SYLLABUS

1. Information regarding the programme

1.1 Higher education institution	**Babeş-Bolyai University**
1.2 Faculty	**Faculty of Mathematics and Computer Science**
1.3 Department	**Department of Computer Science**
1.4 Field of study	**Computer Science**
1.5 Study cycle	**Master**
1.6 Study programme / Qualification	**Data Science for Industry and Society**

2. Information regarding the discipline

2.1 Name of the discipline		**Basics of Statistics**					
2.2 Course coordinator		**Prof. Sanda Micula, PhD. Habil.**					
2.3 Seminar coordinator		**Prof. Sanda Micula, PhD. Habil.**					
2.4. Year of study	**1**	2.5 Semester	**1**	2.6. Type of evaluation	**E**	2.7 Type of discipline	**DF Compulsory**

2.8 Course Code	**MME8180**		

3. Total estimated time (hours/semester of didactic activities)

3.1 Hours per week	4	Of which: 3.2 course	2	3.3 seminar/laboratory	1 + 1 pr.
3.4 Total hours in the curriculum	56	Of which: 3.5 course	28	3.6 seminar/laboratory	28

Time allotment:	hours
Learning using manual, course support, bibliography, course notes	32
Additional documentation (in libraries, on electronic platforms, field documentation)	8
Preparation for seminars/labs, homework, papers, portfolios and essays	26
Tutorship	10
Evaluations	18
Other activities:	-

3.7 Total individual study hours	94
3.8 Total hours per semester	150
3.9 Number of ECTS credits	6

4. Prerequisites (if necessary)

4.1. curriculum	- Basic notions of Algebra - Basic notions of Mathematical Analysis
4.2. competencies	- Logical thinking - Basic logical programming skills

5. Conditions (if necessary)

5.1. for the course	• Lecture room with large blackboard and video projector, laptop, beamer
5.2. for the seminar /lab	• Laboratory with computers having Office and Matlab installed

6. Specific competencies acquired

Professional competencies	C4.1 Defining basic concepts, theory and mathematical models C4.2 Interpretation of mathematical models C4.3 Identifying the appropriate models and methods for solving real-life problems C4.5 Embedding formal models in applications from various areas
Transversal competencies	CT1 Ability to conform to the requirements of organized and efficient work, to develop a responsible approach towards the academic and scientific fields, in order to make the most of one's own creative potential, while obeying the rules and principles of professional ethic CT3 Using efficient methods and techniques for learning, information, research and developing capabilities for using knowledge, for adapting to a dynamic society and for communicating in Romanian and in a worldwide spoken language

7. Objectives of the discipline (outcome of the acquired competencies)

7.1 General objective of the discipline	• Understand the broad directions of Descriptive and Inferential Statistics in order to perform analysis of data • Acquire the ability to use statistical analysis features of various software
7.2 Specific objective of the discipline	• Acquire the ability to collect, organize and display data in order to perform statistical analysis • Become familiar and be able to work with various statistical models and algorithms • Acquire the ability to use statistical software and interpret the results

8. Content

8.1 **Course**	Teaching methods	Remarks
1. **Review of Probability theory notions.** Random variables. Probability distribution function. Cumulative distribution function. Common discrete and continuous distributions.	• Interactive exposure • Explanation • Conversation • Description	
2. Numerical characteristics of random variables. Expectation, median, mode, variance, standard deviation, covariance, correlation coefficient. Properties. Examples.	• Interactive exposure • Explanation • Conversation • Description	
3. **Introduction.** Populations and samples, terminology. Declaring the objectives. Data collection. Statistical data mining	• Interactive exposure • Explanation • Conversation	

techniques. Graphical display of data, pie charts, bar graphs. Frequency and grouped frequency tables. Histograms, frequency polygons, stem-and-leaf plots.	• Description	
4. **Descriptive Statistics.** Parameters of a statistical distribution. Measures of central tendency. Measures of variability. Variance and standard deviation, interpretation. Coefficient of variation.	• Interactive exposure • Explanation • Conversation • Description	
5. Percentiles, deciles, quartiles, interquartile range. Outliers, detection of outliers, the 3σ rule. Boxplots. Sample theory. Sample functions. Standard errors of estimates. Examples and applications. Interpretation of data.	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
6. **Inferential Statistics**. Notions of estimation theory. Point estimators, properties. Method of moments. Examples. Estimation of standard eroors.	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
7. The Normal and Standard Normal distribution, Z-quantiles. The Student T-distribution and T- quantiles. One-sided and two-sided confidence intervals. Estimating the mean and the proportion by confidence intervals. Examples. Selecting the sample size.	• Interactive exposure • Explanation • Conversation • Description	
8. Two-sample statistics, pooled proportion. Estimating the difference of proportions. Confidence intervals for paired data. Pooled variance of two samples. Estimating the difference of means by confidence intervals. Examples.	• Interactive exposure • Explanation • Conversation • Description	
9. Hypothesis testing. Basic concepts, general framework. Rejection region. Type I and type II errors. Significance testing and P-values.	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
10. Z-tests for the mean. Selecting the sample size. Examples. T (Student)-tests for the mean. Tests for proportions.	• Interactive exposure • Explanation • Conversation	
11. The Fisher F-distribution and F-quantiles. F-tests for the ratio of variances. Tests for the difference of means. Paired data tests. Computation of type II errrors probabilities. Examples.	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
12. **Correlation and Regression.** Two-sample statistics. Scatter plots and time plots.	• Interactive exposure • Explanation	

Conditional mean. Curves of regression. Method of least squares. Linear regression. Examples.	• Conversation	
13. Overfitting a model. Polynomial regression. Examples and applications. Fitting models. Univariate analysis of variance (ANOVA) and R-square. Prediction. Examples.	• Interactive exposure • Explanation • Conversation • Description	
14. Multivariate analysis of variance (ANOVA), and F-test. Coefficient of determination. Adjusted R-square. Categorical predictors and dummy variables. Significant correlation. Examples.	• Interactive exposure • Explanation • Conversation • Didactical demonstration	

Bibliography
1. Micula, S., Probability and Statistics for Computational Sciences, Cluj University Press, 2009.
2. Miller, J.D., Statistics for Data Science, Packt Publishing, Birmingham, UK, 2017.
3. Bruce P., Bruce A., Practical Statistics for Data Scientists, 50 Essential Concepts, O'Reilly Media, CA, USA, 2017.
4. Baron, M., Probability and Statistics for Computer Scientists, 3rd edition, CRC Press, Taylor and Francis, Boca Raton, FL, USA, 2019.
5. Milton, J.S., Arnold, J. C., Introduction to Probability and Statistics: Principles and Applications for Engineering and the Computing Sciences, 3rd Edition. McGraw-Hill, New York, 1995.
6. Gentle, J. E., Elements of Computational Statistics, Springer-Verlag, New York, 2002.

8.2 **Seminar /Laboratory**	Teaching methods	Remarks
1. Introduction to Matlab (Octave).	• Interactive exposure • Explanation • Conversation	The seminar is structured as 2 hours per week, every other week
2. Random variables and their characteristics. Statistics and Machine Learning toolbox in Matlab.	• Interactive exposure • Explanation • Conversation • Individual and group work	
3. Descriptive Statistics. Grouped frequency table. Computation of statistical measures. Graphical display of data. Histogram, frequency polygon, boxplot.	• Interactive exposure • Conversation • Synthesis • Individual and group work	
4. Confidence intervals. Interpretation of results.	• Interactive exposure • Explanation • Conversation • Individual and group work	
5. Hypothesis testing. Interpretation of results.	• Interactive exposure • Explanation • Conversation • Individual and group work	
6. Correlation and regression.	• Interactive exposure • Explanation • Conversation	

	• Individual and group work	
7. Presentation. A project of statistical analysis of data.	• Conversation • Individual and group work	

Bibliography
1. Micula, S., Probability and Statistics for Computational Sciences, Cluj University Press, 2009.
2. Miller, J.D., Statistics for Data Science, Packt Publishing, Birmingham, UK, 2017.
3. Bruce P., Bruce A., Practical Statistics for Data Scientists, 50 Essential Concepts, O'Reilly Media, CA, USA, 2017.
4. Baron, M., Probability and Statistics for Computer Scientists, 3rd edition, CRC Press, Taylor and Francis, Boca Raton, FL, 2019.
5. Milton, J.S., Arnold, J. C., Introduction to Probability and Statistics: Principles and Applications for Engineering and the Computing Sciences, 3rd Edition. McGraw-Hill, New York, 1995.
6. Gentle, J. E., Elements of Computational Statistics, Springer-Verlag, New York, 2002.

9. Corroborating the content of the discipline with the expectations of the epistemic community, professional associations and representative employers within the field of the program

- The course follows the ACM and IEEE Curriculum Recommendations for studying Computer Science;
- The course exists in the studying program of all major universities in Romania and abroad;
- The knowledge and skills acquired in this course give students a foundation for launching a career in scientific research;
- The statistical analysis abilities acquired in this course are useful in any career path students may choose.

10. Evaluation

Type of activity	10.1 Evaluation criteria	10.2 Evaluation methods	10.3 Share in the grade (%)
10.4 Course	- acquire the basic principles in Statistics, with emphasis on real life applications; - be able to apply correctly the course concepts on various problems - be able to use statistical exploratory data analysis tools	**Written exam**	60%
10.5 Seminar/Lab activities	- be able to apply course concepts and techniques on practical problems - be able to solve numerical statistical problems in Excel and Matlab - be able to perform statistical analysis of data	- participation in discussing, solving and implementing problems throughout the semester - individual presentation of solutions - **presentation of a project of statistical analysis of data**	40%
10.7 Minimum performance standards			

A grade of 5 or above (on a scale from 1 to 10) on **each** activity mentioned above (written test, seminar/ lab evaluation)

Date	Signature of course coordinator	Signature of seminar coordinator
23.04.2023	Prof. Sanda Micula, PhD. Habil.	Prof. Sanda Micula, PhD. Habil.

Date of approval   Signature of the head of department

....…...........................