

- Zapisać plik **szablon.R** pod nazwą **nazwisko.R** (gdzie **nazwisko** to nazwisko piszącego kolokwium). Na początku pliku należy wpisać w komentarzu własne imię i nazwisko.
- Rozwiązania zadań, wraz ze swoimi autorskimi komentarzami, należy wpisywać do pliku **nazwisko.R**. Plik należy we własnym interesie często zapisywać.
- Należy pamiętać o jawnym i wyczerpującym formułowaniu odpowiedzi na zadane poniżej pytania. Wszystkie udzielane odpowiedzi powinny być poparte stosownym uzasadnieniem.
- Po zakończeniu pracy, należy wpisać do górnego paska Eksploratora Windows nazwę katalogu `\\secundus\upload\glm`, a następnie skopiować i wkleić do tego katalogu stworzony przez siebie podczas rozwiązywania zadań plik **nazwisko.R** (tylko ten plik). We własnym interesie należy poczekać na ustne potwierdzenie odbioru pliku.

1. Zbiór **z1.txt** zawiera dane dotyczące badania oceny jakości sprzętów gospodarstwa domowego pewnej marki (zmienna *qual* z wartościami 1 - *jakość niezadowalająca*, 2 - *jakość umiarkowanie zadowalająca*, i 3 - *jakość zadowalająca*). Badane osoby klasyfikowano ze względu na płeć (zmienna *sex*: kobiety - wartość 0, mężczyźni - wartość 1) oraz grupę wiekową (zmienna *age*). Wartości zmiennej *y* to liczby respondentów o danych poziomach zmiennych *sex*, *age* i *qual*.
  - (a) Narysować na jednym wykresie trzy łamane ilustrujące zależność proporcji kobiet oceniającej jakość sprzętów od wieku. Pierwsza łamana odpowiadać ma proporcji kobiet oceniających jakość jako niezadowalającą, druga - jako umiarkowanie zadowalającą, a trzecia - jako zadowalającą. Sporządzić analogiczny wykres dla mężczyzn.
  - (b) Dopasować do danych model wielomianowy (nazwijmy go *modwiel*) z predyktorami *sex* i *age*. Następnie, korzystając z uporządkowania wartości zmiennej *qual*, dopasować do danych model proporcjonalnych szans (nazwijmy go *modpropsz*) z takimi samymi predyktorami, co *modwiel*. Ocenić, porównać i skomentować jakość dopasowania modeli *modwiel* i *modpropsz*.
  - (c) Ile wynosi prawdopodobieństwo tego, że kobieta z najmłodszej grupy wiekowej uznaje jakość sprzętu za zadowalającą (podać liczby wynikające z obydwu modeli)?
  - (d) Ile wynoszą oczekiwane na podstawie obu modeli liczebności mężczyzn z najstarszej grupy wiekowej uznających jakość sprzętu za niezadowalającą?
2. Przebadano pewną grupę osób pod kątem palenia papierosów i stosowania leków obniżających ciśnienie krwi. Zaobserwowano, że wśród kobiet stosujących leki liczba palaczek wynosi 14, a niepalących 12, zaś wśród kobiet nie stosujących tych leków liczba palaczek to 7, a niepalących 25. Analogiczne wielkości dla badanych mężczyzn wyniosły kolejno 2, 8, 22 i 84. Niech *pal* oznacza zmienną klasyfikującą osoby ze względu na palenie papierosów, *lek* - ze względu na stosowanie leków, a *plec* - ze względu na płeć.
  - (a) Do otrzymanej powyżej trójdzielczej tablicy kontyngencji  $2 \times 2 \times 2$  dopasować model *pal plec/lek plec*. Zbadać jego dopasowanie i zinterpretować je odpowiednio w terminach zależności/niezależności (warunkowej/bezwarunkowej) odpowiednich zmiennych.
  - (b) Który z trzech modeli hierarchicznych z jedną interakcją drugiego rzędu ma najmniejszą dewiancję, a który najmniejszą wartość statystyki  $X^2$  Pearsona?
  - (c) Zbadać warunkową niezależność zmiennych *pal* i *plec*.
  - (d) Obliczyć i zinterpretować (brzegowy) stosunek szans dla brzegowej tablicy kontyngencji klasyfikującej osoby ze względu na płeć i stosowanie leków.
3. Plik **vis.rda** zawiera dane dotyczące pewnej grupy pacjentów. Zmienne:
  - *Visits* - liczba wizyt u lekarza,
  - *Hospital* - liczba pobyków szpitalnych,
  - *HealthStatus* - stan zdrowia (na podstawie samooceny),
  - *Chronic* - liczba chronicznych przypadłości zdrowotnych,
  - *Sex* - płeć,
  - *Edu* - wykształcenie,
  - *PrivInsur* - wskaźnik korzystania z prywatnego ubezpieczenia zdrowotnego.

Każdy z 4406 wierszy zawiera informację na temat pojedynczej osoby. Celem zadania jest modelowanie zapotrzebowania na opiekę lekarską (mierzonego liczbą wizyt u lekarza - zmienna *Visits*) w zależności od danych dostępnych w pozostałych zmiennych.

- (a) Dopasować do danych model poissonowski ze zmienną odpowiedzi *Visits* i pozostałymi zmiennymi jako predyktorami.
- (b) Zaproponować model bardziej adekwatny niż model z poprzedniego punktu. Uzasadnić swój wybór, zwracając uwagę na formalne wskaźniki jakości dopasowania, porównanie oczekiwanej na podstawie obu modeli liczebności zer w zmiennej *Visits*, ewentualną nadwyżkę bądź niedobór rozproszenia.