Lab FDSAI

1. In order to ensure efficient usage of a server, it is necessary to estimate the mean number of concurrent users. According to records, the average number of concurrent users at 100 randomly selected times is 37.7, with a standard deviation 9.2. Construct a 90% confidence interval for the expectation of the number of concurrent users.

2. Salaries of entry-level computer engineers have a Normal distribution with unknown mean and variance. Three randomly selected computer engineers have yearly salaries (in 1000$): 30, 50, 70

   (a) Construct a 90% confidence interval for the average salary of an entry-level computer engineer.

   (b) Looking at this sample, one may think that the starting salaries have a great deal of variability. Construct a 90% confidence interval for the standard deviation of entry-level salaries.

   (c) A new random sample of 100 entry-level computer engineers gives the same sample mean and sample standard deviation for the salary as the smaller sample above. Construct 90% confidence intervals for the average salary and for the standard deviation of the salary. Comment on the results.

3. A pharmaceutical company is looking to purchase a bottling machine to increase efficiency in the production of cough syrup. Currently, the company uses a robotic system that can fill the corresponding containers with the fixed amount of cough syrup, with a standard deviation of 1.5 ml. The machine under consideration for purchase was subjected to 30 tests, for which a sample variance of 2.5 ml^2/bottle was observed. At a significance level of 95%, is the purchase justified?

4. In this exercise, we will reuse the NHANES dataset from Lab 1. The variables BMXHT and BMXWT, which we will examine, refer to the height (cm) and body weight (kg) of surveyed subjects. We will also consider their gender, recorded in the binary variable RIAGENDR (coded as: 1 = male, 2 = female).

   (a) (Lab 1) Extract only the information about gender, height, and body weight from the dataset. Check if the new dataset contains missing entries (NaN) and remove them if necessary.

   (b) Determine a 95% confidence interval for the mean height and body weight, both separately by gender and for the entire sample. Repeat for a 99% confidence level. What do you observe?

   (c) Now assume the entire statistical population is represented by the NHANES dataset. Randomly extract 20 samples of size 100 from it. For each sample, determine a 95% confidence interval for the mean height and mean body weight, both separately by gender and for the entire sample. Plot the resulting confidence intervals within the same graph and highlight the true mean value of the characteristic obtained in the previous step for the entire population. Comment on the results.

5. Let $X_1, \ldots, X_n$ be a simple random sample from a normal distribution with unknown mean μ and variance σ^2. Consider estimators of σ^2 of the form

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

   and find the value of k that minimizes the mean square error (MSE).