# COURSE SYLLABUS

# APPLIED BIOSTATISTICS IN R

1. **Data about the program**

|  |  |
| --- | --- |
| 1.1 Higher education institution | Babeș-Bolyai University |
| 1.2 Faculty | Faculty of Biology and Geology |
| 1.3 Doctoral school | Integrative Biology |
| 1.4 Field of study | BIOLOGY |
| 1.5 Study cycle | Doctorate |
| 1.6 Study program / Qualification | Doctoral training / PhD in Biology |

1. **Course data**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2.1 Name of discipline | | | **Applied biostatistics in R** | | | | | | |
| 2.2 Teacher responsible for lectures | | | | | Dr. Fülöp Attila / Conf. Dr. László Zoltán | | | | |
| 2.3 Teacher responsible for seminars | | | | | Dr. Fülöp Attila / Conf. Dr. László Zoltán | | | | |
| 2.4 Year of study | 1 | 2.5 Semester | | 1 | | 2.6. Type of evaluation | C | 2.7 Course framework | O |

C – colloquium; O – optional

1. **Estimated total time of teaching activities** (hours persemester)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 3.1 Hours per week | 4 | | Out of which: 3.2 Lectures | | 2 | 3.3 Seminars / Laboratory classes | 2 |
| 3.4 Total hours in the curriculum | 48 | | Out of which: 3.5  Lectures | | 24 | 3.6 Seminars /  Laboratory classes | 24 |
| Allocation of study time: | | | | | | | hrs |
| Study supported by textbooks, other course materials, recommended bibliography and personal student notes | | | | | | | 64 |
| Additional learning activities in the library, on specialized online platforms and in the field | | | | | | | 64 |
| Preparation of seminars / laboratory classes, topics, papers, portfolios and essays | | | | | | | 38 |
| Tutoring | | | | | | | 34 |
| Examinations | | | | | | | 4 |
| Other activities: - | | | | | | | - |
| 3.7 Individual study (total hours) | | 204 | |  | | | |
| 3.8 Total hours per semester | | 252 | |
| 3.9 Number of credits | | 10 | |

1. **Preconditions** (where applicable)

|  |  |
| --- | --- |
| 4.1 Curriculum | * Not applicable |
| 4.2 Competences | * Basic knowledge of mathematics * User-level computing skills * Speaking and writing skills in English |

1. **Conditions** (where applicable)

|  |  |
| --- | --- |
| 5.1 Conducting lectures | * Audio-video logistics, whiteboard, access to WiFi internet |
| 5.2 Conducting seminars/laboratory classes | * Audio-video logistics, whiteboard, access to WiFi internet |

# Specific competences acquired

|  |  |
| --- | --- |
| **Professional competences** | * Competences in designing scientific studies, both on the field and in the laboratory. * Competences in collecting, analyzing and interpreting scientific data. * Programming and data analysis competences in the R statistical environment. |
| **Transversal competences** | * Competences in the statistical analysis of data, a requisite of scientific research in any field which involves a quantitative approach. |

1. **Course objectives** (based on the acquired competencies grid)

|  |  |
| --- | --- |
| 7.1 The general objective of the course | * To gain quantitative analytical skills of biological data in the R statistical environment. |
| 7.2 Specific objectives | * To learn to design scientific studies and to collect data in biology and ecology, both in the laboratory and on field. * To learn to analyze data using the adequate statistical tests, as implemented in the R statistical environment. |

# Content

|  |  |  |
| --- | --- | --- |
| 8.1 Lectures | Teaching methods  Hybrid teaching: onsite (40%) and online (60% - MS Teams, Zoom) | Comments |
| 1. Introduction:  - The scientific method  - Data, observations and variables  - Probability distributions | Presentation, discussion, case studies | 2 hours |
| 2. Estimation:  - Samples and populations  - Common parameters and statistics  - Standard errors and confidence intervals for the mean  - Methods for estimating parameters | 2 hours |
| 3. Graphical exploration of data:  - Exploratory data analysis  - Transforming data  - Standardization  - Outliers  - Censored and missing data | 2 hours |
| 4. Hypothesis testing:  - Statistical hypothesis testing  - Decision errors  - Multiple testing  - Critique of statistical hypothesis testing | 2 hours |
| 5. Non-parametric statistics:  - Comparing means  - Comparing variances  - Comparing frequencies | 2 hours |
| 6. Correlation and regression:  - Correlation analysis  - Simple and multiple linear regression analysis | 2 hours |
| 7. Design and power analysis:  - Sampling, experimental design | 2 hours |
| 8. Analysis of variance 1:  - Single factor (one-way) design  - Assumptions and diagnostics  - Factor effects | 2 hours |
| 9. Analysis of variance 2:  - Multi-factor design  - Repeated measures design | 2 hours |
| 10. Analyses of covariance:  - ANCOVA | 2 hours |
| 11. Generalized linear models and logistic regression:  - Logistic regression  - Poisson regression | 2 hours |
| 12. Multivariate analysis of variance  - Principal components analysis (PCA)  - Factor analysis (FA)  - Discriminancy analysis (DA) | 2 hours |
|  |  | Total: 24 hours |
| **Bibliography:** | | |
| (1) Quinn, G.P., Keough, M.K. (2002) Experimental design and data analysis for biologists. Cambridge Univ. Press, UK.  (2) Field, A., Miles, J., Field, Z. (2012) Discovering statistics using R. SAGE, UK.  (3) Crawley, M.J. (2013) The R Book, 2nd edition. John Wiley, UK. | | |
|  |  |  |
| 8.2 Seminars / laboratory classes | Teaching methods  Hybrid teaching: onsite (40%) and online (60% - MS Teams, Zoom) | Comments |
| 1. Introduction to R:  - Setting up the R environment (R and R Studio)  - Installing and loading packages  - Help and troubleshooting  - Introduction to the R language | Practical work. Problem-solving and discussion. | 2 hour |
| 2. Data types and structures in R:  - Operations with different data types and structures | 2 hours |
| 3. R programming:  - Conditionals, loops, and functions  - Debugging  - Organizing and commenting code | 2 hours |
| 4 and 5. Data manipulation, exploration and cleaning:  - Reading and writing data files  - Sorting, filtering, merging, cleaning data  - Preparing the data set for analyses | 4 hours |
| 6. Descriptive statistics and exploratory analysis:  - Distributions  - Exploratory plots and summary statistics  - Data transformation  - Missing values  - Handling outliers | 2 hours |
| 7. Non-parametric statistics:  - t, Wilcoxon, Welch, sign, binomial, Kruskall-Wallis, chi-square, Fisher tests  - Correlation analysis | 2 hours |
| 8 and 9. Linear regression:  - Simple and multiple linear regression  - Model checking  - Post-hoc comparisons  - Analysis of variance | 4 hours |
| 10. Multivariate statistics:  - PCA, FA, DA | 2 hours |
| 11. Plotting:  - Preparing plots for publications | 2 hours |
| 12. Reproducible research, open science tools for R:  - R Markdown  - git, GitHub, GitLab, etc. | 2 hours |
|  |  | Total: 24 hours |
| **Bibliography:** | | |
| (1) Crawley, M.J. (2013) The R Book, 2nd edition. John Wiley, UK.  (2) Ekstrøm, C.T. (2016) The R Primer, Second Edition. CRC Press, FL.  (3) Chang, W. (2013) R Graphics Cookbook. O’Reilly Media, CA.  (4) Gandrud, C. (2020) Reproducible Research with R and RStudio, Third Edition. CRC Press, FL.  (5) Wickham, H., Grolemund, G. (2017) R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. O’Reilly Media, CA. | | |

1. **Aligning the contents of the discipline with the expectations of the epistemic community representatives, professional associations and standard employers operating in the program field**

* The course has a similar content to courses from other European universities, and considers the level of training and abilities of doctoral students.
* The content of the course is regularly updated and incorporates the most novel approaches from the field of statistics.
* The course is fundamental for doctoral students, as quantitative skills are essential for scientific research activities.

1. **Examination**

|  |  |  |  |
| --- | --- | --- | --- |
| Activity type | 10.1 Evaluation criteria | 10.2 Evaluation methods | 10.3 Weight in the final grade |
| 10.4 Lectures | Assessment of knowledge | Colloquium (written) | 50% |
| 10.5 Seminars/laboratory classes | Skills in understanding and reviewing the latest  scientific information | Colloquium (written) | 50% |
| 10.6 Minimum performance standard | | | |
| * Knowledge of 50% of the information content of the course * Knowledge of 50% of the information content of the laboratory work. | | | |

Date of issue

21.09.2022

Signature of the teacher responsible for lectures

Dr. Fülöp Attila

Conf. Dr. László Zoltán

Signature of the teacher responsible for seminars

Dr. Fülöp Attila

Conf. Dr. László Zoltán

Date of approval by the doctoral school council Signature of the doctoral school director 20.09.2022 Prof. Pap Péter László