

**1. Analiza korelacji** → badanie stopnia i kierunku zależności liniowej między dwiema zmiennymi

**Współczynnik korelacji liniowej Pearsona  $r_{xy}$**  przyjmuje wartości w przedziale od -1 do 1, czyli

siła zależności mieści się w przedziale [0-1] → im bliżej 1 tym silniejsza zależność, im bliżej 0 tym słabsza zależność. Można posłużyć się następującą tabelą:

$r_{xy} < 0,2$  – brak związku liniowego  
 $0,2-0,4$  – zależność wyraźna liniowa ale niska  
 $0,4 - 0,7$  zależność umiarkowana  
 $0,7 - 0,9$  –zależność znacząca  
 powyżej  $0,9$  – zależność bardzo silna

Za kierunek zależności odpowiada znak + lub -. + oznacza korelację dodatnią, wprost proporcjonalną, czyli wraz ze wzrostem wartości jednej cechy rosną wartości drugiej. – oznacza zależność ujemną, odwrotnie proporcjonalną, czyli wraz ze wzrostem wartości jednej cechy maleją wartości drugiej.

Wzór na współczynnik korelacji liniowej Pearsona:

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

Lub

$$r_{xy} = \frac{cov(X,Y)}{s(X)s(Y)}, \text{ gdzie } cov(X,Y) = \frac{1}{N} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Przykład obliczeniowy:

$X_i$	$Y_i$	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	85	=1-3=-2	=85-110=-25	50	4	625
3	105	=3-3=0	=105-110=-5	0	0	25
2	100	=2-3=-1	-10	10	1	100
3	110	0	0	0	0	0
4	125	1	15	15	1	225
3	115	0	5	0	0	25
5	130	2	20	40	4	400
<b>21</b>	<b>770</b>			<b>115</b>	<b>10</b>	<b>1400</b>

$$\bar{x} = \frac{21}{7} = 3 \quad \bar{y} = \frac{770}{7} = 110$$

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{115}{\sqrt{10 * 1400}} = 0,972$$

**Zadanie 1.** W tabeli podano roczne spożycie 2 artykułów spożywczych w kg: tłuszczów (art. A) – zmienna zależna i cukrów (art. B)-zmienna niezależna. Oblicz i zinterpretuj współczynnik korelacji liniowej Pearsona.

Art. A	Art. B.
10	20
20	25
25	35
30	30
40	45
45	50
60	60

**Zadanie 2.** Za pomocą Pearsona obliczyć i zinterpretować zależność pomiędzy czasem przeznaczonym na reklamę telewizora marki XX w min/msc, a miesięczną ich sprzedażą.

Czas (X)	Sprzedaż w tys. (Y)
10	2,5
18	4,6
13	5,2
14	4,0
20	5,6
15	3,2
8	1,5

**Zadanie 3.** Do badania wylosowano 6 sklepów branży konfekcyjnej z Łodzi, odnotowując ich przeciętny dzienny obrót w tys. (Y) oraz powierzchnię w m<sup>2</sup> (X). Wyznacz i zinterpretuj współczynnik korelacji liniowej Pearsona.

Y	X
2	30
4	45
6	50
9	60
14	75
25	100

**Zadanie 4.** W pewnym obserwatorium obserwowano średnią temperaturę nocy w lutym C (X) oraz maksymalną głębokość zamrożonej gleby w cm (Y). Wyznacz i zinterpretuj współczynnik korelacji liniowej Pearsona.

Y	X
20	-3
5	-1
20	0
25	0
15	1
5	1

