

Statystyka dla Inżynierów
Laboratorium 13
Analiza Korelacji

Dane do zadania 4 znajdują się w pliku mieszkania.csv

Wygenerowanie skorelowanych zmiennych z dwuwymiarowego rozkładu normalnego: Wynik 1:
Niech X, Y będą niezależnymi zmiennymi losowymi oraz $U = c_1X + c_2Y$, gdzie $c_1^2 + c_2^2 = 1$. Wtedy, $\rho(X, U) = c_1$

Wynik 2: Skalowanie liniowe zmiennych X i U nie zmienia współczynnika korelacji.

1. Poniżej podano liczbę punktów zdobytych na egzaminach z analizy oraz algebry przez 6 studentów.
 - a) Wyznaczyć współczynnik korelacji według i) Pearsona ii) Spearmana, iii) Kendalla
 - b) Korzystając z odpowiednich testów, zweryfikować hipotezę iż wyniki z tych egzaminów są nieskorelowane.

Student	1	2	3	4	5	6
Analiza	28	26	23	18	14	12
Algebra	25	27	20	24	16	13

2. Niech X oraz Y będą niezależnymi zmiennymi losowymi o rozkładzie normalnym standardowym.
 - a) Za pomocą generatora wylosować 100 realizacji pary (X, Y)
 - b) Wyznaczyć współczynnik korelacji $\rho(X, Y)$ i wykonać test hipotezy $H_0: \rho(X, Y) = 0$ według i) Pearsona, ii) Spearmana, iii) Kendalla za pomocą odpowiedniego polecenia.
 - c) Wykonać test hipotezy $H_0: \rho(X, Y) = 0$ według i) Pearsona, ii) Spearmana, iii) Kendalla za pomocą testu permutacyjnego (za pomocą 1000 symulacji)
 - d) Wyznaczyć przedział ufności na poziomie ufności 95% dla $\rho(X, Y)$ i wykonać test hipotezy $H_0: \rho(X, Y) = 0$ według i) Pearsona, ii) Spearmana, iii) Kendalla za pomocą Bootstrapu (za pomocą 1 000 symulacji – w każdej symulacji należy wybrać 100 par (X_i, Y_i) z oryginalnej próby z zwracaniem).
 - e) Sporządzić rozrzut zmiennych X i Y .
3. Niech $V = 0,2X + \sqrt{0,96} Y$ (X i Y tak jak w zad. 2). Stworzyć 100 realizacji pary (X, V) i zbadać korelację między zmiennymi X i V analogicznie jak w zad. 2.
4.
 - a) Wyznaczyć macierz współczynnik korelacji Pearsona między następującymi zmiennymi: metraż, liczba pokoi, cena, cena za m².
 - b) Wyznaczyć macierz współczynnik korelacji Spearmana między zmiennymi: cena, metraż, cena za m², liczba pokoi.
 - c) Wyznaczyć macierz współczynnik korelacji Kendalla między zmiennymi: cena, metraż, cena za m², liczba pokoi.
 - d) Opisać relacje między tymi zmiennymi.
 - e) Wykonać test hipotezy że współczynnik korelacji między ceną za m² a metrażem wynosi zero według i) Pearsona, ii) Spearmana, iii) Kendalla.

5. Wygenerowanie „wzrostu” i „wagi” z dwuwymiarowego rozkładu normalnego

- a) Niech X oraz Y będą niezależnymi zmiennymi losowymi o rozkładach normalnych standardowych oraz $V = \rho X + Y\sqrt{1 - \rho^2}$. Za pomocą generatora wylosować 100 realizacji pary (X, V) przy $\rho = 0,7$.
- b) Niech $H = 170 + 12X$ oraz $W = 65 + 10V$. Wyznaczyć i) średnie oraz odchylenia standardowe tych zmiennych, ii) współczynnik korelacji Pearsona, $r_P(H, W)$, ii) współczynnik korelacji Spearmana, $r_S(H, W)$, iii) współczynnik korelacji Kendalla, $r_K(H, W)$.
- c) Narysować estymatory gęstości zmiennych H i W .

- d) Zweryfikować hipotezy że H i W pochodzą z rozkładów normalnych.
- e) Sporządzić rozrzut zmiennych H i W .