Rachunek prawdopodobieństwa i statystyka

Problem set №11. May/June 2020

1. The data includes the daily number of ads in newspapers A, B and C. We test the hypothesis: **the average number of ads in three newspapers is the same**.
2. Data: daily number of entries in four departments (A, B, C, D) of the company. Tested hypothesis: **the number of entries is on average the same in all departments**.
3. Consider recording the number of earthquakes in four observation points. Tested hypothesis is of the form: **points are located within the same geological structure**.
4. Data: strength of three types of tires from five manufacturers. Tested hypotheses: **average tire strength is independent of the manufacturer** and **different tire types have different durability**.

   Ex. 5–8. Random variable $Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$ is of $N(\mu, \Sigma)$ distribution, with parameters $\mu = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 4 & 1 \\ 1 & 9 \end{bmatrix}$. Let $U = \frac{1}{2\sqrt{15}}(-3Y_1 + 2Y_2)$ and $V = \frac{1}{2\sqrt{21}}(3Y_1 + 2Y_2 - 12)$.
5. (2 p.) Prove that $Z_1 = U^2 + V^2$. Z_1 like in exercise 8.
6. Check that random variables U, V are independent.
7. Check that $U, V \sim N(0, 1)$.
8. What is distribution of random variable $Z_1 = (Y - \mu)^T \Sigma^{-1} (Y - \mu)$?
9. Given data in file rp11-09.csv
   (a) find confidence interval with respect to mean (expected value), confidence level $\alpha = 0.05$,
   (b) find p-value.

---

10. **(E3)** Data: factor A - normal (1) or higher level of sugar (2); factor B - normal weight (1) or above normal (2). The dependent variable is blood pressure. **is there a relationship between level of sugar and blood pressure**, **is there a relationship between weight and blood pressure**, and **is there any interaction between weight and sugar level**?
11. **(E2)** Factor A is an advertising medium, factor B is the main topic of advertising, data is a percentage of positive reactions to advertising. **Is there a relationship between the ad medium and the performance of the ad**, **which type of ad is most effective** and **is there a relationship between the theme of the ad and its media**?
12. **(E2)** Data (in columns) presents the measurement of weight before and after the period of application of a specific diet for 16 people. Perform test of the hypothesis: **diet affects weight**.
13. **(E2)** 10 experimental plots were divided into two parts, in one of them additional agrotechnical activities were performed. The row contains the yield of the part subjected to additional treatments and the part of the field traditionally grown. Test the hypothesis: **additional factor affects crop yields**, i.e. give the form of the null hypothesis and specify p-value.
14. **(E1)** Independent observations $x_1, \ldots, x_n,\ y_1, \ldots, y_k$ are from the $N(\mu, \sigma^2)$ distribution. Find distribution of random variable Z:

$$Z = \frac{\bar{X} - \bar{Y}}{\sigma \sqrt{\frac{n+k}{nk}}}.$$

Witold Karczewski