SYLLABUS

1. Information regarding the programme

1.1 Higher education institution	**Babeş Bolyai University**
1.2 Faculty	**Faculty of Biology and Geology**
1.3 Department	**Department of Molecular Biology and Biotechnology**
1.4 Field of study	**Biology**
1.5 Study cycle	**Master**
1.6 Study programme / Qualification	**Bioinformatics applied in life sciences**

2. Information regarding the discipline

2.1 Name of the discipline	**Basics of Statistics** **Bazele statisticii**						
2.2 Course coordinator	**Assoc. Prof. PhD. Habil. Sanda Micula**						
2.3 Seminar coordinator	**Assoc. Prof. PhD. Habil. Sanda Micula**						
2.4. Year of study	**1**	2.5 Semester	**1**	2.6. Type of evaluation	**E**	2.7 Type of discipline	**Elective**
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	**2**
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)					12
Preparation for seminars/labs, homework, papers, portfolios and essays					26
Tutorship					10
Evaluations					18
Other activities:					-

3.7 Total individual study hours	98
3.8 Total hours per semester	154
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. Introduction. Populations and samples, terminology. Declaring the objectives. Data collection. Statistical data mining techniques. Graphical display of data, pie charts, bar graphs.	• Interactive exposure • Explanation • Conversation • Description	
2. Exploring and visualizing data. Frequency and grouped frequency tables. Histograms, frequency polygons, stem-and-leaf plots.	• Interactive exposure • Explanation • Conversation • Description	
3. Descriptive Statistics. Parameters of a statistical distribution. Measures of central tendency. Measures of variability. Variance and standard deviation, interpretation. Coefficient of variation. Standard errors of estimates. Examples and applications.	• Interactive exposure • Explanation • Conversation • Description	
4. Percentiles, deciles, quartiles, interquartile range. Outliers, detection of outliers, the 3σ rule. Boxplots. Interpretation of data.	• Interactive exposure • Explanation • Conversation • Description	

5. Inferential Statistics. Notions of estimation theory. The Normal and Standard Normal distribution, Z-quantiles. The Student T-distribution and T- quantiles. Point estimators, basic properties.	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
6. One-sided and two-sided confidence intervals. Estimating the mean and the proportion by confidence intervals. Examples.	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
7. Selecting the sample size. Two-sample statistics, pooled proportion. Estimating the difference of proportions. Confidence intervals for paired data.	• Interactive exposure • Explanation • Conversation • Description	
8. 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. Examples.	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
12. Correlation and Regression. Two-sample statistics. Covariance and correlation coefficient. Scatter plots and time plots. Conditional mean. Curves of regression. Examples.	• Interactive exposure • Explanation • Conversation	
13. Method of least squares. Linear regression. Overfitting a model. Polynomial regression. Examples and applications.	• Interactive exposure • Explanation • Conversation • Description	
14. Fitting models. Univariate analysis of variance (ANOVA) and R-square. Prediction. 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, CRC Press, Taylor and Francis, Boca Raton, FL, USA, 2014.

<table>
<tr><td colspan="3">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.</td></tr>
<tr><td>8.2 Seminar /Laboratory</td><td>Teaching methods</td><td>Remarks</td></tr>
<tr><td>1. Introduction to Matlab (Octave) and Excel.</td><td>• Interactive exposure
• Explanation
• Conversation</td><td>The seminar is structured as 2 hours per week, every other week</td></tr>
<tr><td>2. Descriptive Statistics. Grouped frequency table. Computation of statistical measures in Matlab and Excel.</td><td>• Interactive exposure
• Explanation
• Conversation
• Individual and group work</td><td></td></tr>
<tr><td>3. Graphical display of data. Histogram, frequency polygon, boxplot in Matlab and Excel.</td><td>• Interactive exposure
• Conversation
• Synthesis
• Individual and group work</td><td></td></tr>
<tr><td>4. Confidence intervals. Interpretation of results.</td><td>• Interactive exposure
• Explanation
• Conversation
• Individual and group work</td><td></td></tr>
<tr><td>5. Hypothesis testing in Matlab and Excel. Interpretation of results.</td><td>• Interactive exposure
• Explanation
• Conversation
• Individual and group work</td><td></td></tr>
<tr><td>6. Correlation and regression.</td><td>• Interactive exposure
• Explanation
• Conversation
• Individual and group work</td><td></td></tr>
<tr><td>7. Presentation. A project of statistical analysis of data.</td><td>• Conversation
• Individual and group work</td><td></td></tr>
<tr><td colspan="3">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, CRC Press, Taylor and Francis, Boca Raton, FL, 2014.
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.</td></tr>
</table>

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** on problems and applications	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**
16.01.2023	**Assoc. Prof. PhD. Habil. Sanda Micula**	**Assoc. Prof. PhD. Habil. Sanda Micula**

Date of approval **Signature of the head of department**

Prof. dr. Laura Dioşan