

1.1 Wczytaj zbiór danych *daneSoc.csv* (przykład z książki 'Przewodnik po pakiecie R', P. Biecek)

- a) Podaj podstawowe informacje dotyczące zbioru danych i zmiennych (wymiary danych, nazwy zmiennych, typy zmiennych).
- b) Wyznacz tablicę kontyngencji dla zmiennych **wykształcenie** i **praca** (funkcja `table()`).
- c) Wyznacz podstawowe statystyki (średnia, odchylenie standardowe, mediana, zakres) dla zmiennej opisującej ciśnienie skurczowe w grupie mężczyzn ze średnim wykształceniem.
- d) Sporządź wykresy typu boxplot dla zmiennej opisującej ciśnienie skurczowe w grupach mężczyzn zatrudnionych i bez pracy.
- e) Znajdź pacjentów z wykształceniem średnim których ciśnienie skurczowe jest pomiędzy 140 i 150.
- f) Znajdź pacjenta (pacjentów) z największą wartością ciśnienia skurczowego.
- g) Znajdź pacjenta (pacjentów) których ciśnienie skurczowe jest większe niż empiryczny kwantyl rzędu 0.8 tej zmiennej.

1.2 Sporządź wykresy kwantylowe normalne dla prób losowych o licznosciach 10, 50, 100 i 500 pochodzących z rozkładów:

- a) normalnego $N(0, 1)$,
- b) gamma z parametrami 2 i 2,
- c) Cauchy'ego z parametrami 0 i 1.

1.3 W pliku *skorelowana_probka.txt* znajdują się sztucznie wygenerowane obserwacje (x_i, y_i) $i = 1, 2, \dots, 105$. Chcemy zbadać zależność x od y .

- a) Narysuj wykres zależności y od x , czy widzimy zależność między x , a y
- b) Oblicz z definicji, a także za pomocą funkcji `cor` współczynnik korelacji Pearsona. Co możemy wywnioskować z uzyskanego wyniku?
- c) Korzystając z faktu:

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{d} N(0, (1 - \rho^2)^2)$$

zweryfikuj na poziomie istotności $\alpha = 0.05$ hipotezę o nieskorelowaniu zmiennych.

- d) Korzystając z faktu:

$$\sqrt{n}(\text{arctanh } \hat{\rho} - \text{arctanh } \rho) \xrightarrow{d} N(0, 1)$$

wyznacz 95% asymptotyczny przedział ufności dla współczynnika korelacji.

- e) Dokonaj rangowania zmiennych x i y . Sporządź wykres rozproszenia rang y w zależności od rang x .
- f) Oblicz z definicji współczynnik ρ Spearmana i współczynnik τ Kendalla. W jaki sposób można wytłumaczyć wyższe wartości uzyskanych współczynników w porównaniu z korelacją pearsonowską?

g) Zduplicuj ostatni wiersz. Oblicz współczynniki ρ Spearmana i τ Kendalla. W jaki sposób jest rangowana powtarzająca się obserwacja?