|  |  |
| --- | --- |
| **Title and Code** of the subject: **Statistics,** **MTB7028A** | **ECTS Credit Points: 3** |
| **Type** of the subject: **compulsory** / optional | |
| **Ratio of theory and practice: 50/50** (credit%) | |
| **Type and number of classes per semester**: 14 hour(s) lecture and 14 hour(s) practice per **semester**  Number of teaching hours / week : eg.:1+1 (lecture and practice) | |
| **Type of exam**: practical | |
| **Subject in the curriculum:** V. semester | |
| Preliminary requirements:- | |

|  |
| --- |
| **Summary of content - theory**: The students will get acquinted with the most modern measurement, data-logging, data processing and analyitical methods. They attain the skills of analyzing environmental data and huge databases on their own, as well as drawing technically proper consequences from these. |
| **Schedule:**   1. The aim and role of Statistics, basic concepts. Probability, independence, criterion, population, sample. Random sampling, systematic error, parameter. 2. Organizing data into a database, database management systems, characteristics of data, levels of measurement. 3. Determining the characteristic values of data belonging to different levels of measurement. Mean, variance, standard deviation, median, range, geometric mean, harmonic mean, variation coefficient. 4. Reports, OLAP cubes and pivot diagrams, OLAP Online Analytical Processing. 5. Normal distribution as a model. Cumulative probability and density function. Standard values and regularities of normal distribution. 6. Standard distribution, t- distribution, F- distribution, Chi-square distribution. 7. Hypothesis analyses. 8. General linear models (GLM) 1. 9. General linear models (GLM) 2. 10. Post hoc multiple comparisons for observed means. 11. Correlation analyses. Bivariate correlations. 12. Regression analysis, concept, bivariate linear regression. 13. Multiple linear regression, technics of initiating independent variables. 14. Nonlinear regression. Model expressions. Estimation method: sequential quadratic programming or Levenberg-Marquard algorithm. Parameter estimates and residual sum of squares. |
| **Summary of content - practice**: |
| Schedule:   1. Randomising. Collecting environmental data, sampling methods 2. Data transformation procedures: sort cases, select cases, aggregate data, merge files. 3. Mean, variance, standard deviation, median, range, geometric mean, harmonic mean, variation coefficient. Standard error of mean. Confidence intervals. 4. Data portrayal, graphs and other diagrams. 5. Standard values and regularities of normal distribution. Tests of normal distribution. 6. Standard distribution, t- distribution, F- distribution, Chi-square distribution. 7. . Compare means: u-test, one sample t-test, independent-samples t-test, paired-samples t-test, one-way ANOVA. 8. One-way models 9. Multifactoral (two and three factors) models 10. Post hoc multiple comparisons for observed means. 11. Bivariate correlations. Applicability conditions. 12. Bivariate linear regression. 13. Multicollinearity diagnostics, problem. 14. Sum of squares for regression, residual, uncorrected total and corrected total, parameter estimates, asymptotic standard errors, and asymptotic correlation matrix of parameter estimates. |
| **Literature, handbooks in English** |
| 1. Robert I. Kabacoff: R in Action, Second Edition Data analysis and graphics with R. May 2015 ISBN 9781617291388 608 pages printed in black & white 2. Joaquim P. Marques de Sá (2007): Applied Statistics Using SPSS, STATISTICA, MATLAB and R |
| **Competencies gained** *(acc. to the Regulation on training and outcome requirements)* |
| 1. **Knowledge:**  * xx * xx  1. **Skills:**  * xx * xx  1. **Attitude:**  * xx  1. **Autonomy and responsibility:**  * xx |

|  |
| --- |
| **Responsible lecturer: Dr. László Huzsvai** |
| **Other lecturer(s): ………..., ………………….** |

|  |
| --- |
| **Terms of course completion:** |
| 1. Completing assignments / exercises 2. Submitting essay 3. Giving presentation |
| **Form of examination:** |
|  |
| **Requirement(s) to get signature:** |
|  |

|  |
| --- |
| **Exam questions:** |
| 1. Describe the role of Statistics, the criterion of independence and the criteria of good sample. 2. Summarize the characteristics of the collection and sampling procedures of environmental data. List some modern data collecting procedures. 3. Describe the features and criteria of databases. 4. Explain the significance and aims of normalization steps taken during database designing. 5. Describe the main difference between listing, filtering, aggregating and merging databases. 6. List the measurement levels of data. Determine its standard values. 7. Describe the calculation methods of different mean values. List practical examples for mean values calculated by different methods. 8. Define the importance of variance, standard deviation, variational coefficient and the possible means of their calculation. 9. Describe the importance of determining the standard error of mean and the confidence range. List some means of putting them into practice. 10. What is the decision you would make if confidence ranges were to partially overlap, or were not overlap at all? 11. Summarize the importance of reports, pivot-diagrams and the technique of preparing them. 12. Describe the most important methods of data portrayal and their features. 13. Summarize what the significant differences are between the various methods of data portrayal. 14. Explain the statistical importance of normal distribution. Raise an example for its practical use. 15. Describe the importance and method of determining the distribution and density function of normal distribution. List its standard values. 16. Outline the most important normality analysis methods. Determine the applicability conditions. 17. Describe the features, expected values and practical importance of binomial distribution. 18. Summarize the importance and application field of t-distribution. 19. Describe the importance, parameters and application field of F-distribution. 20. Determine the features and parameters of Chi-square distribution. List some practical applications of it. 21. Describe the significance of hypothesis analyses. Explain the concept and probability of appearance of alpha error and beta error. 22. Determine the strength of a statistical test. Raise examples how the strength of the statistical test can be improved. 23. Explain the easiest mean value comparison tests and their applicability conditions. 24. Describe the conditions within which you would apply u-test, t-test and paired t-test. Raise some examples for their applications. 25. Summarize the importance of general linear models, thier most important concept and mathematical model. 26. Describe the structure of one-way linear model and the displaying of the effect of the factor. Summarize the applicability conditions. 27. Describe the stucture of the multifactoral (two and three factors) models and the displaying of the effect the factors. Summarize the applicability conditions. 28. List the main comparison tests of multiple mean values and describe their applicability conditions. 29. Outline the possibilities of compare analyses of variables belonging to different measurement levels and their applicability conditions. 30. Describe the process of correlation analysis in case of scale and ordinal data. 31. Demonstrate the application of multivariable linear regression and describe its applicability conditions. Summarize the problem of multicollinearity and heteroscedasticity. 32. Describe the means of portraying multicollinearity and heteroscedasticity. Describe the mistakes one can make during constructing the model if the applicability conditions are not realized. 33. List the main quasi nonlinear functions and their typical features. List some practical usage methods. 34. Describe the aligning techniques and applicability conditions of the real nonlinear regression. 35. Describe the measurement possibilities of the reliability and accuracy of the real nonlinear function alignments. |