

Uogólnione modele liniowe

Laboratorium nr 8

- 8.1 Narysować symulowaną obwiednię wykresu typu half-normal dla standaryzowanych rezyduów dewiancyjnych dla danych ze zbioru **bliss**.
- 8.2 (Dane wielomianowe - wstępna analiza zbioru danych) Zbiór `miner2.data` to dane dotyczące trzech możliwych stanów (`normal`, `mild`, `severe`) pewnej choroby płuc wśród pewnej grupy górników. Zmienna `year` oznacza czas pracy w kopalni. Dane są w postaci: pierwszych 8 rekordów dotyczy statusu `normal`, następnych 8 statusu `mild` i ostatnich 8 statusu `severe`.
- (a) Utworzyć macierz liczości `Freq`, z trzema kolumnami (`normal`, `mild`, `severe`).
 - (b) Obliczyć proporcje każdego stanu choroby w każdym wierszu (dla każdego okresu zatrudnienia w kopalni).
 - (c) Wyrysować wykresy proporcji jako funkcji `year`.
 - (d) Wykres sugeruje połączenie kategorii `mild` i `severe`. Połączyć te kategorie (utworzyć zbiór z dwoma kolumnami, jedną odpowiadającą statusowi `normal` i drugą odpowiadającą `mild` i `severe`) i dopasować model logitowy. Ocenić dopasowanie modelu.
 - (e) Zmienić zmienną objaśniającą na $\log(\text{year})$ i dopasować na nowo model logitowy. Porównać p -wartości w obu dopasowanych modelach.
- 8.3 Analizować będziemy dane pochodzące z 1996 American National Election Study (zbiór **nes96** z biblioteki **faraway**). Dla uproszczenia będziemy uwzględniać jedynie wiek, poziom wykształcenia i zarobki w badanej grupie respondentów. Zmienną odpowiedzi będzie wskaźnik identyfikacji partyjnej respondenta (w zbiorze, w zmiennej `PID`, przyjmuje więcej niż trzy wartości, my jednak przekształcimy go do zakresu: `Demokraci`, `niezależni`, `Republikanie`).
- (a) Przekształcić zmienną `PID` do zmiennej zawierającej tylko trzy czynniki, poprzez połączenie kategorii: `"strDem"` i `"weakDem"` do pojedynczej kategorii `"Democrats"`, `"indDem"`, `"indind"` i `"indRep"` do pojedynczej kategorii `"Independent"`, a `"strRep"` i `"weakRep"` do pojedynczej kategorii `"Republican"`. Obliczyć częstości występowania otrzymanych trzech kategorii w zbiorze.
 - (b) Obejrzeć poziomy zmiennej `income`, a następnie zamienić zmienną `income` na zmienną numeryczną (powiedzmy `nincome`) za pomocą wektora zawierającego średnie poszczególnych zakresów:

```
inca <- c(1.5,4,6,8,9.5,10.5,11.5,12.5,13.5,14.5,16,18.5,21,23.5,27.5,32.5,37.5,42.5,47.5,55,67.5,82.5,97.5,115)
```


oraz komendy `unclass`. Wyznaczyć podstawowe statystyki otrzymanej w ten sposób zmiennej `nincome`.
 - (c) Obliczyć proporcje sympatyków każdej z trzech opcji politycznych na każdym poziomie wykształcenia. Narysować wykresy proporcji (w zależności od poziomu wykształcenia).
 - (d) Za pomocą `cut` na podstawie zmiennej `nincome` stworzyć 7 grup i przypisać im etykiety będące przybliżonymi środkami zakresów:

```
il <- c(8,26,42,58,74,90,107)
```


W każdej z grup obliczyć proporcje sympatyków każdej z trzech opcji politycznych.
 - (e) Wykonać polecenia z poprzedniego punktu dla zmiennej `Age`, z utworzeniem siedmiu grup, etykietowanych przez

```
a1 <- c(24,34,44,54,65,75,85)
```
 - (f) Za pomocą funkcji `multinom` z biblioteki **nnet** dopasować model wielomianowy (nazwać go `mmod`) ze zmiennymi objaśnianymi `Age`, `Educ` i `nincome`.
 - (g) Za pomocą `step(mmod)` wybrać zmienne, które powinny być zachowane w modelu (`step` bazuje na kryterium AIC). Jaki model wybiera `step` (nazwać go `mmodi`)?
 - (h) Dopasować model ze zmiennymi `Age` i `nincome`. Porównać jakość jego dopasowania z modelem `mmod`. Wyciągnąć wniosek na temat istotności zmiennej `Education`.
 - (i) Za pomocą `predict` obliczyć wartości prognozowane w modelu `mmodi` dla wartości zarobków z wektora `il`.
 - (j) Wykazać, że wyrazy wolne (intercept) w modelu `mmodi` modelują prawdopodobieństwa identyfikowania się z daną partią dla respondenta o zerowych zarobkach.
 - (k) Wykazać, że współczynniki kierunkowe w modelu `mmodi` reprezentują szanse logarytmiczne (log-odds) przy przejściu od kategorii bazowej (`Democrat`) do kategorii `Independent` lub `Republican` przy jednostkowej zmianie \$1000 w zarobkach. W tym celu obliczyć wartości prognozowane w modelu `mmodi` dla `nincome=0` i `nincome=1`, a następnie obliczyć stosowne logarytmy szans.