HANDONG UNITWIN FELLOWSHIP

SYLLABUS

Lecturer's Information	
Name	Nataliia Kruhlova
University	National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

Course Information			
Course Name	R Language in Computational Probability and Statistics	Prerequisite Course	"Probability Theory"
Course Date	2024-02-19 ~ 2024-05-14	Course Language	Ukrainian
Keyword			

Course Description (100 ~200 words)	The "R Language in Computational Probability and Statistics" course focuses on teaching basic concepts and algorithms of mathematical statistics using the R programming language. Targeted at undergraduate students of NTUU "Igor Sikorsky Kyiv Polytechnic Institute," the course caters to those studying probability, statistics, and financial mathematics, with career paths in Financial Analytics, Data Science, and Machine Learning. Its objective is to acquaint students with essential topics in computational statistics through hands-on experience, offering a modern perspective on probability and its applications in Data Science. Utilizing R, a widely-used free programming language for scientific computing, students can immediately visualize their results. Moreover, R serves as a foundation for future courses in Machine Learning, Data Science, and Computer Statistics. The course comprises both lectures and practical sessions, providing a comprehensive learning experience.
Course Goals and Objectives (Approximately 100 words)	Studying the course "R Language in Computational Probability and Statistics" provides the following: - basic knowledge of the statistical programming language R; - the ability to work independently with the basic structures of the language and graphically display data, the implementation of specialized packages, and creation programs and functions. - improvement of the knowledge and skills in the field of modern statistical methods, specifically in the subjects of descriptive statistics; -basic typical laws of probability distribution; -calculating the probabilities of random events using R; -describing stochastic processes using probability distribution laws; -modeling the behavior of stochastic objects; -selecting and using methods for processing and analyzing data or research results.
Textbook	**(The format : Title, Author/Editor, Publisher, Year of Publication)** Probability and statistics with R, Ugarte M.D., Militino A.F., Arnholt A.T/ Boca Raton, London, New York: CRC Press, Taylor&Francis Group, 2008. –700 p.
References	1. http://cran.us.r-project.org 2. Statistics With R, Jenine K. Harris/ Washington University in St.Louis, USA, 2020. 3. The Art of R Programming: A Tour of Statistical Software Design, Matloff N./ No Starch Press, 2011. 4. Data Visualisation with R, Thomas Rahlf/ Springer International Publishing, New York, 2017. 5. Introduction to Scientific Programming and Simulation using R, Jones O., Maillardet R., Robinson A.P./ Chapman, Hall, CRC, 2009. 6. An introduction to data cleaning with R, Jonge E. de, Loo M. van der/ Statistics Netherlands, 2013. 7. Using R for Statistics, Sarah Stowell/ Apress, 2014. 8. Biostatistics with R: An Introduction to Statistics Through Biological Data, Shahbaba B./ Springer, 2012.
Course Requirements and Grades	Knowledge of the probability theory and mathematical statistics. Knowledge of the basic structures of the language R.

Weekly Schedule		
Week	Main Topics	Keywords
1	Basics of working with the R. RStudio.	R programming language, RStudio, Interface, Console, Script editor, Workspace, Working directory, Packages, Installing packages
2	Vectors. Vector functions. Arithmetic operators and special values.	Vector functions, Vectorization, Arithmetic operators, Special values, NA (Not Available), NaN (Not a Number), Inf (Infinity)
3	Factors. Arrays and matrices.	Factors, Categorical variables, Levels, Ordered factors, Arrays, Multi-dimensional arrays, Matrices, Row operations, Column operations, Matrix operations, Indexing, Dimensions
4	Frames. Lists.	Data frames, Lists, Data structure, Columns, Rows, Indexing, Hierarchical data, Data organization, Manipulation
5	Data import/export. Creating own functions.	CSV files, Excel files, Text files, Data frames, Reading data, Writing data, User-defined functions, Function arguments
6	Branches, cycles. Vectorized calculations using apply functions.	Looping, Branching, Control structures, for, while, if-else statements, lapply, sapply, vapply, mapply, Iteration
7	Data manipulation. Package dplyr.	Data frame, Filtering, Selecting, Mutating, Summarizing, Grouping
8	High-level plotting functions.	Plot, ggplot2, data visualization, grammar of graphics, Layers, Aesthetics, Geoms, Facets, Themes, Coordination systems
9	Low-level plotting functions.	Plotting, Graphics, Base graphics, Scatter plots, Line plots, Bar plots
10	Discrete densities and distributions.	Binomial distribution, Poisson distribution Geometric distribution, Hypergeometric distribution, Bernoulli distribution
11	Continuous distributions and densities.	Normal distribution, Uniform distribution, Exponential distribution, Beta distribution, Gamma distribution, Log-normal distribution Weibull distribution, Chi-square distribution
12	Descriptive statistics.	Mean, median, mode, variance and standard deviation, skewness and kurtosis, percentiles and quartiles, range
13	Graphical methods of statistics.	Data visualization, Histograms, Scatter plots, Box plots
14		
15		
16		