

Analiza danych jakościowych

Lista 9

1. Dla rozkładu maksimum Gumbela o dystrybuancie $F(u)=\exp(-\exp(-u))$ wyznacz lewostronną $g_L(p)$ i prawostronną $g_P(p)$ skalę kwantylową. Sprawdź, że skala ta nie jest symetryczna.

2. Uzupełnij brakujące wartości:

Prawdop.	0,1	0,95								
logit			-1	2						
probit					-1	2				
$g_L(p)$							-1	2		
$g_P(p)$									-1	2

3. Dane w tabeli [Źródło: Agresti, 106] podają liczbę owadów zniszczonych za pomocą pewnego środka owadobójczego. Dawka środka jest stężeniem w mg/kg ciała owadów, tu wyrażonym w skali logarytmicznej.

log(dawka)	1,691	1,724	1,755	1,784	1,811	1,837	1,861	1,884
n_owadów	59	60	62	56	63	59	62	60
n_zniszczonych	6	13	18	28	52	53	61	60

a) Oblicz prawdopodobieństwa zniszczenia owada dla każdej badanej dawki środka owadobójczego.

b) Wyraż te prawdopodobieństwa w każdej ze skal z zadania 2.

c) Dla każdej ze skal dopasuj metodą najmniejszych kwadratów model liniowy, opisujący relację między h_i a l_i gdzie h_i jest prawdopodobieństwem zniszczenia owada dla logarytmu dawki l_i , wyrażonym w odpowiedniej skali. Wybierz najlepszy z tych modeli. Przyjmując, że te parametry są bliskie parametrom uzyskanym metodą największej wiarygodności, zweryfikuj hipotezę, że wybrany model pasuje do danych.

4. W toksykologii miarą toksyczności danego środka jest LD50: jest to stężenie środka, które powoduje zatrucie 50% populacji. Z najbardziej znanych trucizn arszenik ma LD50 = 20 mg/kg cyjanowodor LD50 = 1.5 mg/kg; jad skorpiona hodowlanego ma LD50=60 mg/kg a typowego skorpiona jadowitego 2-30 mg/kg.

Dla liniowych modeli, opisujących relację między h_i a l_i gdzie h_i jest prawdopodobieństwem zatrucia dla logarytmu dawki l_i , wyrażonym w skalach logitowej, probitowej, g_L , , g_P , zapisz wzór na wyznaczenie LD50. Wyznacz LD50 dla modelu z zadania 3.

5. W historii USA opisany jest, pod nazwą *Donner Party*, przypadek wędrówki dwóch rodzin amerykańskich Donnerów i Reedów.¹ W jednym z opracowań, na podstawie danych 45 osób zbudowano dwa modele logistyczne przeżycia członków tej grupy (zmienna *KOBIETA* ma wartość 1 gdy osobą była kobieta, 0 – gdy mężczyzną):

Model 1

$$Lgt(przeżycia)=0,318-0,032*wiek+6,927*KOBIETA-0,162*wiek*KOBIETA$$

Statystyka $G^2=47,346$

Model 2

$$Lgt(przeżycia)=1,633-0,078*wiek+1,597*KOBIETA$$

Statystyka $G^2=51,256$

Dla obu modeli:

a) Podaj interpretację współczynników równania logistycznego przy pomocy ilorazów krzyżowych

b) Jaki jest iloraz krzyżowy dla przeżycia osób 50-letnich i 20 letnich?

c) Jakie jest prawdopodobieństwo przeżycia kobiet a jakie mężczyzn 50-letnich i 20-letnich?

d) W jakim wieku prawdopodobieństwo przeżycia wynosi 50% dla kobiet a w jakim dla mężczyzn?

e) Jaka jest wartość p dla każdego z modeli w teście zgodności z danymi?

f) Korzystając z twierdzenia Pitagorasa przeprowadź test czy modele 1 i 2 są od siebie istotnie różne

¹ 87 ludzi wyruszyło 20 wozami konnymi w lipcu 1846 z Illinois do Kalifornii. Do celu dotarło 47 spośród nich (ostatniego członka tych rodzin uratowano w kwietniu 1847).