

SAD - Sprawozdanie I

24 marca 2019

Sprawozdanie należy wykonać w oparciu o jeden z następujących zbiorów danych:

- `globalBreastCanserRisk2002.csv` - dane, które składają się z 11 zmiennych i dotyczą liczby nowych rozpoznań raka piersi w poszczególnych krajach. Zmienna objaśniana - **NCOBC_2002** (liczba rozpoznań raka piersi). Analizowane zmienne objaśniające - Country, Continent, LifeExp (oczekiwana długość życia), Pop (populacja), gdpPer-cap, AlcoholComsumption, BloodPressure, BodyMassIndex, Cholestrol, Smoking (wartości średnie).
- `HOMA.csv` - dane, które składają się z 10 zmiennych. Opisują 194 kobiety i 84 mężczyzn w wieku od 20 do 40 lat o prawidłowej masie ciała, czyli o indeksie masy ciała [BMI] pomiędzy 18 a 25. Zmienna objaśniana - **HOMA** (wartość liczona na podstawie pomiaru insuliny i glukozy na czczo). Analizowane zmienne objaśniające - Plec, Wiek, BMI, WHR (stosunek obwodu talii do obwodu bioder), Trojglicerydy, LDL/HDL (stosunek złego cholesterolu do dobrego), FAT ALL P, FAT A P, FAT G P (wartości procentowe tłuszczu odpowiednio całkowitego, androidalnego (w okolicy talii), gynoidalnego (w okolicy bioder)) .

Sprawozdanie powinno zawierać:

1. Sprawdzenie danych pod kątem obserwacji o błędnych wartościach (niezgodnych z istotą zmiennej, np. ujemna populacja) oraz wartości brakujących. Obsługa takich przypadków.
2. Wstępną analizę danych: podstawowe statystyki (np. średnia, mediana, odchylenie standardowe, korelacja) oraz wykresy (np. rozrzutu, box-plot),
3. Wylosowanie 10 obserwacji, które nie wezmą udziału w dopasowywaniu modelu, a dla których zostanie wykonana predykcja punktowa i przedziałowa przy użyciu ostatecznie wybranego modelu. (Należy pamiętać, aby przed losowaniem ustawić wartość ziarna - *set.seed.*)
4. Zbudowanie modelu regresji w oparciu o wszystkie zmienne objaśniające,

5. Analizę reszt dla pełnego modelu,
6. Identyfikację obserwacji odstających i wpływowych dla pełnego modelu, a następnie usunięcie ich ze zbioru danych,
7. Wybór modelu przy użyciu metody krokowej w oparciu o test istotności współczynnika oraz o wybrane kryterium (AIC, BIC, itp.),
8. Interpretację współczynników wybranego modelu,
9. Przeprowadzenie ponownej analizy reszt dla wybranego modelu,
10. Wnioski oraz odpowiednie komentarze do wyników.