Urząd Statystyczny w Lublinie

Lubelski KONKURS statystyczno-demograficzny

Współczynnik korelacji Pearsona

XIV Lubelski Konkurs Statystyczno-Demograficzny Sigma Kwadrat

Urząd Statystyczny w Lublinie

TPS

UMCS

NBP NARODOWY BANK POLSKI

Współczynnik korelacji Pearsona

Lubelski KONKURS statystyczno-demograficzny

Co to jest korelacja?

Sprawdzamy, czy dwie cechy zmieniają się razem.

Przykład

Czy więcej godzin nauki wiąże się zwykle z lepszym wynikiem testu?

Każda kropka oznacza jednego ucznia.

Korelacja opisuje współwystępowanie dwóch cech. Sama nie dowodzi, że jedna cecha powoduje drugą.



Pytanie:

Czy punkty układają się wzdłuż rosnącej lub malejącej linii?

Urząd Statystyczny w Lublinie | TPS | UMCS | NBP Narodowy Bank Polski

Współczynnik korelacji Pearsona

Lubelski KONKURS statystyczno-demograficzny

Jak czytać współczynnik r?

Wartość r mieści się od -1 do +1.

r = +1

Idealna dodatnia zależność liniowa.

r ≈ 0

Brak lub słaba zależność liniowa.

r = -1

Idealna ujemna zależność liniowa.



Im |r| bliżej 1, tym silniejsza zależność liniowa; znak wskazuje kierunek.

Urząd Statystyczny w Lublinie | TPS | UMCS | NBP Narodowy Bank Polski

Współczynnik korelacji Pearsona

Lubelski KONKURS statystyczno-demograficzny

Wzór Pearsona

Najpierw poznaj wzór, potem użyj go w tabeli.

$$r = \Sigma[(x_i - \bar{x})(y_i - \bar{y})] \,/\, \sqrt{\{\Sigma(x_i - \bar{x})^2 \cdot \Sigma(y_i - \bar{y})^2\}}$$

Oznaczenia:

- x, y — badane cechy
- $\bar{x}$, $\bar{y}$ — średnie z obu cech
- Σ — suma wszystkich obserwacji
- i — kolejna para danych

Plan obliczeń

1. Oblicz średnie $\bar{x}$ i $\bar{y}$.
2. Oblicz odchylenia od średnich.
3. Pomnóż odpowiadające sobie odchylenia i zsumuj.
4. Oblicz sumy kwadratów odchyleń.
5. Podziel licznik przez pierwiastek z iloczynu sum kwadratów.

Dane muszą występować w parach: każdy x odpowiada temu samemu obiektowi co y.

Urząd Statystyczny w Lublinie · TPS · UMCS · NBP Narodowy Bank Polski

Współczynnik korelacji Pearsona

Lubelski KONKURS statystyczno-demograficzny

Przykład – obliczamy krok po kroku

Godziny nauki (x) i wynik testu (y) dla 4 uczniów.

Uczeń	x	y	$x-\bar{x}$	$y-\bar{y}$	Iloczyn	$(x-\bar{x})^2$	$(y-\bar{y})^2$
1	1	2	−1,5	−1,5	2,25	2,25	2,25
2	2	3	−0,5	−0,5	0,25	0,25	0,25
3	3	4	0,5	0,5	0,25	0,25	0,25
4	4	5	1,5	1,5	2,25	2,25	2,25

Średnie

$\bar{x} = 2{,}5$ oraz $\bar{y} = 3{,}5$

Sumy

Σ iloczynów = 5; obie sumy kwadratów = 5

$r = 5 / \sqrt{(5 \cdot 5)} = 1$ → idealna dodatnia korelacja liniowa

Urząd Statystyczny w Lublinie · TPS · UMCS · NBP Narodowy Bank Polski

Współczynnik korelacji Pearsona

Lubelski KONKURS statystyczno-demograficzny

Szybszy wzór obliczeniowy

Przy większej liczbie danych wygodna jest postać z sumami.

Wzór obliczeniowy

$$r = [n\Sigma xy - (\Sigma x)(\Sigma y)] \,/\, \sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}$$

Przygotuj kolumny:

- x
- y
- x^2
- y^2
- xy

Na końcu zsumuj każdą kolumnę.

Sprawdź

✓ Tyle samo wartości x i y.

✓ Każda para danych jest poprawna.

✓ Wynik należy do $[-1; 1]$.

✓ Nie dziel przez zero.

Współczynnik korelacji Pearsona

Lubelski KONKURS statystyczno-demograficzny

Jak interpretować wynik?

Przykładowe przedziały – granice są umowne, zależą od dziedziny.

r od 0,70 do 1,00	silna dodatnia korelacja
r od 0,70 do 1,00	umiarkowana dodatnia korelacja
r od −0,29 do +0,29	słaba lub brak korelacji liniowej
r od −0,29 do +0,29	umiarkowana ujemna korelacja
r od −0,29 do +0,29	silna ujemna korelacja

Pamiętaj:

Korelacja nie oznacza przyczynowości. r bliskie 0 nie wyklucza zależności nieliniowej.

Urząd Statystyczny w Lublinie | TPS | UMCS | NBP Narodowy Bank Polski

Współczynnik korelacji Pearsona

Lubelski KONKURS statystyczno-demograficzny

Zadanie do samodzielnego rozwiązania

Oblicz r dla pięciu par danych.

Godziny nauki (x)	1	2	3	4	5
Wynik testu (y)	8	7	5	4	1

Podpowiedź

1. Oblicz $\bar{x}$ i $\bar{y}$.
2. Dodaj kolumny x^2, y^2 i xy.
3. Zastosuj wzór obliczeniowy.
4. Zinterpretuj znak i siłę.

Pytanie: czy zależność jest dodatnia, czy ujemna?

Odpowiedź do sprawdzenia po obliczeniu: r ≈ −0,98 — silna ujemna korelacja.

Urząd Statystyczny w Lublinie | TPS | UMCS | NBP Narodowy Bank Polski

Współczynnik korelacji Pearsona

Lubelski KONKURS statystyczno-demograficzny

Zapamiętaj 4 rzeczy

1

r zawsze mieści się od −1 do +1.

2

Znak + lub − pokazuje kierunek zależności liniowej.

3

Im |r| bliżej 1, tym silniejsza zależność liniowa.

4

Korelacja nie dowodzi związku przyczynowego.

Lubelski Ośrodek Badań Regionalnych

Urząd Statystyczny w Lublinie