

Information regarding the exam from MV013 Statistics for Computer Science

The exam consists of two parts: data analysis and explanation of concepts and theory.

Data analysis

- Time to complete: 60 minutes.
- Open notes format.
- The aim of this part is to demonstrate the application of methods on data – selecting an appropriate method, constructing a suitable mathematical model, verifying its assumptions, conducting data analysis, and commenting on the results.
- It consists of practical tasks to solve in R.
- The provided solution will consist of your (functional) R code accompanied by comments on the model, its assumptions, and the results.
- Solution: Rmd file together with readable pdf or html file.
- Solutions lacking explanations will be disregarded.
- The use of external materials is allowed, but must be properly cited.
- The use of chatbots is prohibited.
- At least 10 points out of 30 is needed to pass this part.

Explanation of concepts and theory

- Time to complete: 60 minutes.
- Closed notes format.
- The objective of this part is to be able to explain the mathematical concepts behind the methods being used.
- You don't need to memorize the formulas, but you might be asked to explain one of them if given.
- Solution: handwritten answers on paper.
- Unreadable solutions will be disregarded.
- At least 10 points out of 30 is needed to pass this part.

Key topics to remember:

- Data preprocessing – data cleaning, dummy variables.
- One-dimensional exploratory data analysis – numerical characteristics and visualization techniques for all kind of data (numerical, nominal, ordinal).
- Multi-dimensional exploratory data analysis for numerical data – correlations and scatterplots, PCA.
- Parametric modelling – point estimates and its properties, model selection.
- Introduction to confidence intervals and hypotheses testing.
- One-sample, two-sample and k -sample problem.
- Tests about parameter p of Bernoulli distribution.
- Tests of independence for both numerical and categorical data.
- Methods for normality assessment.
- Nonparametric tests - ranks and their properties, one-sample and two-sample rank tests, rank tests of independence.
- Linear regression model - description and interpretation, statistical tests, checking model assumptions, model selection.
- Generating pseudorandom numbers.
- Monte Carlo methods, nonparametric and parametric bootstrap.