

SAR 2020/2021 Laboratorium 1

6-7.10.2020

1.1

Wczytaj zbiór danych *daneSoc.csv* (przykład z książki *Przewodnik po pakiecie R*, P. Biecek) `Data <- read.csv("ściezka-do-pliku/daneSoc.csv", sep=";")`

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

1.2 wykresy kwantylowe

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

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

1.3 wykresy

Wczytaj zbiór danych *mtcars* `data(mtcars)`.

- Narysuj histogram i wykres skrzynkowy zmiennej `mpg`.
- Narysuj oddzielne wykresy skrzynkowe dla `mpg` dla każdej liczby cylindrów oddzielnie (zmienna `cyl`).
- Narysuj wykres punktowy zmiennej `mpg`.
- Narysuj wykres punktowy zmiennej `mpg` w zależności od zmiennej `hp` (horsepower).

1.4 model liniowy

Skorzystaj z danych *mtcars* `data(mtcars)`.

- Objasnij za pomocą modelu liniowego wpływ pojemności skokowej (`disp`) na spalanie (`mpg`).
- Wypisz współczynnik wolny i kierunkowy.

- c) Narysuj wykres rozproszenia branych pod uwagę zmiennych i nanieś na niego wyestymowaną prostą.
- d) Na podstawie wykresu oceń, czy model jest dobrze dopasowany.