Warsaw University
of Technology

COURSE OFFERED IN THE DOCTORAL SCHOOL

Code of the course	4606-ES-CDEGHKL-0267	Name of the course	Polish	**Analiza i wizualizacja danych w języku R**	
			English	**Data analysis and visualization in R**	
Type of the course	Special courses				
Course coordinator	prof. dr hab. inż. Janusz Hołyst	Course teacher	dr inż. Robert Paluch		
Implementing unit	Faculty of Physics	Scientific discipline / disciplines*	Chemical engineering , materials engineering, chemical sciences, physical sciences, biomedical engineering, mechanical engineering, information and communication technology		
Level of education	Doctoral studies	Semester	Winter		
Language of the course	English				
Type of assessment	Graded credit	Number of hours in a semester	30	ECTS credits	2
Minimum number of participants	10	Maximum number of participants	20 (10 per group)	Available for students (BSc, MSc)	No

Type of classes		Lecture	Auditory classes	Project classes	Laboratory	Seminar
Number of hours	in a week	0	0	0	3	0
	in a semester	0	0	0	30	0

* does not apply to the Researcher's Workshop

1. Prerequisites

None

2. Course objectives

The aim of the course is to familiarize students with the basics of the R language and to teach them the methods of data analysis and visualization in this language.

3. Course content (separate for each type of classes)

Lecture

Does not apply

Laboratory

1. Introduction to R.
   a) Variables, instructions, file input and output.
   b) Vectors and matrix operations.
   c) Random values generation from selected distribution.
   d) Generating documents in R Markdown.
2. Graphics in R.
   a) Scatter and box plots.
   b) Histograms, binning.
   c) ggplot2 library.
   d) Output to PNG, JPG and EPS file formats.
3. Data preprocessing, statistical models and tests.
   a) Data transformation (tidyr and dplyr packages).

b) Hypothesis testing with parametric and nonparametric methods.
c) Distributions fitting to experimental data.
d) Chi2 test.
e) One and multidimensional analysis of variance.
f) Linear, logistic and Poisson regressions.
4. Data-mining.
a) Classification efficiency.
b) Cross-validation.
c) Discriminant analysis.
d) Classification trees
e) Cluster analysis.
f) Multidimensional scaling.
g) Principal component analysis.

4. Learning outcomes

Type of learning outcomes	Learning outcomes description	Reference to the learning outcomes of the WUT DS	Learning outcomes verification methods*
Knowledge			
K01	Absolwent zna i rozumie podstawy teoretyczne metod statystycznej analizy danych.	SD_W2	homework
K02	Absolwent zna główne trendy rozwojowe związane z metodami statystycznej analizy danych.	SD_W3	homework
Skills			
S01	Absolwent potrafi wyciągać poprawne wnioski na podstawie wyników uzyskanych podczas analizy i wizualizacji danych	SD_U1	homework
Social competences			
SC01	Absolwent jest gotów do niezależnego prowadzenia badań naukowych obejmujących analizę i wizualizację danych.	SD_K5	homework

*Allowed learning outcomes verification methods: exam; oral exam; written test; oral test; project evaluation; report evaluation; presentation evaluation; active participation during classes; homework; tests

5. Assessment criteria

Assignments (started at class and completed at home): 10 x 5 points = 50 points
Marks: (25; 30) - 3.0; [30; 35) – 3.5; [35; 40) – 4.0; [40; 45) – 4.5; [45; 50] – 5.0

6. Literature

Primary references:

[1] P. Biecek, Przewodnik po pakiecie R, Oficyna Wydawnicza Gis, Wrocław 2017.

[2] T. Górecki, Podstawy statystyki z przykładami w R, Wydawnictwo BTC, Legionowo 2011.

[3] M. Walesiak, E. Gatnar , Statystyczna analiza danych z wykorzystaniem pakietu R, Wydawnictwo Naukowe PWN, Warszawa 2013.

Secondary references:

[1] K. Seefeld, E. Linder, Statistics Using R with Biological Examples, https://cran.r-project.org/doc/contrib/Seefeld_StatsRBio.pdf

[2] A. Coghlan, A Little Book of R For Biomedical Statistics, http://a-little-book-of-r-for-biomedical-statistics.readthedocs.io/en/latest/

[3] A. Coghlan, A Little Book of R For Multivariate Analysis, http://a-little-book-of-r-for-biomedical-statistics.readthedocs.io/en/latest/

7. PhD student's workload necessary to achieve the learning outcomes**		
No.	Description	Number of hours
1	Hours of scheduled instruction given by the academic teacher in the classroom	30
2	Hours of consultations with the academic teacher, exams, tests, etc.	5
3	Amount of time devoted to the preparation for classes, preparation of presentations, reports, projects, homework	15
4	Amount of time devoted to the preparation for exams, test, assessments	0
	Total number of hours	**50**
	ECTS credits	**2**

** 1 ECTS = 25-30 hours of the PhD students work (2 ECTS = 60 hours; 4 ECTS = 110 hours, etc.)

8. Additional information	
Number of ECTS credits for classes requiring direct participation of academic teachers	1.4
Number of ECTS credits earned by a student in a practical course	1.2