją *zmiennych losowych* X *i* Y, *dla których istnieją* EX, EY, $E(XY)$, *nazywamy liczbę*

$$cov(X, Y) = E[(X - EX)(Y - EY)].$$

Uwaga

Miernikiem liniowej zależności między X i Y jest współczynnik korelacji.

Definicja

Niech X *i* Y *będą zmiennymi losowymi takimi, że* $VX > 0$ *i* $VY > 0$. **Współczynnikiem korelacji** *zmiennych* X *i* Y *nazywamy liczbę*

$$\rho(X, Y) = \frac{cov(X, Y)}{\sqrt{VX \cdot VY}}.$$

Kowariancja i współczynnik korelacji dla próby losowej

Jeśli dysponujemy próbą n-elementową $(x_1, y_1), \ldots, (x_n, y_n)$, to:

- $cov(X, Y) = \frac{1}{n-1} \cdot \sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y})$

Kowariancja i współczynnik korelacji dla próby losowej

Jeśli dysponujemy próbą n-elementową $(x_1, y_1), \ldots, (x_n, y_n)$, to:

- $cov(X, Y) = \frac{1}{n-1} \cdot \sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y})$
- Współczynnik korelacji (**współczynnik korelacji próbkowej Pearsona**) obliczamy zgodnie ze wzorem:

$$r(X, Y) = \frac{\sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^{n} (x_k - \bar{x})^2 \cdot \sum_{k=1}^{n} (y_k - \bar{y})^2}},$$

przy czym

Kowariancja i współczynnik korelacji dla próby losowej

Jeśli dysponujemy próbą n-elementową $(x_1, y_1), \ldots, (x_n, y_n)$, to:

- $cov(X, Y) = \frac{1}{n-1} \cdot \sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y})$
- Współczynnik korelacji (**współczynnik korelacji próbkowej Pearsona**) obliczamy zgodnie ze wzorem:

$$r(X, Y) = \frac{\sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^{n} (x_k - \bar{x})^2 \cdot \sum_{k=1}^{n} (y_k - \bar{y})^2}},$$

przy czym

$$\sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y}) =$$

Kowariancja i współczynnik korelacji dla próby losowej

Jeśli dysponujemy próbą n-elementową $(x_1, y_1), \ldots, (x_n, y_n)$, to:

- $$cov(X, Y) = \frac{1}{n-1} \cdot \sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y})$$
- Współczynnik korelacji (**współczynnik korelacji próbkowej Pearsona**) obliczamy zgodnie ze wzorem:

$$r(X, Y) = \frac{\sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^{n} (x_k - \bar{x})^2 \cdot \sum_{k=1}^{n} (y_k - \bar{y})^2}},$$

przy czym

$$\sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y}) = \sum_{k=1}^{n} (x_k - \bar{x}) \cdot y_k =$$

Kowariancja i współczynnik korelacji dla próby losowej

Jeśli dysponujemy próbą n-elementową $(x_1, y_1), \ldots, (x_n, y_n)$, to:

- $cov(X, Y) = \frac{1}{n-1} \cdot \sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y})$
- Współczynnik korelacji (**współczynnik korelacji próbkowej Pearsona**) obliczamy zgodnie ze wzorem:

$$r(X, Y) = \frac{\sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^{n} (x_k - \bar{x})^2 \cdot \sum_{k=1}^{n} (y_k - \bar{y})^2}},$$

przy czym

$$\sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y}) = \sum_{k=1}^{n} (x_k - \bar{x}) \cdot y_k = \sum_{k=1}^{n} (y_k - \bar{y}) \cdot x_k,$$

gdzie

Kowariancja i współczynnik korelacji dla próby losowej

Jeśli dysponujemy próbą n-elementową $(x_1, y_1), \ldots, (x_n, y_n)$, to:

- $cov(X, Y) = \frac{1}{n-1} \cdot \sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y})$
- Współczynnik korelacji (**współczynnik korelacji próbkowej Pearsona**) obliczamy zgodnie ze wzorem:

$$r(X, Y) = \frac{\sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^{n} (x_k - \bar{x})^2 \cdot \sum_{k=1}^{n} (y_k - \bar{y})^2}},$$

przy czym

$$\sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y}) = \sum_{k=1}^{n} (x_k - \bar{x}) \cdot y_k = \sum_{k=1}^{n} (y_k - \bar{y}) \cdot x_k,$$

gdzie

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

Kowariancja i współczynnik korelacji dla próby losowej

Jeśli dysponujemy próbą n-elementową $(x_1, y_1), \ldots, (x_n, y_n)$, to:

- $cov(X, Y) = \frac{1}{n-1} \cdot \sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y})$
- Współczynnik korelacji (**współczynnik korelacji próbkowej Pearsona**) obliczamy zgodnie ze wzorem:

$$r(X, Y) = \frac{\sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^{n} (x_k - \bar{x})^2 \cdot \sum_{k=1}^{n} (y_k - \bar{y})^2}},$$

przy czym

$$\sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y}) = \sum_{k=1}^{n} (x_k - \bar{x}) \cdot y_k = \sum_{k=1}^{n} (y_k - \bar{y}) \cdot x_k,$$

gdzie

$$\bar{x} = \frac{1}{n} \sum_{k=1}^{n} x_k, \quad \bar{y} = \frac{1}{n} \sum_{k=1}^{n} y_k.$$

Korelacja

- **Dodatnia korelacja między cechami X i Y:**
  $cov(X, Y) > 0, \ \rho(X, Y) > 0$

Korelacja

- **Dodatnia korelacja między cechami X i Y:**
  $cov(X, Y) > 0, \ \rho(X, Y) > 0$
- **Ujemna korelacja między cechami X i Y:**
  $cov(X, Y) < 0, \ \rho(X, Y) < 0$

Korelacja

- **Dodatnia korelacja między cechami X i Y:** $cov(X,Y) > 0,\ \rho(X,Y) > 0$
- **Ujemna korelacja między cechami X i Y:** $cov(X,Y) < 0,\ \rho(X,Y) < 0$
- **Cechy X i Y nieskorelowane:** $cov(X,Y) = 0,\ \rho(X,Y) = 0$

Współczynnik korelacji

Własności współczynnika korelacji

- $|\rho(X, Y)| \leqslant 1$;

Współczynnik korelacji

Własności współczynnika korelacji

- $|\rho(X, Y)| \leqslant 1$;
- $|\rho(X, Y)| = 1 \longleftrightarrow$ istnieją $a, b \in \mathbb{R}$ takie, że $Y = aX + b$

Współczynnik korelacji

Własności współczynnika korelacji

- $|\rho(X, Y)| \leqslant 1$;
- $|\rho(X, Y)| = 1 \longleftrightarrow$ istnieją $a, b \in \mathbb{R}$ takie, że $Y = aX + b$
- X i Y są nieskorelowane $\longleftrightarrow \rho(X, Y) = 0$.

Współczynnik korelacji

Własności współczynnika korelacji c.d.

Jeśli:

- $0 < |\rho(X, Y)| < 0.25$, to korelacja jest **bardzo słaba (pomijalna)**

Współczynnik korelacji

Własności współczynnika korelacji c.d.

Jeśli:

- $0 < |\rho(X, Y)| < 0.25$, to korelacja jest **bardzo słaba (pomijalna)**
- $0.25 \leqslant |\rho(X, Y)| < 0.5$, to korelacja jest **słaba**

Współczynnik korelacji

Własności współczynnika korelacji c.d.

Jeśli:

- $0 < |\rho(X, Y)| < 0.25$, to korelacja jest **bardzo słaba (pomijalna)**
- $0.25 \leqslant |\rho(X, Y)| < 0.5$, to korelacja jest **słaba**
- $0.5 \leqslant |\rho(X, Y)| < 0.75$, to korelacja jest **umiarkowana**

Współczynnik korelacji

Własności współczynnika korelacji c.d.

Jeśli:

- $0 < |\rho(X, Y)| < 0.25$, to korelacja jest **bardzo słaba (pomijalna)**
- $0.25 \leqslant |\rho(X, Y)| < 0.5$, to korelacja jest **słaba**
- $0.5 \leqslant |\rho(X, Y)| < 0.75$, to korelacja jest **umiarkowana**
- $0.75 \leqslant |\rho(X, Y)| < 1$, to korelacja jest **silna**

Współczynnik korelacji

Własności współczynnika korelacji c.d.

Jeśli:

- $0 < |\rho(X,Y)| < 0.25$, to korelacja jest **bardzo słaba (pomijalna)**
- $0.25 \leqslant |\rho(X,Y)| < 0.5$, to korelacja jest **słaba**
- $0.5 \leqslant |\rho(X,Y)| < 0.75$, to korelacja jest **umiarkowana**
- $0.75 \leqslant |\rho(X,Y)| < 1$, to korelacja jest **silna**
- $|\rho(X,Y)| = 1$, to korelacja jest **pełna**.

Kowariancja

Uwaga

Jeśli zmienne losowe X i Y są niezależne, to $cov(X, Y) = 0$.

Kowariancja

Uwaga

Jeśli zmienne losowe X i Y są niezależne, to $cov(X, Y) = 0$.
Ale: z tego, że X i Y są nieskorelowane nie wynika, że są niezależne!

Kowariancja

Uwaga

Jeśli zmienne losowe X i Y są niezależne, to $cov(X, Y) = 0$.
Ale: z tego, że X i Y są nieskorelowane nie wynika, że są niezależne!

Wyjątkiem jest rozkład normalny:

Jeśli (X, Y) ma rozkład normalny, to X i Y są niezależne $\longleftrightarrow$ $cov(X, Y) = 0$.

Kowariancja

Twierdzenie

$$cov(X, Y) = E(X \cdot Y) - EX \cdot EY.$$

Kowariancja

Własności kowariancji

1. $cov(X, Y) = cov(Y, X)$

Kowariancja

Własności kowariancji

1. $cov(X, Y) = cov(Y, X)$
2. $cov(X, X) = VX$

Kowariancja

Własności kowariancji

1. $cov(X, Y) = cov(Y, X)$
2. $cov(X, X) = VX$
3. $cov(a, X) = 0$ dla $a \in \mathbb{R}$

Kowariancja

Własności kowariancji

1. $cov(X, Y) = cov(Y, X)$
2. $cov(X, X) = VX$
3. $cov(a, X) = 0$ dla $a \in \mathbb{R}$
4. $cov(aX + b, cY + d) = a \cdot c \cdot cov(X, Y)$ dla $a, b, c, d \in \mathbb{R}$

Kowariancja

Własności kowariancji

1. $cov(X, Y) = cov(Y, X)$
2. $cov(X, X) = VX$
3. $cov(a, X) = 0$ dla $a \in \mathbb{R}$
4. $cov(aX + b, cY + d) = a \cdot c \cdot cov(X, Y)$ dla $a, b, c, d \in \mathbb{R}$
5. $$\begin{aligned} & cov(aX_1 + bX_2, cY_1 + dY_2) = \\ = \; & a \cdot c \cdot cov(X_1, Y_1) + a \cdot d \cdot cov(X_1, Y_2) + \\ + \; & b \cdot c \cdot cov(X_2, Y_1) + b \cdot d \cdot cov(X_2, Y_2) \end{aligned}$$

Wariancja sumy i różnicy zmiennych losowych

Twierdzenie

Jeśli X i Y są zmiennymi losowymi, dla których istnieją wariancje, to

$$\begin{aligned} V(X+Y) &= VX + 2cov(X,Y) + VY \\ V(X-Y) &= VX - 2cov(X,Y) + VY. \end{aligned}$$

Wariancja sumy i różnicy zmiennych losowych

Twierdzenie

Jeśli X i Y są zmiennymi losowymi, dla których istnieją wariancje, to

$$\begin{aligned} V(X+Y) &= VX + 2cov(X,Y) + VY \\ V(X-Y) &= VX - 2cov(X,Y) + VY. \end{aligned}$$

Wniosek

Jeśli X i Y są nieskorelowane, to

$$V(X+Y) = V(X-Y) = VX + VY.$$