NMST545 – Mathematical statistics 4 (2025–26)

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 or AI-generated solutions** 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 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 Iq-en.RData (available **here**).

```
(load("Iq-en.RData"))
```

For $n = 111$ children, the dataset contains their gender (`Iq$Gender`), IQ (`Iq$iq`), and the average marks in the seventh (`Iq$mark7`) and the eighth grade (`Iq$mark8`). Use this data for the following tasks.

(i) First, we focus on the variable `iq`. Recall that the kurtosis of a random variable X is defined as $\kappa(X) = \mathbb{E}(X-\mu)^4/\sigma^4$, where $\mu = \mathbb{E}(X)$ and $\sigma^2 = \mathbb{E}(X-\mu)^2$. Provide a point estimate of the kurtosis κ of IQ in our population, and use nonparametric bootstrap methods to construct a confidence interval for κ. You may find the function kurtosis from the moments package helpful for this task.

Note that for a normally distributed random variable X, it holds that $\kappa(X) = 3$. IQ scores were originally designed to follow a normal distribution. Based on our results, can we expect this to be consistent with our data?

(ii) As mentioned in task (i), IQ scores were designed to follow a normal distribution. Therefore, we aim to test whether the variable `iq` follows a normal distribution $\mathcal{N}(\mu, \sigma^2)$ for some unknown parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Describe and perform a parametric bootstrap test for this hypothesis, and provide the final conclusion.

(iii) Describe and perform a permutation test to evaluate the null hypothesis that the distributions of average marks in the seventh (`mark7`) and eighth (`mark8`) grades are the same.

(iv) Assume the model $Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \varepsilon_i$, where

- Y_i is the average mark in the seventh grade (`mark7`) for student i,
- X_i is the IQ (`iq`) of student i,
- Z_i is an indicator variable equal to 1 if the student i is a girl (`Gender` = "girl"), and 0 otherwise (this is stored in variable `Iq$gender`),
- $\beta_0, \beta_1, \beta_2 \in \mathbb{R}$ are unknown parameters,
- ε_i is the error term.

Using a model-based bootstrap, construct confidence intervals for β_1 and β_2.

Do the bootstrap confidence intervals differ from the classical Wald confidence intervals which rely on the validity of the assumed model? These can be obtained, for example, using

```
confint(lm(Iq$mark7 ~ Iq$iq + Iq$gender)).
```