Rachunek prawdopodobieństwa i statystyka

Lista zadań nr 10a. 11. maja 2023

Zadania

1. Niech $X \sim \text{Geom}(p)$. Udowodnić, że $P(X > k+l) = P(X > l) \cdot P(X > k)$.
2. Niech $X \sim \text{Exp}(\lambda)$. Udowodnić, że $P(X > s+t) = P(X > s) \cdot P(X > t)$.
3. Zmienne losowa X to czas pracy jakiegoś urządzenia (do momentu awarii, t to miesiące). $P(X \geqslant t) = \exp(-0.1t)$. Jakie jest ppb, że urządzenie zepsuje się w czwartym (piątym) roku? (Odp: 0.095)

   [Zadania 4–5] W pliku klimat.csv znajduje się: szerokość i długość geograficzna, roczna suma opadów (mm), średnia temperatura roczna (°C) i wysokość nad poziomem morza miast wojewódzkich. Po rozwiązaniu omówić wyniki zadań.

4. Wyznaczyć prostą regresji temperatury względem wysokości npm.
5. Wyznaczyć prostą regresji temperatury względem długości i szerokości. (Z zależy od X oraz od Y).
6. Zmienna losowa X ma dyskretny rozkład jednostajny

$$P(X = i) = \frac{1}{100}, \quad i \in \{1, 2, \ldots, 99, 100\}.$$

   Zmienne losowe Y oraz Z określone są następująco

$$Y = \begin{cases} 1, & 2|X \vee 3|X, \\ 0, & \text{wpw}, \end{cases} \qquad Z = \begin{cases} 1, & 3|X, \\ 0, & \text{wpw}. \end{cases}$$

   Znaleźć wartość współczynnika korelacji ρ zmiennych Y i Z. (Odp.: $\rho = {}^{33}/_{67}$)

7. Dane są punkty $(x_1, y_1), \ldots, (x_n, y_n)$. Szukamy krzywej regresji w postaci $y = a + bx + cx^2$. Uzasadnić, że parametry a, b, c są rozwiązaniem układu równań:

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{bmatrix}.$$

8. Dane są punkty $(x_1, y_1, z_1), \ldots, (x_n, y_n, z_n)$. Szukamy równania regresji w postaci $z = a + bx + cy$. Uzasadnić, że parametry a, b, c są rozwiązaniem układu równań:

$$\begin{bmatrix} n & \sum x_i & \sum y_i \\ \sum x_i & \sum x_i^2 & \sum x_i y_i \\ \sum y_i & \sum x_i y_i & \sum y_i^2 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum z_i \\ \sum x_i z_i \\ \sum y_i z_i \end{bmatrix}.$$

9. Dla danych z pliku rp10-09a.csv
   (a) wyznaczyć przedział ufności dla wartości średniej,
   (b) obliczyć wartość_p H_0: $\mu = 2840$, następnie H_0: $\mu = 2850$ i w końcu H_0: $\mu = 2875$.

10. Dane (w kolumnach) przedstawiają pomiar wagi przed i po okresie stosowania określonej diety dla 16 osób. Testujemy hipotezę: **dieta ma wpływ na wagę**.
11. **(2p.)** 10 poletek doświadczalnych podzielono na dwie części, w jednej z nich przeprowadzono dodatkowe czynności *agrotechniczne*. W wierszu znajduje się wydajność części poddanej dodatkowym zabiegom i części poletka uprawianej tradycyjnie. Testujemy hipotezę: **dodatkowy czynnik ma wpływ na wydajność uprawy**, tzn. podać postać hipotezy zerowej i podać wartość_p.
12. Niezależne obserwacje $x_1, \ldots, x_n,\ \ y_1, \ldots, y_k$ pochodzą z rozkładu $N(\mu, \sigma^2)$. Znaleźć rozkład zmiennej Z:
$$Z = \frac{\bar{X} - \bar{Y}}{\sigma\sqrt{\frac{n+k}{nk}}}.$$

Witold Karczewski