

Współczynnik Korelacji Między Dwoma Zmiennymi Losowymi

Gabriel Tyszk

25 czerwca 2025

Spis treści

1	Wprowadzenie do Korelacji	1
2	Współczynnik Korelacji Pearsona	1
2.1	Definicja i Obliczanie	1
2.2	Interpretacja Wartości	1
3	Inne Miary Zależności i Ich Zastosowanie	2
4	Zastosowania Korelacji	2

1 Wprowadzenie do Korelacji

W statystyce, korelacja lub zależność to statystyczny związek, niezależnie od tego, czy jest przyczynowy, czy nie, między dwiema zmiennymi losowymi lub danymi dwuwymiarowymi. Chociaż w najszerszym sensie "korelacja" może wskazywać na każdy rodzaj związku, w statystyce zazwyczaj odnosi się do stopnia, w jakim para zmiennych jest liniowo powiązana. Nieformalnie, "korelacja" jest synonimem "zależności".

Formalnie, zmienne losowe są zależne, jeśli nie spełniają matematycznej właściwości niezależności probabilistycznej. W sensie technicznym, korelacja odnosi się do kilku specyficznych typów zależności matematycznych, gdzie oczekiwana wartość jednej zmiennej warunkowa na drugą nie jest stała. Istnieje kilka współczynników korelacji, często oznaczanych jako ρ lub r , mierzących stopień korelacji.

2 Współczynnik Korelacji Pearsona

Najbardziej znaną miarą zależności między dwoma wielkościami jest współczynnik korelacji Pearsona. Mierzy on stopień oraz kierunek liniowego związku między dwiema zmiennymi.

2.1 Definicja i Obliczanie

Współczynnik Pearsona jest uzyskiwany poprzez wzięcie stosunku kowariancji dwóch zmiennych w zbiorze danych, znormalizowanego przez iloczyn ich odchyłeń standardowych. Matematycznie, dla populacji zmiennych losowych X i Y z wartościami oczekiwanymi μ_X i μ_Y oraz odchyleniami standardowymi σ_X i σ_Y , współczynnik korelacji $\rho_{X,Y}$ jest zdefiniowany jako:

$$\rho_{X,Y} = \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}, \quad \text{jeśli } \sigma_X \sigma_Y > 0.$$

Gdzie E jest operatorem wartości oczekiwanej, a cov oznacza kowariancję. Współczynnik Pearsona jest zdefiniowany tylko wtedy, gdy oba odchylenia standardowe są skończone i dodatnie.

2.2 Interpretacja Wartości

Wartość współczynnika korelacji Pearsona zawiera się w przedziale od -1 do +1.

- **+1:** Wskazuje na doskonały, bezpośredni (rosnący) związek liniowy.
- **-1:** Wskazuje na doskonały, odwrotny (malejący) związek liniowy.

- **Wartość bliska 0:** Wskazuje na brak zależności liniowej.
- **Wartości w otwartym przedziale $(-1, 1)$:** Wskazują na stopień zależności liniowej między zmiennymi; im bliżej -1 lub 1 , tym silniejsza korelacja.

Współczynnik Pearsona próbuje ustalić linię najlepszego dopasowania przez zestaw danych, a jego znak wskazuje na korelację negatywną lub pozytywną.

3 Inne Miary Zależności i Ich Zastosowanie

Oprócz współczynnika Pearsona, istnieją inne ważne miary zależności:

- **Współczynniki korelacji rang (Rank correlation coefficients):** Np. współczynnik korelacji rang Spearmana i współczynnik korelacji rang Kendalla (τ). Mierzą one, w jakim stopniu wzrost jednej zmiennej wiąże się ze wzrostem drugiej, nie wymagając, aby ten wzrost był liniowy. Są używane, aby być bardziej odporne niż współczynnik Pearsona i wykrywać mniej ustrukturyzowane relacje między zmiennymi.
- **Korelacja odległościowa (Distance Correlation) i Randomized Dependence Coefficient (RDC):** Te miary adresują niedostatek korelacji Pearsona polegający na tym, że może ona wynosić zero dla zmiennych zależnych. Ważną ich właściwością jest to, że wartość zero dla tych miar zazwyczaj oznacza niezależność. Wadą może być ich niższa moc w porównaniu do korelacji Pearsona, gdy dane pochodzą z rozkładu normalnego wielowymiarowego.
- **Iloraz szans (Odds Ratio):** Dla dwóch zmiennych binarnych mierzy ich zależność.
- **Współczynnik determinacji (R kwadrat):** W przypadku modelu liniowego z jedną zmienną niezależną, współczynnik determinacji jest kwadratem współczynnika Pearsona (r_{xy}^2). Generalizuje on współczynnik korelacji na regresję wielokrotną.

4 Zastosowania Korelacji

Korelacje są przydatne, ponieważ mogą wskazywać na związek predykcyjny, który można wykorzystać w praktyce. Na przykład, zakład energetyczny może produkować mniej energii w łagodny dzień, bazując na korelacji między zapotrzebowaniem na energię elektryczną a pogodą. W tym konkretnym przypadku istnieje związek przyczynowy, ponieważ ekstremalne warunki pogodowe powodują, że ludzie zużywają więcej energii elektrycznej do ogrzewania lub chłodzenia.