Elementy statystyki - DEST LIO

Test przykładowy

Zadanie 1. Zmienna `Wynik` w pliku `Z1.txt` podaje wyniki odpowiedzi na pytanie: Jak oceniasz sytuację materialną swojej rodziny? Dane pochodzą z 1994 roku. Odpowiedzi zakodowano w następujący sposób: 1 – bardzo dobra, 2 - przeciętna, 3 - zła, 4 - fatalna. Rozkład odpowiedzi dla dziewcząt urodzonych w 1980 roku przedstawia tabela:

(A)

Ocena sytuacji materialnej	% odpowiedzi
bardzo dobra	$14,09$
przeciętna	$79,08$
zła	$6,68$
fatalna	$0,15$

(B)

Ocena sytuacji materialnej	Liczebność
bardzo dobra	83
przeciętna	450
zła	35
fatalna	1

(C)

Ocena sytuacji materialnej	% odpowiedzi
bardzo dobra	$8,55$
przeciętna	$81,20$
zła	$9,40$
fatalna	$0,85$

(D)

Ocena sytuacji materialnej	Liczebność
bardzo dobra	11
przeciętna	95
zła	10
fatalna	1

Zadanie 2. Niech $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$ będzie próbą prostą z populacji o rozkładzie opisanym funkcją gęstości f_θ, gdzie $\theta \in \Theta$ jest parametrem. Wtedy funkcja wiarogodności wyraża się wzorem:

(A) $L(\theta; \mathbf{x}) = \sum_{i=1}^{n} f_\theta(x_i)$

(B) $L(\theta; \mathbf{x}) = \prod_{i=1}^{n} f_\theta(x_i)$

(C) $L(\mathbf{x}; \theta) = \sum_{i=1}^{n} f_\theta(x_i)$

(D) $L(\mathbf{x}; \theta) = \prod_{i=1}^{n} f_\theta(x_i)$

Zadanie 3. Porównywano długości czasów świecenia dwóch typów żarówek (w h). Wyniki zawarto w pliku `Z3.RData`. Na poziomie istotności 0,05 zweryfikowano hipotezę zerową, że wartości przeciętne czasów świecenia żarówek tych typów nie różnią się istotnie (przyjmując założenia o normalności rozkładów badanej cechy i jednorodności wariancji). Prawdziwe jest stwierdzenie:

(A) wartość statystyki testowej wynosi $-0,1461$, p-wartość wynosi $0,8866$,
(B) p-wartość wynosi $0,8833$, zatem odrzucamy hipotezę zerową,
(C) p-wartość wynosi $0,8833$, zatem nie ma podstaw do odrzucenia hipotezy zerowej,
(D) p-wartość wynosi $0,8866$, zatem przyjmujemy hipotezę zerową.

Zadanie 4. W teorii weryfikacji hipotez statystycznych:

(A) jeżeli p-wartość jest większa od wartości statystyki testowej, to nie ma podstaw do odrzucenia hipotezy zerowej,
(B) jeżeli p-wartość jest mniejsza od poziomu istotności testu, to odrzucamy hipotezę zerową,
(C) p-wartość jest największym poziomem istotności testu, przy którym odrzucamy hipotezę zerową,
(D) jeżeli p-wartość jest większa od wartości krytycznej, to nie ma podstaw do odrzucenia hipotezy zerowej.

Zadanie 5. Dane w pliku `Z5.RData` opisują 5 cech fizycznych 12 minerałów z grupy kryształów jednoosiowych. Podzielono analizowane obiekty na cztery skupienia. Zastosowano metodę hierarchiczną przyjmując jako miarę niepodobieństwa obiektów odległość euklidesową oraz jako metodę aglomeracji metodę średnich połączeń. Skupienia składają się z elementów:

(A)

Skupienie 1:	$1, 2, 5, 11$
Skupienie 2:	3
Skupienie 3:	$8, 9, 10, 12$
Skupienie 4:	$4, 6, 7$

(B)

Skupienie 1:	$1, 3, 5, 11$
Skupienie 2:	2
Skupienie 3:	$8, 9, 10, 12$
Skupienie 4:	$4, 6, 7$

(C)

Skupienie 1:	$1, 5, 11$
Skupienie 2:	2
Skupienie 3:	$8, 9, 10, 12$
Skupienie 4:	$3, 4, 6, 7$

(D)

Skupienie 1:	$1, 5, 11$
Skupienie 2:	2
Skupienie 3:	$3, 8, 9, 10, 12$
Skupienie 4:	$4, 6, 7$

Zadanie 6. Badano zależność pomiędzy miesięcznymi dochodami w rodzinie w przeliczeniu na jedna osobę, a miesięczną wartością wydatków konsumpcyjnych w rodzinie w przeliczeniu na jedna osobę. Wyniki zawarto w pliku `Z6.RData`. Przyjęto liniowy model zależności pomiędzy badanymi dochodami i wydatkami. Następnie oszacowano współczynniki prostej regresji oraz wyznaczono prognozowaną miesięczną wartość wydatków konsumpcyjnych w rodzinie w przeliczeniu na jedna osobę, przy miesięcznym dochodzie równym 350 złotych na jedna osobę. 95% przedział ufności dla tej prognozy ma postać:

....

Zadanie 7. Napisz funkcję realizującą test istotności dla wariancji w modelu normalnym. Zwracany wynik ma być klasy `htest`. **Wskazówka:** Hipoteza zerowa ma postać $H_0 : \sigma^2 = \sigma_0^2$. Statystyka testowa

$$T = \frac{nS^2}{\sigma_0^2}$$

ma przy prawdziwości hipotezy H_0 rozkład $\chi^2(n-1)$, gdzie n jest liczbą obserwacji a

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

wariancją z próby. Hipotezy alternatywne wraz z odpowiadającymi im obszarami krytycznymi zostały podane w Tabeli 1, gdzie $\alpha \in (0, 1)$ jest zadanym poziomem istotności.

Hipoteza alternatywna	Obszar krytyczny
$H_1 : \sigma^2 > \sigma_0^2$	$B = \{\mathbf{x} : T(\mathbf{x}) \geqslant \chi^2(1-\alpha, n-1)\}$
$H_1 : \sigma^2 < \sigma_0^2$	$B = \{\mathbf{x} : T(\mathbf{x}) \leqslant \chi^2(\alpha, n-1)\}$
$H_1 : \sigma^2 \neq \sigma_0^2$	$B = \{\mathbf{x} : T(\mathbf{x}) \geqslant \chi^2(1-\alpha/2, n-1)$ lub $T(\mathbf{x}) \leqslant \chi^2(\alpha/2, n-1)\}$

Tablica 1. Test istotności dla wariancji w modelu normalnym.

Zadanie 8. W pliku `Z8.RData` zawarto dane z roku 2010, dotyczące ochrony środowiska (emisja zanieczyszczeń, odprowadzanie ścieków, itp.) dla 16 województw. Wykorzystując zebrane dane, przeprowadzono analizę składowych głównych. Dwie pierwsze składowe główne wyjaśniają:

......................% zmienności układu

Odpowiedzi:

1. (C)
2. (B)
3. (C)
4. (B)
5. (D)
6. $(211{,}5352; 334{,}1937)$
7. –
8. 91,2%