

Programowanie narzędzi analitycznych - przykładowa kartkówka - Z05

Zadanie 1

Przeprowadzić eksperyment składający się z 10 tysięcy powtórzeń. W każdym powtórzeniu wygenerować zmienną o 100 obserwacjach z rozkładu t-Studenta z 5 stopniami swobody oraz przeprowadzić test Kołmogorowa-Smirnowa zgodności z rozkładem t-Studenta z 5 stopniami swobody. W każdym powtórzeniu zapisujemy wynik (statystykę testową lub p-value). Sprawdzić w jakim procencie przypadków test wskazuje prawidłowy wynik. **Podpowiedź:** polecenie `ks.test(zmienna, 'pt', df=5)` wykonuje test Kołmogorowa-Smirnowa na zgodność z rozkładem t-Studenta z 5 stopniami swobody.

Zadanie 2

Za pomocą polecenia `curve` narysować wykres funkcji gęstości rozkładu:

- Gamma z parametrami 2 i 3,
- wykładniczego z parametrem 2.

Zadanie 3

Przeprowadzić eksperyment składający się z 15 tysięcy powtórzeń wygenerowania zmiennej o 200 obserwacjach z rozkładu Gamma z parametrami 2 i 3 i przeprowadzeniu Kołmogorowa-Smirnowa na zgodność z rozkładem wykładniczym z parametrem 2. W każdym powtórzeniu zapisujemy wynik (statystykę testową lub p-value). Sprawdzić w jakim procencie przypadków test wskazuje prawidłowy wynik.

Zadanie 4

Wygenerować zmienną `wek`, która ma 10000 obserwacji, które pochodzą z rozkładu χ^2_2 . Zapisać zmienną `wek` do pliku `wek.csv` poleceniem `write.csv`.

Zadanie 5

Statystyka testowa w pewnym teście ma rozkład χ^2_2 . Wyznaczyć dolną i górną granicę 95% przedziału ufności.

Zadanie 6

Zbiór danych `beef.csv` zawiera informacje o konsumpcji i cenach mięsa oraz dochodu rozporządzalnego w Australii w latach 1949-1965. Bez wykorzystywania funkcji zewnętrznych napisać kod, który pozwoli oszacować model regresji liniowej dla spożycia wołowiny per capita (QB). Zmiennymi objaśniającymi mają być: cena wołowiny (PB), cena baraniny (PL), cena wieprzowiny (PP) oraz dochód rozporządzalny per capita (INC). Postać modelu:

$$\ln(QB) = \beta_0 + \beta_1 \ln(PB) + \beta_2 \ln(PL) + \beta_3 \ln(PP) + \beta_4 \ln(INC) + \varepsilon. \quad (1)$$

Sprawdzić czy oszacowanie parametru β_1 jest istotnie różne od zera. Policzyc wartość współczynnika determinacji liniowej R^2 .

Zadanie 7

Rozważmy zagadnienie relacji między godzinową stawką a liczbą przepracowanych godzin.

Życiowy przykład (za The Radical Statistician by Jim Higgins, Ed. D.): obserwujemy płace za 1 h pracy i ciekawi nas ile godzin pracy jest wykonywane za tę stawkę w tygodniu. Dla obserwacji (5\$, 50h), (10\$, 40h), ..., (10000\$, 2h) zauważyć można ujemną korelację między płacą a liczbą przepracowanych godzin.

Dla wygody obliczeniowej rozważmy zbiór punktów

$$(stawka, placa) = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}. \quad (2)$$

Znaleźć najlepiej dopasowaną do nich funkcję $y = ax$. Policzyc R^2 według wzoru $R^2 = 1 - \frac{RSS}{TSS}$. Dlaczego w tym przypadku nie można stosować tego wzoru?