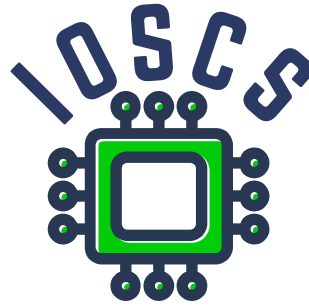


Project: Innovative Open Source Courses for Computer Science

# Pravdepodobnosť a štatistika s programovaním v R Sylabus

Aleš Kozubík  
Žilinská univerzita v Žiline

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

This project has been funded with support from the European Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

### Copyright Notice

This content was created by the IOSCS consortium: 2019–2022. The content is Copyrighted and distributed under Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

# POPIS KURZU

**Študijný odbor:** Informatika

**Úroveň:** Prvý kurz

**Názov kurzu:** Pravdepodobnosť a štatistika s programovaním v R

**ECTS kredity:** 5

**Druh, rozsah a metóda vzdelávacích činností:** Prednášky, laboratórne cvičenia

**Vyučovacie hodiny:** 24, 24

**Typ, rozsah a spôsob výučby:** 2 – 0 – 2 (prednášky – cvičenia – laboratórne cvičenia) hodín týždenne, prezenčné štúdium.

**Prerekvizity:** matematická analýza

**Výsledky vzdelávania:** Študent získa základné vedomosti z teórie pravdepodobnosti a matematickej štatistiky, ktoré umožňujú ich použitie pri riešení technických problémov v inžinierskej praxi a porozumenie jednotlivých pokročilých štatistických metód.

**Po absolvovaní kurzu študent:** Spozná/zopakuje si základné pojmy z pravdepodobnosti a štatistiky. Získa nové vedomosti z uvedených oblastí. Osvojí si základné nástroje a metódy pre praktické a teoretické riešenie analytických problémov. Dokáže identifikovať stochastický problém. Získa schopnosť aplikovať získané vedomosti pri riešení praktických úloh pomocou nástrojov Open Source.

**Obsah kurzu rozdelený do rôznych foriem výučby (s počtom hodín):**

týždeň	prednášky (2 hodiny týždenne)	laboratórne cvičenia (2 hodiny týždenne)
1	ZÁKLADNÉ POJMY Z PRAVDEPODOBNOTI • Pojem náhody a náhodnej udalosti, • Náhodná premenná, • Rozdelenie pravdepodobnosti.	ÚVOD DO PROGRAMOVACIEHO PROSTREDIA R, • Prvé kroky, inštalácia R, • R ako kalkulátor, • Algebraické operácie, • Balíčky a repozitáre, • Získanie nápovedy, • Ukončenie práce v R.
2	ROZDELENIE NÁHODNEJ PREMENNEJ • Hustota pravdepodobnosti, distribučná funkcia a jej vlastnosti, • Číselné charakteristiky náhodnej premennej, • Začiatočné a centrálné momenty, • Stredná hodnota a jej vlastnosti, • Rozptyl, šikmosť, • Kvantily, medián a modus.	UKLADANIE DÁT V R, DÁTOVÉ TYPY A ŠTRUKTÚRY, • Dátové typy numeric, integer, complex, logical a character, • Vektory, matice a polia v R, • Štruktúry zoznam, frame a faktory, • Vstup a výstup dát.

3	<p>SPOJITÉ ROZDELENIA PRAVDEPODOBNOSTI</p> <ul style="list-style-type: none"> <li>• Rovnomerné, exponenciálne a normálne rozdelenie,</li> <li>• Náhodný vektor,</li> <li>• Združená a marginálna distribučná funkcia,</li> <li>• Združená hustota,</li> </ul>	<p>ROZDELENIA IMPLEMENTOVANÉ V R</p> <ul style="list-style-type: none"> <li>• Implementované hustoty, distribučné a kvantilové funkcie,</li> <li>• Rozdelení obsiahnuté v prostredí R,</li> <li>• Simulácia a náhodný výber z daného rozdelenia.</li> </ul>
4	<p>DVOJROZMERNÉ NÁHODNÉ VEKTORY</p> <ul style="list-style-type: none"> <li>• Momenty dvojrozmerných náhodných vektorov,</li> <li>• Kovariancia a korelačný koeficient,</li> <li>• Podmienené rozdelenie pravdepodobnosti,</li> <li>• Charakteristiky podmieneného rozdelenia.</li> </ul>	<p>FUNKCIE V R A PROGRAMOVANIE V R</p> <ul style="list-style-type: none"> <li>• Zabudované funkcie,</li> <li>• Definovanie vlastných funkcií,</li> <li>• Riadení behu programu – podmienky a cykly,</li> <li>• Aplikácia na pojmoch z prednášky.</li> </ul>
5	<p>LIMITNÉ VETY</p> <ul style="list-style-type: none"> <li>• Zákon veľkých čísel,</li> <li>• Centrálna limitná veta,</li> <li>• Základní pojmy štatistiky – štatistický súbor, základný súbor, výberový súbor, štatistický znak.</li> </ul>	<p>ZÁKLADY PROGRAMOVANIA GRAFIKY</p> <ul style="list-style-type: none"> <li>• Vykreslenie grafov funkcií a symbolov,</li> <li>• Typy grafov,</li> <li>• Ofarbenie prvkov grafu, osí, popisov a textu v grafe.</li> </ul>
6	<p>ZÁKLADNÉ VÝBEROVÉ CHARAKTERISTIKY</p> <ul style="list-style-type: none"> <li>• Induktívna štatistika – náhodný výber a jeho realizácia,</li> <li>• Výberový priemer, výberový rozptyl,</li> <li>• Rozdelenie výberových charakteristík,</li> <li>• Výberový modus a medián.</li> </ul>	<p>ZÁKLADY ŠTATISTIKY</p> <ul style="list-style-type: none"> <li>• Popisné štatistické charakteristiky,</li> <li>• Výberový priemer, rozptyl a šikmosť,</li> <li>• Metódy výpočtu charakteristík v rôznych situáciách,</li> <li>• Práca s dátovými súborami, programové výstupy do súboru.</li> </ul>
7	<p>BODOVÉ ODHADY</p> <ul style="list-style-type: none"> <li>• Vlastnosti odhadov (nevychýlenosť, efektívnosť, konzistencia),</li> <li>• Momentová metóda,</li> <li>• Metóda maximálnej vierohodnosti.</li> </ul>	<p>VIZUALIZÁCIA DÁT</p> <ul style="list-style-type: none"> <li>• Pokročilá grafika a vizualizácia dát,</li> <li>• Histogramy,</li> <li>• Box ploty,</li> <li>• Odhady parametrov.</li> </ul>
8	<p>INTERVALOVÉ ODHADY</p> <ul style="list-style-type: none"> <li>• Intervaly spoľahlivosti pre parametre normálneho rozdelenia,</li> <li>• Jednostranné a obojstranné odhady.</li> </ul>	<p>VIZUALIZÁCIA DÁT</p> <ul style="list-style-type: none"> <li>• Stĺpcové grafy,</li> <li>• Koláčové grafy,</li> <li>• Kvantilové grafy,</li> <li>• Základy práce s balíčkom ggplot.</li> </ul>
9	<p>TESTOVANIE HYPOTÉZ</p> <ul style="list-style-type: none"> <li>• Princípy štatistických testov hypotéz,</li> <li>• Chyba 1. a 2.druhu,</li> <li>• Sila testu,</li> <li>• <math>p</math>-hodnota.</li> </ul>	<p>INTERVALY SPOĽAHLIVOSTI A TESTY HYPOTÉZ</p> <ul style="list-style-type: none"> <li>• Intervaly spoľahlivosti,</li> <li>• Štatistické testy,</li> <li>• Parametrické testy,</li> <li>• Párové testy.</li> </ul>

<b>10</b>	TESTY PRE PARAMETRE NORMÁLNEHO ROZDELENÍ <ul style="list-style-type: none"> <li>• Jednovýberové testy,</li> <li>• Testy na zhodu parametrov dvoch základných súborov,</li> <li>• Studentov <math>t</math>-test,</li> <li>• Fisherov <math>F</math>-test.</li> </ul>	PARAMETRICKÉ TESTY HYPOTÉZ <ul style="list-style-type: none"> <li>• Jedno a dvojjvýberové testy,</li> <li>• Testy pre rozptyl,</li> <li>• Viacnásobné porovnaní.</li> </ul>
<b>11</b>	NEPARAMETRICKÉ TESTY <ul style="list-style-type: none"> <li>• <math>\chi^2</math>-kvadrát test dobrej zhody,</li> <li>• <math>\chi^2</math>-kvadrát test nezávislosti.</li> </ul>	NEPARAMETRICKÉ TESTY <ul style="list-style-type: none"> <li>• Testy dobrej zhody,</li> <li>• Testy nezávislosti.</li> </ul>
<b>12</b>	KORELAČNÁ A REGRESNÁ ANALÝZA <ul style="list-style-type: none"> <li>• Koeficient korelácie,</li> <li>• Test významnosti pre koeficient korelácie,</li> <li>• Lineárna regresia – jednoduchá lineárna regresia,</li> <li>• Metóda najmenších štvorcov.</li> </ul>	REGRESIA A LINEÁRNE MODELY <ul style="list-style-type: none"> <li>• Miery štatistickej závislosti,</li> <li>• Kovariančná matica,</li> <li>• Testy pre korelačný koeficient,</li> <li>• Jednoduchý lineárny regresný model.</li> </ul>

**Záťaž študenta – formy činnosti:** Samostatná práca s počítačom v prostredí R, riešenie problémov zo štatistiky a programovanie v programe R, práca s reálnymi dátami.

**Vyučovacie metódy/nástroje:** Prednášky a laboratórne cvičenia, ľubovoľný OS (linux OS, Win, OS2), nainštalované prostredie R (Open Source pre akýkoľvek OS) a pripojenie na internet.

**Metódy hodnotenia:** Hodnotenie je založené na dvoch zložkách – priebežné hodnotenie počas semestra a záverečná skúška (spolu 100 bodov).

*Priebeh hodnotenia:*

- Semester – 60 bodov: overenie vedomostí (písané v 9. týždni semestra) – max. 40 bodov, špeciálne činnosti – max. 20 bodov.
- Skúška – 40 bodov: test – max. 20 bodov, teoretické otázky/úlohy – max. 20 bodov.

Na prihlásenie na skúšku musí študent získať najmenej 30 bodov počas semestra.

*Záverečné hodnotenie:*

Podmienkou úspešného absolvovania kurzu je získanie minimálne 61 bodov. To znamená najmenej 30 bodov počas semestra, najmenej 10 bodov za test počas skúšky a najmenej 10 bodov za teoretické otázky. Výsledné hodnotenie kurzu:

- 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.