

## Uogólnione modele liniowe

### Laboratorium nr 11

- 11.1 (Modele logliniowe dla tablicy trójdzielczej) Dane zawarte w pliku `uzywki.txt` odnoszą się do badania przeprowadzonego w 1992 roku przez Wright State University School of Medicine oraz United Health Services z Dayton, Ohio (USA). Przepytano 2276 uczniów ostatnich klas liceów zlokalizowanych w podmiejskich okolicach Dayton – czy kiedykolwiek stosowali oni używki, tzn. alkohol (w oznaczeniach modeli - A), papierosy (C), marihuana (M).
- (a) Wyliczyć wartości dopasowane dla następujących modeli logliniowych: (A,C,M), (AC,M), (AM,CM) i modelu ze wszystkimi interakcjami drugiego rzędu.
  - (b) Wyestymować stosunki szans opisujące warunkowe stowarzyszenie zmiennych występujących we wszystkich parach (AC, AM, CM) dla modeli z powyższego punktu.
  - (c) Wyestymować stosunki szans (bezwarunkowe) dla zmiennych występujących we wszystkich parach dla modeli z punktu pierwszego.
  - (d) Obliczyć i porównać wartości dewiancji i statystyki  $X^2$  Pearsona dla modeli z (a).
  - (e) Spośród badanych modeli wybrać najlepszy (na podstawie formalnego testu dopasowania). Który model jest sugerowany przez kryterium AIC?
  - (f) Przetestować warunkową niezależność pomiędzy używaniem alkoholu a paleniem papierosów poprzez porównanie jakości dopasowania stosownych modeli.
- 11.2 (Analiza korespondencji) Zbiór `HairEyeColor` (wbudowany w R) zawiera dane 592 studentów sklasyfikowanych ze względu na ich płeć, kolor oczu i kolor włosów.
- (a) Przeprowadzić test niezależności  $\chi^2$  Pearsona dla koloru oczu i koloru włosów (dla danych zagregowanych ze względu na płeć) (jednym ze sposobów przeprowadzenia tego testu jest przekształcenie danych do tablicy kontyngencji `hair*eye` i skorzystanie z `summary(tablica_kontyngencji)`).
  - (b) Obejrzyć dwie możliwości wyświetlania danych z tablic kontyngencji:
    - wykres "dotchart": za pomocą polecenia `dotchart`,
    - wykres "mosaic plot": za pomocą polecenia `mosaicplot` (wykonać także dla tabelki `eye*hair`).
  - (c) Przekształcić zbiór `HairEyeColor` do zbioru z czterema kolumnami `y`, `sex`, `eye` i `hair`, w którym wylistowane są liczby (`y`) osób płci `sex` z każdą kombinacją kolorów oczu (`eye`) i włosów (`hair`). Dopasować do danych model Poissona `modc` :  $y \sim \text{hair} + \text{eye}$  i wywnioskować na tej podstawie zależność koloru oczu i koloru włosów.
  - (d) Zagregować dane ze względu na płeć, tzn. przekształcić zbiór `HairEyeColor` do zbioru z trzema kolumnami `y`, `eye` i `hair` (teraz `y` jest sumą odpowiednich licznosci). Do tak przekształconych danych dopasować model Poissona, z którego można wywnioskować zależność koloru oczu i koloru włosów. Przeanalizować podobieństwa i różnice w otrzymanych współczynnikach, dewiancjach itd. pomiędzy tym modelem, a modelem `modc` z poprzedniego punktu.
  - (e) Przeprowadzić analizę korespondencji zmiennych `hair` i `eye` (na podstawie modelu z punktu (2d)).