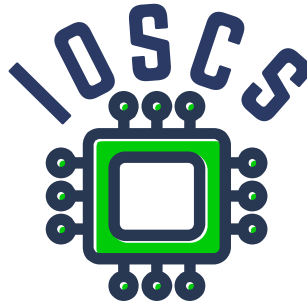


Project: Innovative Open Source Courses for Computer Science

Pravděpodobnost a statistika s programováním 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

Studijní obor: Informatika

Úroveň: První kurz

Název kurzu: Pravděpodobnost a statistika s programováním v R

ECTS kredity: 5

Druh, rozsah a metoda vzdělávacích činností: Přednášky, laboratorní cvičení

Vyučovací hodiny: 24, 24

Typ, rozsah a způsob výuky: 2 – 0 – 2 (přednášky – cvičení – laboratorní cvičení) hodin týdně, prezenční studium.

Prerekvizity: matematická analýza

Výsledky vzdělávání: Student získá základní znalosti z teorie pravděpodobnosti a matematické statistiky, které umožňují jejich použití při řešení technických problémů v inženýrské praxi a porozumění jednotlivých pokročilých statistických metod.

Po absolvování kurzu student: Poznává/zopakuje si základní pojmy z pravděpodobnosti a statistiky. Získá nové znalosti z uvedených oblastí. Osvojí si základní nástroje a metody pro praktické a teoretické řešení analytických problémů. Dokáže identifikovat stochastický problém. Získá schopnost aplikovat získané znalosti při řešení praktických úkolů pomocí nástrojů Open Source.

Obsah kurzu rozdělený do různých forem výuky (s počtem hodin):

týden	přednášky (2 hodiny týdně)	laboratorní cvičení (2 hodiny týdně)
1	ZÁKLADNÍ POJMY Z PRAVDĚPODOBNOСТИ • Pojem náhody a náhodného jevu, • Náhodná proměnná, • Rozdělení pravděpodobnosti.	ÚVOD DO PROGRAMOVACÍHO PROSTŘEDÍ R • První kroky, instalace R, • R jako kalkulátor, • Algebraické operace, • Balíčky a repozitáře, • Získání nápovědy, • Ukončení práce v R.
2	ROZDĚLENÍ NÁHODNÉ VELIČINY • Hustota pravděpodobnosti, distribuční funkce a jejich vlastnosti, • Číselné charakteristiky náhodné veličiny, • Počáteční a centrální momenty, • Střední hodnota a její vlastnosti, • Rozptyl, šikmost, • Kvantily, medián a modus.	UKLÁDÁNÍ DAT V R, DATOVÉ TYPY A STRUKTURY • Datové typy numeric, integer, complex, logical a character, • Vektory, matice a pole v R, • Struktury seznam, frame a faktory, • Vstup a výstup dat.
3	SPOJITÁ ROZDĚLENÍ PRAVDĚPODOBNOСТИ • Rovnoměrné, exponenciální a normální rozdělení, • Náhodný vektor, • Sdružená a marginální distribuční funkce, • Sdružená hustota.	ROZDĚLENÍ ZABUDOVANÁ V R • Zabudované hustoty, distribuční a kvantilové funkce, • Rozdělení obsažená v prostředí R, • Simulace a náhodný výběr z daného rozdělení.

4	<p>DVOUROZMĚRNÉ NÁHODNÉ VEKTORY</p> <ul style="list-style-type: none"> • Momenty dvourozměrných náhodných vektorů, • Kovariance a korelační koeficient, • Podmíněné rozdělení pravděpodobnosti, • Charakteristiky podmíněného rozdělení. 	<p>FUNKCE V R A PROGRAMOVÁNÍ V R</p> <ul style="list-style-type: none"> • Zabudované funkce, • Definice vlastních funkcí, • Řízení běhu programu – podmínky a cykly, • Aplikace na pojmech z přednášky.
5	<p>LIMITNÍ VĚTY</p> <ul style="list-style-type: none"> • Zákon velkých čísel, • Centrální limitní věta, • Základní pojmy statistiky – statistický soubor, základní soubor, výběrový soubor, statistický znak. 	<p>ZÁKLADY PROGRAMOVÁNÍ GRAFIKY</p> <ul style="list-style-type: none"> • Vykreslení grafů funkcí a symbolů, • Typy grafů, • Obarvení prvků grafu, os, popisu a textu v grafu.
6	<p>ZÁKLADNÍ VÝBĚROVÉ CHARAKTERISTIKY</p> <ul style="list-style-type: none"> • Induktivní statistika – náhodný výběr a jeho realizace, • Výběrový průměr, výběrový rozptyl, • Rozdělení výběrových charakteristik, • Výběrový modus a medián. 	<p>ZÁKLADY STATISTIKY</p> <ul style="list-style-type: none"> • Popisné statistické charakteristiky, • Výběrový průměr, rozptyl a šikmost, • Metody výpočtu charakteristik v různých situacích, • Práce s datovými soubory, programové výstupy do souboru.
7	<p>BODOVÉ ODHADY</p> <ul style="list-style-type: none"> • Vlastnosti odhadů (nestrannost, efektivita, konzistence), • Momentová metoda, • Metoda maximální věrohodnosti. 	<p>VIZUALIZACE DAT</p> <ul style="list-style-type: none"> • Pokročilá grafika a vizualizace dat, • Histogramy, • Box ploty, • Odhady parametrů.
8	<p>INTERVALOVÉ ODHADY</p> <ul style="list-style-type: none"> • Intervaly spolehlivosti pro parametry normálního rozdělení, • Jednostranné a oboustranné odhady. 	<p>VIZUALIZACE DAT</p> <ul style="list-style-type: none"> • Sloupcové grafy, • Koláčové grafy, • Kvantilové grafy, • Základy práce s balíčkem ggplot.
9	<p>TESTOVÁNÍ HYPOTÉZ</p> <ul style="list-style-type: none"> • Principy statistických testů hypotéz, • Chyba 1. a 2.druhu, • Síla testu, • p-hodnota. 	<p>INTERVALY SPOLEHLIVOSTI A TESTY HYPOTÉZ</p> <ul style="list-style-type: none"> • Intervaly spolehlivosti, • Statistické testy, • Parametrické testy, • Párové testy.
10	<p>TESTY NA PARAMETRY NORMÁLNÍHO ROZDĚLENÍ</p> <ul style="list-style-type: none"> • Jednovýběrové testy, • Testy na shodu parametrů dvou základních souborů, • Studentův t-test, • Fisherův F-test. 	<p>PARAMETRICKÉ TESTY HYPOTÉZ</p> <ul style="list-style-type: none"> • Jedno a dvouvýběrové testy, • Testy pro rozptyl, • Vícenásobná porovnání.

11	NEPARAMETRICKÉ TESTY χ^2 test dobré shody, χ^2 test nezávislosti.	NEPARAMETRICKÉ TESTY • Testy dobré shody, • Testy nezávislosti.
12	KORELAČNÍ A REGRESNÍ ANALÝZA • Koeficient korelace, • Test významnosti pro koeficient korelace, • Lineární regrese – jednoduchá lineární regrese, • Metoda nejmenších čtverců.	REGRESE A LINEÁRNÍ MODELY • Míry statistické závislosti, • Kovarianční matice, • Testy pro korelační koeficient, • Jednoduchý lineární regresní model.

Zatížení studenta – formy činnosti: Samostatná práce s počítačem v prostředí R, řešení problémů ze statistiky a programování v programu R, práce s reálnými daty.

Vyučovací metody/nástroje: Přednášky a laboratorní cvičení, libovolný OS (linux OS, Win, OS2), nainstalované prostředí R (Open Source pro jakýkoli OS) a připojení k internetu.

Metody hodnocení: Hodnocení je založeno na dvou složkách – průběžné hodnocení během semestru a závěrečná zkouška (celkem 100 bodů).

Průběh hodnocení:

- Semestr – 60 bodů: ověření znalostí (psáno v 9. týdnu semestru) – max. 40 bodů, speciální činnosti – max. 20 bodů.
- Zkouška – 40 bodů: test – max. 20 bodů, teoretické otázky/úkoly – max. 20 bodů.

Pro přihlášení na zkoušku musí student získat nejméně 30 bodů během semestru.

Závěrečné hodnocení:

Podmínkou úspěšného absolvování kurzu je získání minimálně 61 bodů. To znamená nejméně 30 bodů během semestru, nejméně 10 bodů za test během zkoušky a nejméně 10 bodů za teoretické otázky. Výsledné hodnocení 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.