

Uogólnione modele liniowe

Część laboratoryjna egzaminu, 4/02/2015

- 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.

- Powodzenia!

1. Poniższa tabela opisuje pewien zbiór pacjentów uczestniczących w pewnym badaniu i zaklasyfikowanych krzyżowo względem trzech zmiennych: typu osobowości (możliwe wartości to A i B), poziomu cholesterolu (normalny albo wysoki) oraz ciśnienia krwi (normalne albo wysokie).

osobowość	cholesterol	c.normalne	c.wysokie
A	normalny	716	79
A	wysoki	207	25
B	normalny	819	67
B	wysoki	186	22

- (a) Obliczyć stosunek szans posiadania normalnego poziomu cholesterolu dla pacjentów z typami osobowości A i B.
 - (b) Przy założeniu niezależności wszystkich trzech zmiennych, ile wynosi dopasowana wartość liczby pacjentów z osobowością typu B, z normalnym poziomem cholesterolu i z normalnym ciśnieniem krwi?
 - (c) Zbadać hipotezę niezależności wszystkich trzech zmiennych.
 - (d) Zbadać hipotezę (warunkowej) niezależności poziomu cholesterolu i ciśnienia krwi przy danym typie osobowości.
 - (e) Który z modeli jest bardziej odpowiedni do przedstawionych danych: ten opisujący niezależność z punktu (c), czy też ten opisujący niezależność z punktu (d)?
2. Podczas startu każdej misji promu kosmicznego mierzono temperaturę otoczenia i sprawdzano, ile z sześciu części pewnego typu (używanych w promach) jest uszkodzonych. Dane w pliku **czesci.txt** opisują liczbę (zmienna *damage*) części wadliwych i temperaturę otoczenia mierzoną w stopniach Fahrenheita (zmienna *temp*). Można przypuszczać, że liczba uszkodzonych części opisywana jest przez rozkład dwumianowy.
 - (a) Dopasować do danych modele logitowy i probitowy i zbadać ich dopasowanie. Na jednym wykresie narysować dopasowane prawdopodobieństwa wynikające z obydwu modeli.
 - (b) Dokonać predykcji prawdopodobieństwa zniszczenia pojedynczej części przy temperaturze otoczenia 31°F w obydwu modelach.
 - (c) Zbadać istotność zmiennej *temp* w modelu logitowym.
 - (d) Wyznaczyć 95% przedział ufności dla parametru stojącego przy zmiennej *temp* w modelu logitowym.
 - (e) Jak zmieniają się szanse uszkodzenia pojedynczej części przy zmianie temperatury o 1°F?
 - (f) Przekształcając odpowiednio przedział otrzymany w punkcie (c), wyznaczyć 95% przedział ufności dla ilorazu (stosunku) szans przy zmianie temperatury otoczenia o 1°F.
 3. Zbiór **carpref.txt** zawiera dane dotyczące oceny ważności dostępności klimatyzacji i wspomagania układu kierowniczego w samochodzie (zmienna *response* z wartościami: 1 - nieważne/moło ważne, 2 - ważne, 3 - bardzo ważne) przez kobiety (wartość 0 zmiennej *sex*) i mężczyzn (wartość 1 zmiennej *sex*). Zmienna *frequency* podaje liczbę wystąpień danej wartości *response* przy ustalonych wartościach zmiennych *sex* i *age*.
 - (a) Zaproponować model opisujący, w jaki sposób płeć i wiek wpływają na ocenę ważności dostępności klimatyzacji i wspomagania układu kierowniczego. Ocenić jakość jego dopasowania.
 - (b) Na podstawie zaproponowanego przez siebie modelu oszacować prawdopodobieństwo tego, że kobieta w wieku 18-23 uznaje dostępność klimatyzacji i wspomagania za nieważny/moło ważny czynnik.