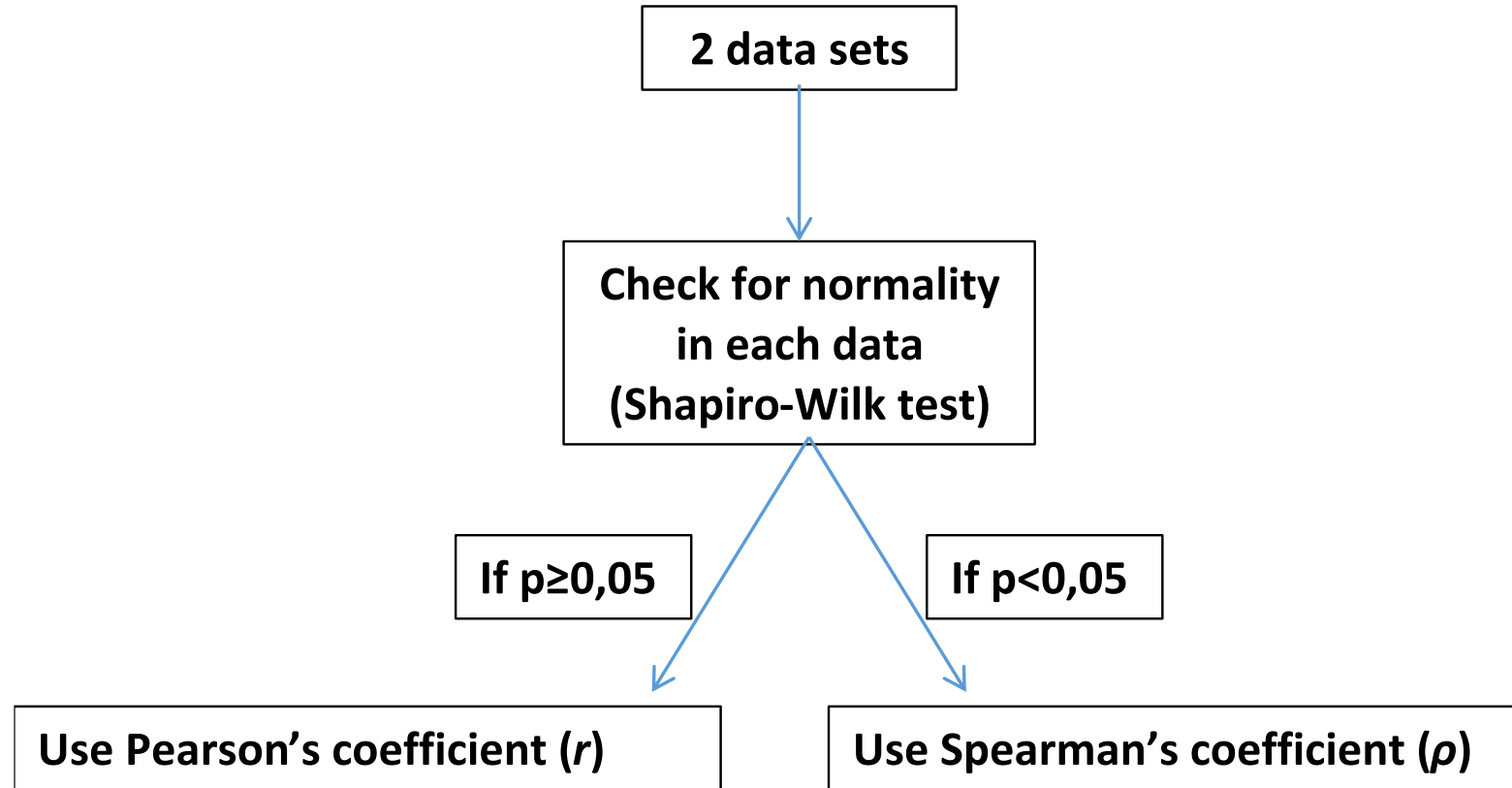


Statystyka i modelowanie w naukach o środowisku

Wykład 10

Analiza korelacji

Korelacja



Based on: Daniel Granato, Verônica Maria de Araújo Calado, Basil Jarvis
„Observations on the use of statistical methods in Food Science and Technology” Food
Research International 55 (2014) 137–149.

Analiza korelacji - współczynnik korelacji Pearsona

Cel: ocena współzależności między dwiema zmiennymi ilościowymi

Ocenia jedynie zależność liniową.

$$r = \frac{cov(X, Y)}{s_x \cdot s_y}$$

gdzie, wartość kowariancji (cov) na podstawie próby liczymy wg następującego wzoru:

$$cov(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

natomiast s_x i s_y są odchyleniami standardowymi dla zmiennych: X i Y

Analiza korelacji - współczynnik korelacji Pearsona

Współczynnik korelacji liniowej przyjmuje zawsze wartości w zakresie $[-1,1]$.
Im większa wartość bezwzględna współczynnika, tym silniejsza jest zależność liniowa między zmiennymi.

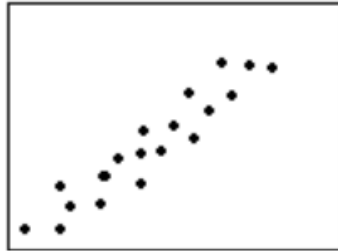
$r_{xy} = 0$ oznacza brak korelacji,

$r_{xy} = 1$ oznacza silną korelację dodatnią, jeżeli jedna zmienna (X) rośnie to również rośnie druga zmienna (Y),

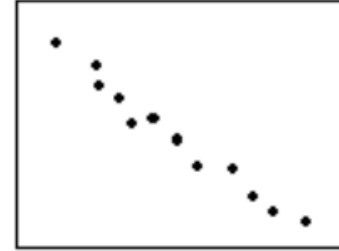
$r_{xy} = -1$ oznacza korelację ujemną (jeżeli zmienna X rośnie, to Y maleje i na odwrót).

Analiza korelacji - współczynnik korelacji Pearsona

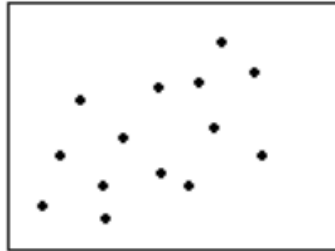
Stopień korelacji



silna dodatnia ($r = 0,8$)



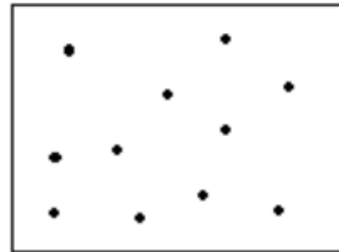
silna ujemna ($r = -0,8$)



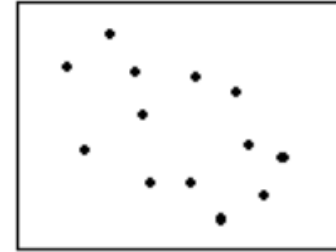
słaba dodatnia ($r = 0,3$)



umiarkowana ujemna ($r = -0,5$)

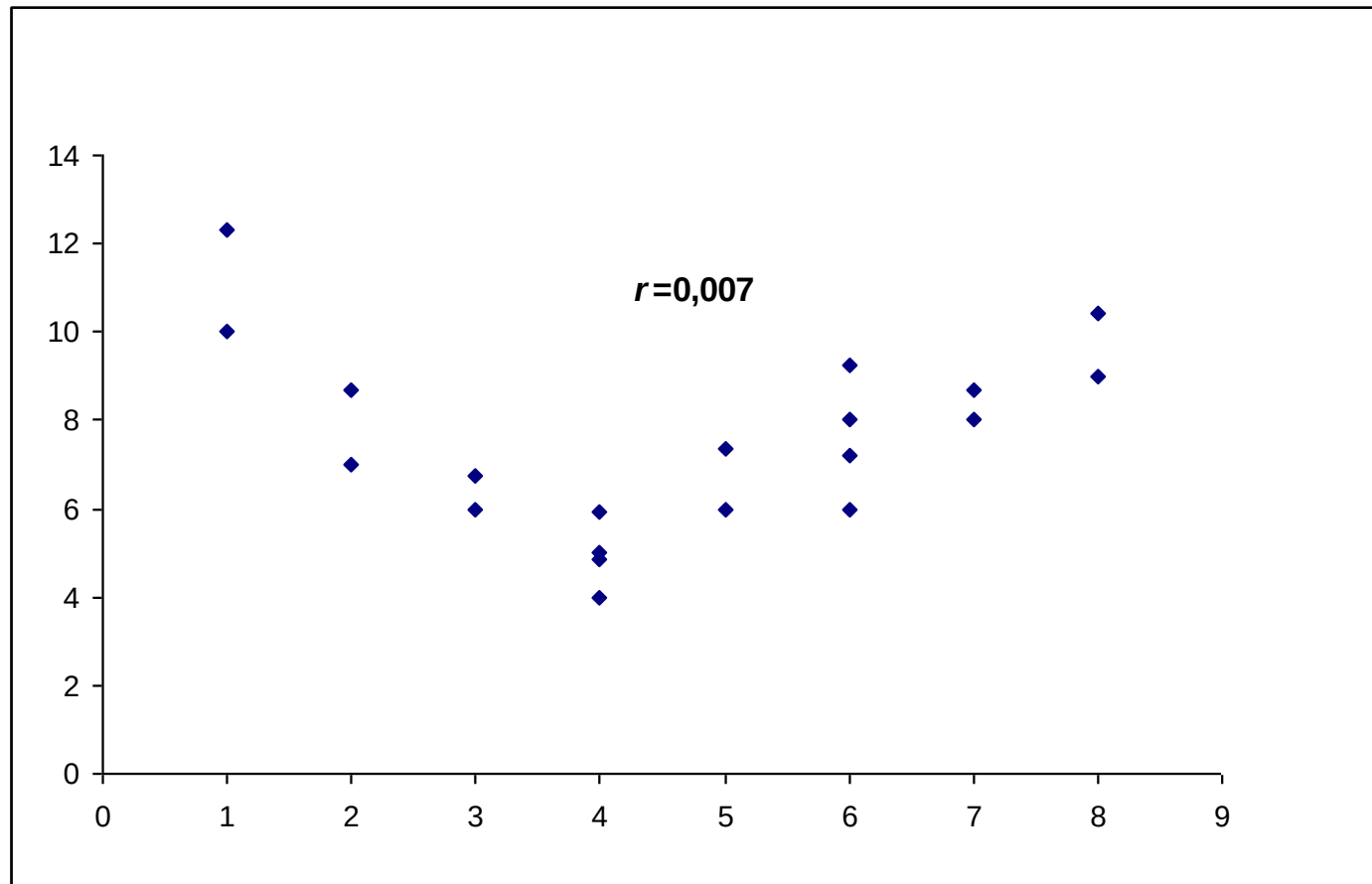


brak korelacji ($r = 0,0$)



słaba ujemna ($r = -0,3$)

Analiza korelacji - współczynnik korelacji Pearsona vs współzależność nieliniowa



Analiza korelacji - współczynnik korelacji Pearsona

Hipoteza zerowa: $H_0: \rho = 0$

Hipoteza zerowa: $H_1: \rho \neq 0$

Test t-Studenta, gdy zmienne pochodzą z rozkładu normalnego

$$t_{emp} = \frac{r}{\sqrt{1 - r^2}} \sqrt{n - 2}$$

$t_{\alpha, n-2}$ - jest wartością krytyczną z rozkładu t-Studenta

Jeżeli $|t_{emp}| > t_{\alpha, n-2}$ lub $p < \alpha$ to H_0 odrzucamy.

Analiza korelacji - współczynnik korelacji Pearsona

Testowanie istotności korelacji

Testowanie jest tylko wtedy uzasadnione, gdy obydwie zmienne mają rozkład normalny

Hipoteza zerowa: $H_0: \rho=0$

ρ - wartość współczynnika korelacji dla całej populacji

Jeżeli $|r_{\text{emp}}| > r_{\alpha, 2, n-2}$ to H_0 odrzucamy.

$r_{\alpha, 2, n-2}$ – jest wartością krytyczną współczynnika korelacji prostej Pearsona

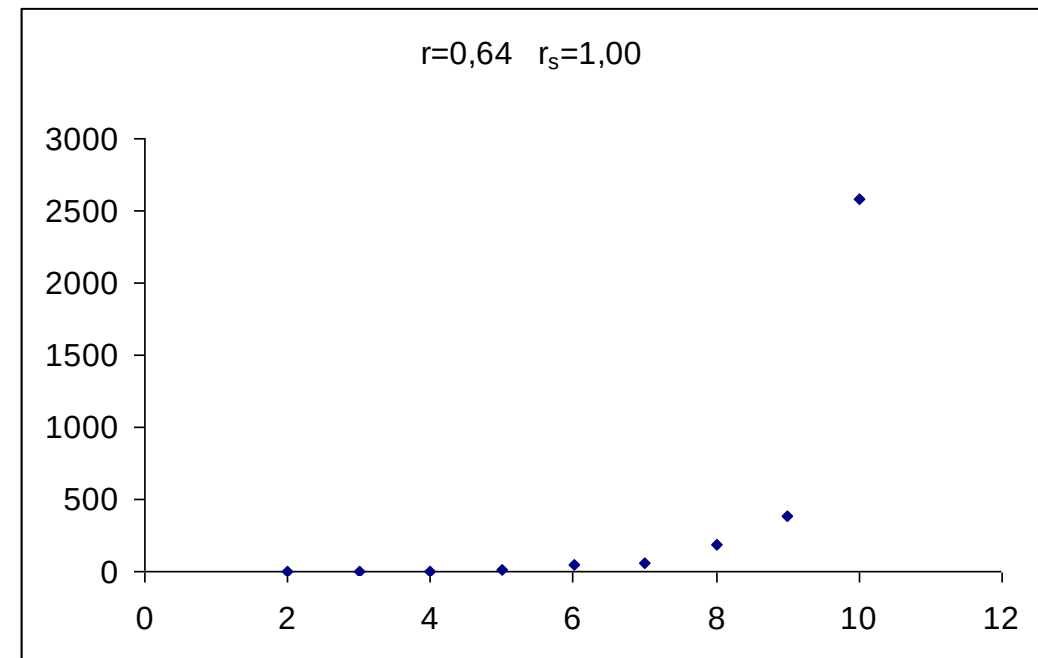
Podobnie jak w przypadku innych hipotez w programach statystycznych (wnioskowanie o istotności współzależności dwóch zmiennych odbywa się na podstawie wartości p ($p < \alpha$ oznacza istotna współzależność))

Należy pamiętać również, że współczynnik korelacji liniowej Pearsona dobrze opisuje jedynie zależności liniowe. W przypadku, gdy zależność istnieje ale jest nieliniowa (np. punkty są położone na paraboli) wartość współczynnika korelacji może być bliska 0.

Analiza korelacji - współczynnik korelacji rang Spearmana

Współczynnik korelacji rang Spearmana (r_s) służy do oceny współzależności między dwiema zmiennymi. W odróżnieniu od współczynnika korelacji Pearsona można przy pomocy współczynnika korelacji Spearmana oceniać zależności nieliniowe. Przy testowaniu nie jest wymagana normalność rozkładu zmiennych, tak więc możliwe jest stosowanie tego współczynnika korelacji wtedy, gdy nie możemy stosować współczynnika korelacji Pearsona.

Wartości współczynnika korelacji rang Spearmana są z zakresu $[-1, 1]$ a ich interpretacja jest podobna jak w przypadku współczynnika korelacji Pearsona, czyli czym wartość r_s jest bliższa 1 tym zależność jest silniejsza, dodatnia, czym bliższa jest -1 tym zależność silniejsza, ujemna, a jeśli wartość r_s jest bliska 0 to oznacza brak zależności lub bardzo słabą zależność.



Analiza korelacji - przykład

Ocenić z użyciem analizy korelacji, czy występuje związek między cechami X i Y.

kalkulatory online

Pearson:

<https://www.socscistatistics.com/pvalues/pearsondistribution.aspx>

Spearman:

<https://www.socscistatistics.com/tests/spearman/default.aspx>

X	Y
0,5	10,3
0,7	12,3
1,2	15,6
1,4	16,8
1,6	17,5
1,55	17,9
1,4	18,5
1,8	18,2
1,7	18,6
1,9	16,2
2,3	15,8
2,4	15,4
2,5	14,6
2,1	17,1
2,8	9,6

Analiza korelacji - przykład

STATISTICA:

Statystyka → statystyki podstawowe → macierze korelacji → jedna lista zmiennych (wybieramy obie zmienne)

W opcjach wybieramy „Wyświetl dokładną tabelę wyników”

H_0 : nie ma korelacji pomiędzy zmiennymi X i Y

$r=0,056$

$t=0,204$

$p=0,842$

$p > \alpha$

H_0 przyjmujemy

korelacja pomiędzy zmiennymi X i Y nie jest istotna

Korelacje i korelacje cząstkowe: Arkusz1

☒ Jedna lista zmien. ☐ Dwie listy zmiennych ☐ Podsumowanie

Pierw. lista: X-Y
Druga lista: brak

Podstawowe | Więcej | Opcje | Kolory

Zawartość macierzy korelacji

☐ Wyświetl macierz (podświetl p)
☐ Wyświetl r, p i N
☒ Wyświetl dokładną tabelę wyników

☐ Pokaż długie nazwy zmiennych
☐ Obliczenia zwiększonej precyzji
poziom p dla podświetlania: .05
☒ Średnie i odch. std. w macierzy dla 1 listy

Anuluj

Opcje

Grupami...

SELECT CASES S W

☐ Momenty ważone

DF = ☒ W-1 ☐ N-1

Usuwanie BD
☒ Przypadkami
☐ Parami

Analiza korelacji - przykład

STATISTICA:

Statystyka → nieparametryczne → korelacje → jedna lista zmiennych (wybieramy obie zmienne)