Rachunek prawdopodobieństwa i statystyka

Lista zadań nr 11. 18 i 19 maja 2017

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

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

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

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

3. Znaleźć prostą regresji Y względem X dla poniższych danych:

4. Poniższa tabela zawiera dane dotyczące ciśnienia P i objętości V pewnej stałej masy gazu. Równanie łączące te dwie wielkości ma postać PV^k = C, gdzie k, C są pewnymi stałymi.

Stosując regresję liniową znaleźć wartości C oraz k. Jaka jest przewidywana wartość P dla V=100?

5. Dla poniższych danych znaleźć równanie regresji kwadratowej:

6. Znaleźć równanie regresji liniowej zmiennej Z względem zmiennych X,Y.

- 7. Dana jest prosta regresji y = a + bx. Niech $\hat{y}_k = a + bx_k$ (wartość przewidywana). Definiujemy błąd przybliżenia jako $s_{y.x}^2 = \frac{\sum (y_k \hat{y}_k)^2}{n}$. Wykazać, że $s_{y.x}^2 = \frac{\sum y_k^2 a \sum y_k b \sum x_k y_k}{n}$.
- 8. Udowodnić, że $\sum (y_k \bar{y})^2 = \sum (y_k \hat{y}_k)^2 + \sum (\hat{y}_k \bar{y})^2$. Słownie: zmienność całkowita = zmienność losowa + zmienność regresji.
- 9. Dla poniższych danych obliczyć zmienność całkowitą, zmienność losową i zmienność regresji.

$$x_k$$
 - ojciec | 165 | 160 | 170 | 163 | 173 | 157 | 178 | 168 | 173 | 170 | 175 | 180 | 173 | syn | 173 | 168 | 173 | 165 | 175 | 168 | 173 | 165 | 180 | 170 | 173 | 178 |

- 10. Znaleźć estymator największej wiarygodności parametru a, dla niezależnych obserwacji x_1, \ldots, x_n z rozkładu o gęstości $f(x) = 2a\sqrt{a/\pi} \, x^2 \mathrm{e}^{-ax^2}, -\infty < x < \infty$.
- 11. Znaleźć estymator największej wiarygodności parametru k, dla niezależnych obserwacji x_1, \ldots, x_n z rozkładu o gęstości $f(x) = (k+1) x^k$, 0 < x < 1.
- 12. **(E1)** 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