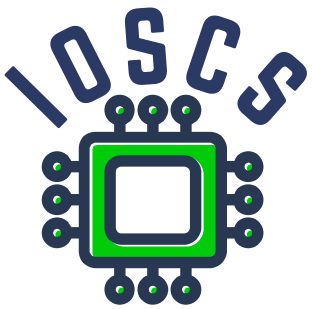
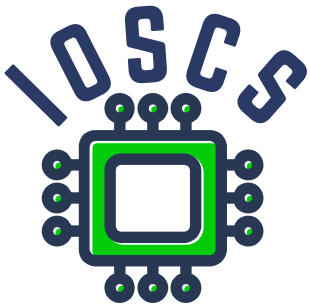


Project: Innovative Open Source Courses for Computer Science

Probability and statistics with programming in R Sylabus

Aleš Kozubík
University of Žilina

31. 1. 2020



This teaching material was written as one of the outputs of the project "Innovative Open Source Courses for Computer Science", funded by the Erasmus+ grant no. 2019-1-PL01-KA203-065564. The project is coordinated by West Pomeranian University of Technology in Szczecin (Poland) and is implemented in partnership with Mendel University in Brno (Czech Republic) and University of Žilina (Slovak Republic). The project implementation timeline is September 2019 to December 2022.

Project information

Project was implemented under the Erasmus+.
Project name: **"Innovative Open Source courses for Computer Science curriculum"**
Project nr: **2019-1-PL01-KA203-065564**
Key Action: **KA2 – Cooperation for innovation and the exchange of good practices**
Action Type: **KA203 – Strategic Partnerships for higher education**

Consortium

ZACHODNIOPOMORSKI UNIWERSYTET TECHNOLOGICZNY W SZCZECINIE
MENDELOVA UNIVERZITA V BRNĚ
ŽILINSKÁ UNIVERZITA V ŽILINE

Erasmus+ Disclaimer

Copyright Notice

Co-funded by the
Erasmus+ Programme
of the European Union

COURSE DESCRIPTION

Field of study: Computer science

Level: First cycle

Course name: Probability and statistics with programming in R

ECTS credits: 5

Instruction forms: lecture, laboratory

Instruction hours: 24, 24

Type, extent and method of teaching activities: 2 – 0 – 2 (lectures – exercises – labs) hours weekly, presence study.

Prerequisites: matematical analysis

Module/course unit objective: The student acquires basic knowledge of probability theory and mathematical statistics, which enables their use in solving technical problems in engineering practice and understanding of individual advanced statistical methods.

After completing the course the student: Learns/repeats basic concepts from probability and statistics. Acquire new knowledge in the above areas. Acquire the ability to apply the acquired knowledge in solving practical tasks. He can identify the stochastic problem. Acquire basic tools and methods for practical and theoretical solution of analytical problems using Open Source tools.

Course content divided into various forms of instruction (with number of hours):

Week	Lecture (2h per week)	Laboratory (2h per week)
1	Elements of the probability theory • Concepts random, random event, • Random variable, • The probability distribution of random variables.	Introduction into the R environment • First steps, installing R, • Basic features of R, • R as a calculator, • Algebraic operations, • Packages and repositories, • Getting help, • Quitting R.
2	The probability distribution of random variable • Probability density function, cumulative distribution function and their properties, • Numerical characteristics of the random variable, • Initial and central moments, • The expected value and its properties, • Dispersion, skewness, • quantiles, median and mode.	Data storage in R, data types and structures • Data types numeric, integer, complex, logical and character, • Vectors, matrices and arrays in R, • Lists, frames and factors, • Data input and output.

3	Continuous probability distributions • Uniform, exponential and normal distribution, • Random vector, • The associated distribution function, marginal distribution functions, • Associated density.	Probability distributions in R • The build in probabilistic functions, • Build in distributions, • Simulations and sampling from the given distribution.
4	Two dimensional random vectors • Moments of two dimensional random variables, • Covariance, correlation coefficient, • Conditional probability distribution, • Characteristics of the conditional distributions.	Functions in R and programming in R • The build in functions, • Creating own functions, • Program flow controls – if and loop statements, • Applied onto the concepts corresponding to the lecture.
5	Limit theorems • The law of large numbers, • Central limit theorem, • Basic statistical concepts – set of statistics, basic set, the sample, the statistical character.	Programming elementary graphics • Plotting the functions, plotting symbols, • Types of plots, • Colouring the plots, axes, description and text in the graph.
6	Elementary sample characteristics • The inductive statistics – random sample, • Sample mean, sample variance, • Probability distribution of the sample characteristics, • Sample mode and median.	Basics of the statistics • The descriptive statistical characteristics, • Sample mean, variance, skewness, • Methods of their computations in different situations, • Working with the data files, program outputs into the files.
7	Point estimates • Properties of the estimates (non-deformed, efficiency, consistency), • Moment's method, • The maximum likelihood method.	Data visualisation • Advanced graphics and data visualization, • Histograms, • Box plots, • Parameters estimation.
8	Interval estimates • Confidence intervals for the parameters of the normal distribution, • One-side and both-side estimates.	Data visualisation • Bar plots, • Pie charts, • Q-Q plots, • Basics of the ggplot package.
9	Hypotheses testing • Principles of the statistical hypotheses testing, • Error type 1 and 2, • The strength of the test, • p-value.	Confidence intervals and hypotheses testing • Confidence intervals, • Statistical tests, • Parametric tests, • Pair tests.

10	Tests on the parameters of the normal distribution • One-sample tests, • Two-sample tests, • Student's t-test, • Fisher's F-test.	Parametric hypotheses tests • One and two-sample tests, • Hypotheses tests for variance, • Multiple comparison methods.
11	Non-parametric tests • Goodness of fit test χ^2-test, • χ^2-square test of independence.	Non-parametric tests • Goodness of fit tests, • Independence test.
12	Correlation and regression analysis • Correlation coefficient, • Test of significance of the coefficient of correlation, • Linear regression – Simple linear regression, • Least squares method.	Regression and linear models • Statistical dependence measures, • Covariance matrix, • Correlation coefficient tests, • Simple linear regression model.

Student workload – forms of activity: Individual work with computer in the R, solving problems from statistics and programming in R, working with real data.

Teaching methods/tools: Lectures and laboratories, computer laboratory with anywhere OS (linux OS, Win, OS2), installed R environment (Open Source for any OS) and connection to internet.

Evaluation methods: Evaluation is based on two components – the continuous evaluation during the semester and final exam (total 100 points).

Evaluation process:

- Semester — 60 points: knowledge verification (written in the 9th week of the semester) — max. 40 points, special activities — max. 20 points.
- Exam — 40 points: test — max. 20 points, theoretical questions/tasks — max. 20 points.

To register for the exam the student must obtain at least 30 points during the semester.

Final Evaluation:

The condition for successful completion of the course is obtaining at least 61 points. This means at least 30 points during the semester, at least 10 points per test during the exam and at least 10 points for theoretical questions. Final evaluation of the course:

- **A** 93 — 100,
- **B** 85 — 92,
- **C** 77 — 84,
- **D** 69 — 76,
- **E** 61 — 68.

Bibliography:

- Verzani J., *Using R for Introductory Statistics*, 2014, Second edition, CRC Press, Taylor & Francis Group, Boca Raton, ISBN 9781466590731.
- Crawley M. J., *Statistics: An Introduction Using R*, Addison-Wesley Publishing company, 2015, ISBN 0-201-54199-8.