

---

## Homework assignment

---

- The solution of these assignments is a part of the final examination. It will be counted towards the final grade.
- You are required to send your commented **R** code together with a pdf of a report at least **48 hours before the start of the exam** by e-mail to both `nagy@karlin.mff.cuni.cz` and `bocinec@karlin.mff.cuni.cz`. Scanned hand-written reports are fine, but must be **readable**.
- The language of the homework reports can be either **English** or **Czech/Slovak**.
- **Plagiarism** will result in a failed exam attempt. Further consequences may apply depending on the severity of the plagiarism.
- Before the resampling procedure in each part of your solution, set `set.seed(AAA)`, where **AAA** stands for the number of your student identity card. Include this number at the beginning of your solution of the assignment. From your report, it must be clear how the resampling is done for each particular task.
- In your report, it is expected that you provide a brief description of the method used, its assumptions, and comment on the results. You should also provide a numerical value of the estimators, confidence intervals, test statistics... Do not also forget to **specify the assumed model** and give the formulas so that it is clear how the result is calculated.
- Use 5% as the level (prescribed probability of type I error) of the tests and 95% as the coverage of the confidence intervals.

### Task

Load the dataset `diabetes_patients.csv` (available [here](#)).

```
DATA = read.csv("diabetes_patients.csv")
```

Use the diabetes dataset containing several covariates measured on diabetes patients. In our analysis, we will focus on three variables — **Glucose**, **Insulin**, and **BMI**. Note that for these variables, missing data are indicated by zeros.

- (i) First, we focus on the variable **Glucose**, aiming to describe its dispersion to understand the variability in glucose levels among diabetes patients. Specifically, we are interested in the median absolute deviation (MAD). Recall that the MAD of a random variable  $X$  is defined as  $\text{median}|X - \text{median}(X)|$ . Provide a point estimate of the MAD and use nonparametric bootstrap methods to find a confidence interval for the MAD of the glucose level for diabetes patients.
- (ii) Let  $u_X(\alpha)$  denote the  $\alpha$ -quantile of the **BMI** for diabetes patients. Using nonparametric bootstrap, describe and perform a test of the hypothesis  $H_0: u_X(0.5) = 35$  against  $H_1: u_X(0.5) \neq 35$ . What is your conclusion?

- (iii) Focus on the variable **Insulin** representing each patient's insulin level. We aim to test whether this variable follows a gamma distribution  $\text{Gamma}(\alpha, \beta)$  for some unknown parameters  $\alpha, \beta > 0$ . Describe and conduct a bootstrap test about this hypothesis and provide the final conclusion.
- (iv) Describe and conduct a permutation test to evaluate the null hypothesis that the distribution of insulin levels is the same for patients without obesity (**BMI** < 30) and for obese patients (**BMI**  $\geq$  30).