

**COURSE UNIT (MODULE) DESCRIPTION**

|  |  |
| --- | --- |
| **Course unit (module) title** | **Code** |
| **COMPUTING AND DATA ANALYSIS** |  |

|  |  |
| --- | --- |
| **Lecturer(s)** | **Department(s) where the course unit (module) is delivered** |
| **Coordinator:** Dr. Vaidotas Zemlys-Balevičius  **Other(s):** | The Faculty of Economics and Business Administration |

|  |  |
| --- | --- |
| **Study cycle** | **Type of the course unit (module)** |
| 1st cycle (full-time studies) | Compulsory |

|  |  |  |
| --- | --- | --- |
| **Mode of delivery** | **Period when the course unit (module) is delivered** | **Language(s) of instruction** |
| Face-to-face | Semester 4 | English |

|  |  |
| --- | --- |
| **Requirements for students** | |
| **Prerequisites:** Statistical Theory (First year) | **Additional requirements (if any):** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Course (module) volume in credits** | **Total student’s workload** | **Contact hours** | **Self-study hours** |
| 5 | 130 | 36 | 94 |

|  |  |  |
| --- | --- | --- |
| **Aims of the course unit (module): programme competences to be developed** | | |
| **The aim** of the module – to develop undergraduate students' research skills in data handling, cleaning, analysis, visualisation and presentation by using statistical techniques and a number of computer software packages. | | |
| **Learning outcomes of the course unit (of the programme)** | **Teaching and learning methods** | **Assessment methods** |
| Students:  - will be able to use and undertake programming in the number of statistical software packages (R, MS Excel); (3.4) | Practicing to work with different data types and software packages both at university labs and at home. | Assignments for the lab sessions (4 assignments);  Final group project and its presentation |
| - will undertake basic cleaning of micro, macro and financial datasets and preliminary data description of those datasets; (1.2; 3.4) | Lab sessions, home-works. |
| - will undertake statistical analysis, hypothesis testing of datasets, visualisation of data, pattern recognition; (1.2; 4.2) | Theory lectures, examples, lab sessions, home-works. |
| - will write reports of their data analysis, distilling key insights and conclusions and effectively present their results in a group setup. (4.1; 4.2) | Work in group and individually on real economic and financial data, preparation of reports, their presentations. |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Content: breakdown of the topics** | **Contact hours** | | | | | | | **Self-study work: time and assignments** | | |
| Lectures | Tutorials | Seminars | Exercises | Laboratory work | Internship/work placement | Contact hours | | Self-study hours | **Assignments** |
| 1. Introduction and overview to data analysis, its visualization, main tools and software packages (R, Python, MS Excel). | 2 | 2 |  |  |  |  | 4 | | 6 | Reading lecture notes |
| 2. Particularities of economic and financial data. Cross sections, time series, and panel data, high frequency and big data sets. | 2 | 2 |  |  | 2 |  | 6 | | 12 | First introduction to different economic and financial data types and sets. Simple operations in the laboratory sessions. Assignment 1: Data downloading, reading, cleaning and preparing for analysis. |
| 3. Exploratory data analysis (EDA) using R and MS Excel: Data summaries; visualisation (basic plots, scatter, bar, line, box, histograms and distributions). Script running, coding. | 2 | 2 |  |  | 2 |  | 6 | | 22 | Assignment 2: Exploratory work with two different datasets, using both MS Excel and R. Prepare R script to visualise data. |
| 4.Further data analysis: descriptive statistics, correlation, regression, classification, model selection. Script running, coding. Examples in R. Interpretation of results. | 6 | 2 |  |  | 4 |  | 12 | | 28 | Assignment 3: use two datasets – one from economic growth literature (I Just Ran Two Million Regressions), another one from finance (VIX index). Produce codes and summary of both datasets.  Assignment 4: Replication of the published paper that intensively uses data-analytic techniques. |
| 5. Exporting results to the text processing software. Introduction to LaTeX, writing a report. Structuring and interpreting results. | 2 | 2 |  |  | 4 |  | 8 | | 26 | Preparation of the final data-analytic report. It includes: importing data, cleaning of the dataset, exploratory analysis, descriptive analysis, production of codes as well as final report with economic interpretation of the data patterns. |
| **Total:** | **14** | **10** |  |  | **12** |  | **36** | | **94** |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Assessment strategy** | **Weight,%** | **Deadline** | **Assessment criteria** |
| Final (group) data-analytic project and its presentation | 60 | The exam session | Quality of data analysis, depth and width of statistical and computational tools that are used to conduct analysis, clarity of results, clarity and efficiency of the code, presentation quality. |
| Four assignments during the semester, each worth 10% | 40 | During the semester | Number of correct answers, creative solutions, meticulous and concise presentation of results |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Author** | **Year of publication** | **Title** | **Issue of a periodical**  **or volume of a publication** | **Publishing place and house**  **or web link** |
| **Compulsory reading** | | | | |
| Garrett Grolemund and Hadley Wickham | 2017 | [R for Data Science](https://r4ds.had.co.nz/) | 1st Edition | O'Reilly Media |
| Christian Kleiber and Achim Zeileis | 2008 | [Applied Econometrics with R](http://www.springer.com/us/book/9780387773162) | 1st Edition | Springer-Verlag, New York |
| Roger D. Peng | 2015 | [R Programming for Data Sciences](https://leanpub.com/rprogramming) |  | Learnpub.com |