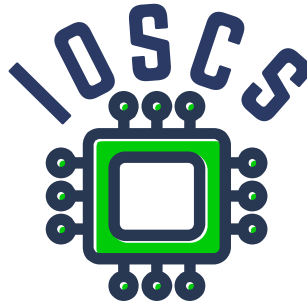


Project: Innovative Open Source Courses for Computer Science

Prawdopodobieństwo i statystyka z programowaniem w R Sylabus

Aleš Kozubík
Uniwersytet Žyliński w Žylinie

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).

OPIS KURSU

Kierunek studiów: Informatyka

Poziom: Pierwszy kurs

Nazwa kursu: Prawdopodobieństwo i statystyka z programowaniem w R

Punkty ECTS: 5

Rodzaj, zakres i metoda zajęć dydaktycznych: Wykłady, ćwiczenia laboratoryjne

Godziny zajęć: 24, 24

Rodzaj, zakres i metoda nauczania: 2 – 0 – 2 (wykłady – ćwiczenia – ćwiczenia laboratoryjne) godzin tygodniowo, studia stacjonarne.

Wymagania wstępne: analiza matematyczna

Efekty uczenia się: Student uzyskuje podstawową wiedzę z zakresu teorii prawdopodobieństwa i statystyki matematycznej, która umożliwi ich wykorzystanie w rozwiązywaniu problemów technicznych w praktyce inżynierskiej oraz zrozumienie poszczególnych zaawansowanych metod statystycznych.

Po ukończeniu kursu student: Rozpoznaje/powtarza podstawowe pojęcia z rachunku prawdopodobieństwa i statystyki. Zdobędzie nową wiedzę z wymienionych obszarów. Otrzyma podstawowe narzędzia i metody praktycznego i teoretycznego rozwiązywania problemów analitycznych. Potrafi zidentyfikować problem stochastyczny. Nabędzie umiejętności zastosowania zdobytej wiedzy w rozwiązywaniu praktycznych zadań z wykorzystaniem narzędzi Open Source.

Treść kursu z podziałem na różne formy nauczania (z podaniem liczby godzin):

Tydzień	Wykład (2 h tydzień)	Laboratoria(2 h tydzień)
1	ELEMENTY TEORII PRAWDOPODOBIENSTWA • Pojęcia losowości, przypadkowe wydarzenie, • Zmienna losowa, • Rozkład prawdopodobieństwa, zmienne losowe.	WPROWADZENIE DO ŚRODOWISKA R • Pierwsze kroki, instalacja R, • Podstawowe cechy R, • R jako kalkulator, • Operacje algebraiczne, • Pakiety i repozytoria, • Uzyskiwanie pomocy, • Kończenie pracy Z R.
2	ROZKŁAD PRAWDOPODOBIENSTWA ZMIENNEJ LOSOWEJ • Gęstość prawdopodobieństwa, funkcje dystrybucji i ich własności, • Charakterystyka zmiennej losowej, • Momenty początkowe i centralne, • Wartość oczekiwana i jej właściwości, • Rozpraszanie, skośność, • Kwantyle, mediany i mody.	PRZECHOWYWANIE DANYCH W R, TYPY DANYCH I STRUKTUR • Numeryczne typy danych w R, liczby całkowite, złożone, • Wektory, macierze i tablice w R, • Struktury, listy i współczynniki, • Wprowadzanie i wyprowadzanie danych.

3	<p>CIĄGŁY ROZKŁADY PRAWDOPODOBIENSTWA</p> <ul style="list-style-type: none"> • Rozkład jednolity, wykładniczy i normalny, • Wektor losowy, • Funkcje dystrybucji, • Funkcja gęstości prawdopodobieństwa. 	<p>ROZKŁADY PRAWDOPODOBIENSTWA W R</p> <ul style="list-style-type: none"> • Wbudowanie funkcji probabilistyczne, • Funkcje prawdopodobieństwa w R, • Symulacje i pobieranie danych z dystrybucji.
4	<p>DWUWYMIAROWE WEKTORY LOSOWE</p> <ul style="list-style-type: none"> • Dwuwymiarowe wektory losowe, • Kowariancja i współczynnik korelacji, • Warunkowe rozkłady prawdopodobieństwa, • Charakterystyka warunkowych rozkładów prawdopodobieństwa. 	<p>FUNKCJE W R I PROGRAMOWANIE W R</p> <ul style="list-style-type: none"> • Wbudowane funkcje, • Definicja własnych funkcji, • Sterowanie przebiegiem programu – warunki i pętle, • Odniesienie do wykładu.
5	<p>TWIERDZENIE GRANICZNE</p> <ul style="list-style-type: none"> • Prawo wielkich liczb, • Centralne twierdzenie graniczne, • Podstawowe pojęcia statystyczne – zbiór, statystyka, zestaw podstawowy, próbka. 	<p>PROGRAMOWANIE ELEMENTARNEJ GRAFIKI</p> <ul style="list-style-type: none"> • Wykreślanie funkcji, kreślenie symboli, • Rodzaje wykresów, • Dostosowywanie wykresów, osi, opisów i tekstu na wykresie.
6	<p>PODSTAWOWE CECHY PRÓBK</p> <ul style="list-style-type: none"> • Statystyka indukcyjna – próba losowa, • Średnia próby, wariancja próby, • Rozkład prawdopodobieństwa próby, • Mediana. 	<p>PODSTAWY STATYSTYKI</p> <ul style="list-style-type: none"> • Opisowe cechy statystyczne, • Średnia próbna, wariancja i skośność, • Metody obliczania charakterystyk w różnych sytuacjach, • Praca z plikami danych, programowa obsługa plików.
7	<p>SZACUNKI PUNKTOWE</p> <ul style="list-style-type: none"> • Właściwości szacunków (bezstronność, wydajność, spójność), • Metoda momentowa, • Metoda największej wiarygodności. 	<p>WIZUALIZACJA DANYCH</p> <ul style="list-style-type: none"> • Zaawansowana grafika i wizualizacja danych, • Histogramy, • Wykresy typu Box, • Estymacja parametrów.
8	<p>SZACUNKI PRZEDZIAŁOWE</p> <ul style="list-style-type: none"> • Przedziały ufności dla parametrów rozkładu normalnego, • Szacunki jednostronne i dwustronne. 	<p>WIZUALIZACJA DANYCH</p> <ul style="list-style-type: none"> • Wykresy słupkowe, • Wykresy kołowe, • Wykresy QQ, • Podstawy pakietu ggplot.
9	<p>TESTOWANIE HIPOTEZ</p> <ul style="list-style-type: none"> • Zasady statystyki, testowanie hipotez, • Typy błędów 1 i 2, • Siła testu, • p-wartość. 	<p>PRZEDZIAŁY UFNOŚCI I TESTOWANIE HIPOTEZ</p> <ul style="list-style-type: none"> • Przedziały ufności, • Testy statystyczne, • Testy parametryczne, • Testy w parach.

10	TESTY PARAMETRÓW W ROZKŁADZIE NORMALNYCH <ul style="list-style-type: none"> • Testy na jednej próbie, • Testy na dwóch próbach, • Student t-test, • Fisher F-test. 	TESTY HIPOTEZ PARAMETRYCZNYCH <ul style="list-style-type: none"> • Testy z jedno i dwoma próbkami, • Testy hipotez pod kątem wariancji, • Metody porównawcze.
11	TESTY NIEPARAMETRYCZNE <ul style="list-style-type: none"> • dopasowanie testu χ^2, • χ^2 niezależność testu. 	TESTY NIEPARAMETRYCZNE <ul style="list-style-type: none"> • Testy zgodności, • Testy niezależności.
12	ANALIZA KORELACJI I REGRESJI <ul style="list-style-type: none"> • Współczynnik korelacji, • Test istotności współczynnika korelacji, • Regresja liniowa, • Metoda najmniejszych kwadratów. 	MODELE REGRESYJNE I LINIOWE <ul style="list-style-type: none"> • Miary zależności statystycznej, • Macierz kowariancji, • Testy współczynników korelacji, • Prosty model regresji liniowej.

Obciążenie studentów – formy aktywności: Samodzielna praca z komputerem w środowisku R, rozwiązywanie zadań ze statystyki i programowanie w programie R, praca z danymi rzeczywistymi.

Metody/narzędzia dydaktyczne: Wykłady i ćwiczenia laboratoryjne, dowolny system operacyjny (linux OS, Win, OS2), zainstalowane środowisko R (Open Source dla dowolnego systemu operacyjnego) i połączenie z Internetem.

Metody oceny: Ocena opiera się na dwóch elementach – ocena śródroczna w trakcie semestru i egzamin końcowy (razem 100 punktów).

Postęp oceny:

- Semestr – 60 pkt.: weryfikacja wiedzy (pisemna w 9 tygodniu semestru) – max. 40 punktów, zajęcia specjalne – maks. 20 punktów.
- Egzamin – 40 punktów: test – max. 20 pkt., pytania/zadania teoretyczne – max. 20 punktów.

Warunkiem wpisu na egzamin jest uzyskanie w ciągu semestru co najmniej 30 punktów.

Końcowa ocena:

Warunkiem pomyślnego zaliczenia kursu jest uzyskanie co najmniej 61 punktów. Oznacza to co najmniej 30 punktów w ciągu semestru, co najmniej 10 punktów za kolokwium podczas egzaminu oraz co najmniej 10 punktów za pytania teoretyczne. Ocena końcowa kursu:

- 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.