

Kolokwium ze Statystycznych modeli liniowych i nieliniowych

Inżynieria i analiza danych III rok, grupa 1, sem. zimowy 2024/25

2025-01-23

Uwaga. Ten plik zawiera dwie strony.

ZADANIE 1. Zbiór danych *datasets::iris* zawiera pomiary w centymetrach długości i szerokości działki kielicha (ang. sepal) oraz długości i szerokości płatk (ang. petal) 50 kwiatów trzech odmian gatunku irys. Zapoznaj się z opisem danych zamieszczonym w oknie Help programu R-Studio, a następnie wykonaj poniższe polecenia i odpowiedz na zadane pytania.

(A) Zbuduj model wyjaśniający zmienność szerokości działki kielicha (Sepal.Width) od długości i szerokości płatk (Petal.Length, Petal.Width) rozważanego gatunku kwiatów. (Nie sprawdzaj założeń.) (2 p.)

(B) Wyznacz przedziały ufności dla wszystkich współczynników modelu (A). (4 p.)

(C) Za pomocą odpowiedniego testu zbadaj, czy długość działki kielicha zależy od gatunku (Species) kwiatów irys. (Pamiętaj o sprawdzeniu założeń.) (6 p.)

(D) Wykorzystując kontrasty typu *contr.sum* zbadaj, czy długość działki kielicha zależy od gatunku kwiatów irys. (8 p.)

ZADANIE 2. Zbiór danych *pima* biblioteki *faraway* zawiera wyniki badania przeprowadzonego na 768 dorosłych Indiankach Pima mieszkających w pobliżu Phoenix przez Narodowy Instytut Cukrzycy oraz Chorób Trawiennych i Nerek. Zbiór ten zawiera następujące zmienne

pregnant – liczba przebytych ciąż,

glucose – stężenie glukozy w osoczu po 2 godzinach w doustnym teście tolerancji glukozy,

diastolic – ciśnienie rozkurczowe krwi (mm Hg),

triceps – grubość fałdu skórno-tricepsu (mm),

insulin – ilość insuliny w surowicy (mu U/ml),

bmi – wskaźnik masy ciała (waga w kg/(wzrost w metrach do kwadratu)),

diabetes – wskaźnik wystąpienia cukrzycy,

age – wiek w latach,

test – wynik testu sprawdzającego, czy u pacjentki występują objawy cukrzycy (0 w przypadku wyniku negatywnego, 1 w przypadku wyniku pozytywnego).

(A) Na podstawie zbioru *pima* zbuduj model za pomocą którego możemy wyjaśnić szanse na to, że losowo wybrana Indianka Pima jest chora na cukrzycę. Do modelu włącz wszystkie zmienne. Na podstawie zbudowanego modelu określ jaki jest wpływ stężenie glukozy w osoczu (glucose) oraz wskaźnika masy ciała (bmi) na to czy dana Indianka Pima jest chora na cukrzycę. (8 p.)

(B) Na podstawie zbioru *pima* zbuduj model liniowy za pomocą którego możemy wyjaśnić wpływ stężenia glukozy we krwi (glucose) oraz ilości insuliny w surowicy (insulin) na ciśnienie rozkurczowe krwi (diastolic). Zbadaj czy w zbudowanym modelu zostały spełnione założenia braku autokorelacji i jednorodności błędów. (Użyj wszystkich poznanych testów na wykładzie.) (7 p.)

(C) Zmodyfikuj model otrzymany w punkcie (B) logarytmując wszystkie zmienne. Podaj interpretacje współczynników w tak otrzymanym modelu. (5 p.)

Rozwiązania wszystkich zadań zamieść w jednym pliku Rmd. Plikowi nadaj nazwę, którą jest twoje **nazwisko bez znaków diakrytycznych**. Tak nazwany plik prześlij jako odpowiedź do zadania na Teamsach. Pamiętaj, aby do każdego podpunktu podać odpowiedź.