TASK LIST NO. 6: Basic Statistics and Their Distributions

Task 1

A sample of $n = 50$ elements was taken from a general population (e.g., server response times in ms). Raw results were obtained: $3.6, 5.0, 4.0, 4.7...$ (full data in the set).

Prepare a **frequency distribution** (frequency table) for the given sample, assuming the number of classes $k = 7$.

Task 2

For an ordered sample (e.g., number of code errors in subsequent modules): $3.0, 3.1, 3.3, 3.4, ..., 6.4$ (a total of 50 results).

Determine:

a) **Median** (m_e) – the middle value (resistant to extreme values, so-called *outliers*).
b) **Mode** (m_o) – the most frequent value.

Task 3

In a certain chemical experiment (or processor production process), the amount of pure substance was investigated. For 5 measurements, the following results were obtained: $3.5, 3.4, 2.1, 5.4, 1.1$.

Calculate:

a) The arithmetic mean of the sample $\bar{x}$.
b) The sample variance s^2 (using the formula for a small sample).
c) The standard deviation s.

Task 4

A vehicle (or a data packet in a network) traveled a path consisting of three sections of the same length but with different speeds: $v_1 = 50, v_2 = 60, v_3 = 70$ km/h. Calculate the average speed over the entire route.

Hint: Use the harmonic mean, not the arithmetic mean.

Task 5

Two six-element samples are given (e.g., access times to two different disks):

- Sample I: $80, 40, 40, 80, 40, 80$
- Sample II: $40, 80, 120, 80, 120, 40$

Calculate the coefficients of variation $v = \frac{s}{\bar{x}}$ for both samples. Which disk operates more stably (has a smaller relative dispersion)?

Task 6

Find the confidence interval (or probability), knowing that the investigated feature X of the population has a normal distribution $N(\mu, \sigma)$. The statistic U:

$$U = \frac{\bar{X} - \mu}{\sigma}\sqrt{n}$$

has a distribution $N(0,1)$.

Task 7

A small sample ($n = 10$) was drawn from a population with a normal distribution. Since we do not know the population standard deviation σ, we must use the sample deviation s. The statistic:

$$t = \frac{\bar{X} - \mu}{s}\sqrt{n-1}$$

follows the **Student's t-distribution**. Read from the tables the critical value for $n - 1$ degrees of freedom and a confidence level of 0.95.

Task 8

To investigate the sample variance (dispersion), the following statistic is used:

$$\chi^2 = \frac{nS^2}{\sigma^2}$$

> *Note: The symbol S^2 denotes the sample variance calculated with the formula dividing by n (according to the definition in Krysicki's textbook).*

which has a chi-square distribution. 15 measurements were taken ($n = 15$). Read from the tables the values between which this statistic will fall with a probability of 0.90.

Task 9

Show on a simple numerical example that the sample mean $\bar{X}$ is an **unbiased estimator** of the population mean (i.e., $E(\bar{X}) = \mu$), whereas the sample variance (divided by n) is biased (which is why we divide by $n - 1$).

Task 10

For a small sample: $0.18, 0.56, 0.87, 1.37, 2.46$, determine the values of the empirical distribution function $S_n(x)$.