

Ćwiczenia 6

Zakładamy, że człowiek ma nadciśnienie tętnicze, jeżeli jego ciśnienie skurczowe przekracza 140 mmHg, a ciśnienie rozkurczowe 100 mmHg. Plik „cisnienie.csv” ma następującą strukturę kolumn:

- 1) ciśnienie skurczowe
- 2) ciśnienie rozkurczowe
- 3) płeć (1 – mężczyzna, 0 – kobieta)
- 4) nadużywanie papierosów (1 – tak, 0 – nie)
- 5) nadużywanie alkoholu (1 – tak, 0 – nie)
- 6) wiek

Zadania:

- 1) Zaproponować model regresji liniowej dla zmiennych zależnych „ciśnienie skurczowe” i „ciśnienie rozkurczowe”.
- 2) Wybrać model z największą wartością współczynnika R^2 .
- 3) Co to jest kryterium AIC?
- 4) Zaproponować model z „najlepszą” wartością kryterium AIC. (Jakie wartości dla tego kryterium są najlepsze?).
- 5) Wybrać zmienne objaśniające do modelu przy pomocy regresji krokowej (zarówno postępującej i wstecznej) – funkcja w **R**: `stepAIC()` z biblioteki MASS
- 6) Przeprowadzić testy istotności dla współczynników regresji.
- 7) Sprawdzić, czy reszty mają rozkład normalny – funkcja w **R**: `shapiro.test()`.
- 8) Wyznaczyć przedziały ufności dla wszystkich estymowanych współczynników regresji na poziomie istotności 0.05 – funkcja w **R**: `confint()`
- 9) Utworzyć nową kolumnę danych o nazwie „nadciśnienie” w taki sposób że, człowiek mający problemy z ciśnieniem skurczowym lub rozkurczowym będzie przyjmował wartość 1, a człowiek nie mający nadciśnienia wartość 0.
- 10) Utworzyć model regresji logistycznej dla zmiennej „nadciśnienie” i sprawdzić czy bardziej prawdopodobne jest że nadciśnienie ma kobieta nadużywająca alkoholu i papierosów czy mężczyzna nadużywający obydwu używek.