

Ćwiczenia nr 8

Plik *piwo.csv* zawiera dane dotyczące 35 marek piwa występujących na rynku amerykańskim. Wśród zmiennych wyróżniamy:

- Zmienną *marka* służącą do opisu obserwacji.
- Zmienną *kalorie*, która jest zmienną celu.
- Zmienną *alkohol*, która jest predyktorem.

Wykonaj następujące polecenia:

- Wczytaj plik *piwo.csv* do R.
- Wykonaj wykres rozrzutu przedstawiający zależność kaloryczności piwa od zawartości alkoholu. Czy jest to zależność liniowa? Wyznacz współczynnik korelacji tych zmiennych.
- Zbuduj model liniowy do danych metodą regresji liniowej. Wykonaj histogram i wykres normalny dla reszt. Wykonaj test Shapiro Wilka i Durбина-Watsona dla reszt. Na wykresie rozrzutu przedstaw zależność reszt standaryzowanych względem standaryzowanych wartości przewidywanych. Czy założenia analizy regresji są spełnione?
- Zwróć uwagę na występowanie punktów odstających. Czy mogą one wpływać na parametry i dopasowanie modelu?
- Utwórz nową zmienną o wartościach 1, jeśli piwo jest light, i 0, jeśli nie.
- Zbuduj model raz jeszcze, tym razem osobno dla piw light i tych, które nie są light. Porównaj parametry i dopasowanie otrzymanych modeli z początkowym. Postaraj się wyjaśnić różnice. Czy założenia są teraz spełnione?
- * Następnie utwórz nowy plik danych, w który wpiszesz kilka „własnych” piw z ich zawartością alkoholu. Zastosuj zbudowany model do przewidywania ich kaloryczności.