Syllabus

Summary

Name of Discipline:	Introduction to Data Analysis with Open Source Software
Teacher(s):	Székedi Levente, PhD
Contact(s):	szekedi.levente@partium.ro
Weekly Hours (course + seminar/lab/practice):	4 (2 course + 2 lab)
Credits:	10
Language of Instruction:	English
Year of Study (recommended):	At least undergraduate (BA) 2 or 3

Prerequisites

Students can come from a variety of backgrounds, though most examples and assignments use real social science datasets. An understanding of basic statistical concepts is important as the discipline has a practical and applied focus and we cannot deal with statistical or probability theory. Students enrolled in this course should have at least average digital competences, and be familiar with an office software package, file structure and operations etc. It is preferable, but not essential, for students to have an intermediate knowledge of a spreadsheet application (e.g. MS Excel) or a basic knowledge of statistical software (e.g. SPSS or PSPP). Students should bring their own laptop to the class.

Course Content

Abstract

The course aims to introduce students to basic concepts and techniques of data analysis, together with the skills to use two specialised software packages for manipulating, cleaning, analysing and visualising data. The programmes chosen are Jamovi and R, which are open source and free software of high quality, so that they can be used by students even after they have lost their university licence to various commercial alternatives. As Jamovi is a more conventional, menu-driven software, the bulk of the course is about using the R statistical software. The course is introductory in nature, and the lecturer will make recommendations about possible further learning paths for students. Data analysis is mostly focused on descriptive statistics, the relationship between the variables and hypothesis testing will be discussed only briefly. Real-world, freely available social science data will be used in the labs and assignments. The course is mostly linked to the methodology of the social sciences, as it also aims to develop critical thinking and problem solving skills in the context of social science research. Data analysis techniques will be used to address relevant research questions.

Topics

Week	Course (Lecture)	Lab
Week 1	Introduction to data analysis: a short review of the main concepts ● The role of data collection in the social sciences ● Types of data (quantitative, qualitative) ● Methods of data collection (offline and online surveys, other methods) ● Variables and levels of measurement Statistical software in the social sciences ● Commercial and free software ● The open source model The Jamovi statistical software ● Installation and interface	Research topics and recommended methodologies ● Students connect topics with methods and explain their choices. Exercises related to the levels of measurement ● Identify the levels of measurement in questionnaires ● Discussion on advantages and disadvantages of the various levels of measurement Commercial and open source ● Which software we use and why? (Classroom discussion) Download and install Jamovi on the laptops
Week 2	Data entry and import in Jamovi ● Defining variables ○ Setting variables and their levels of measurement ○ Text and numeric levels (aka value labels) ● Data import ○ Importing common formats (CSV, SPSS SAV etc.) Defining and handling missing data ● Types of missing data ● Dealing with missing data ● Handling missing data in Jamovi Data transformation ● Recoded and computed variables ● Variables transformations ● Computed variables	Data entry exercises ● Variable creation and data data entry Data import exercises ● Importing databases from free online sources of quality survey data in common formats (e.g. GESIS) ● Importing CSV files (online sources) Missing data exercises ● Defining missing data Data transformation exercises ● Transforming age to age groups ● Recoding satisfaction scales ● Calculating age from year of birth ● Standardisation
Week 3	A short review of the descriptive statistics ● Frequency tables ● Measures of central tendency and dispersion Basics of data visualisation ● Common chart types, their advantages and pitfalls ● The right choice of the charts Descriptive statistic in Jamovi ● Creating frequency tables ● Summarising the numeric variables ● Charts ● Exporting the results in various formats	Generating frequency tables ● Frequency tables for nominal variables ● Frequency tables for ordinal variables (after the reordering of levels) ● Frequency tables for recoded variables ● Exporting the tables and importing them in a spreadsheet app Calculating the measures of central tendency and dispersion Using the visualisation solutions of Jamovi ● Possibilities and limitations ● Exporting the charts
Week	Relationships between variables	Correlation is not causation

Week	Course (Lecture)	Lab
4	• Numerical variables (Correlation and regression) • Categorical variables (Crosstabs) Correlation and crosstabs in Jamovi On the concept of hypothesis testing • Null and alternative hypothesis • Common tests: T-test, Chi-square test, ANOVA Hypothesis testing in Jamovi Some final remarks on the Jamovi • Updating • Installing modules	• Calculating correlation coefficients in examples databases • Anscombe's quartet • Doing linear regression Analysing the relationship between nominal variables • Crosstabs and Chi-square test using example datasets • Recoding variables before crosstab Installing modules in Jamovi
Week 5	Introduction to R • A short history • R and the conventional statistical programs • The notion of reproducible research • Installing R and the RStudio interface • The R console • The R help system • Updating and installing packages The R environment • Objects in memory vs files	Resources on R • Resources on the The Comprehensive R Archive Network (https://cran.r-project.org/) • Online courses • Free books Installing R and RStudio The Rstudio interface and panels, projects • Creating projects The R console • Simple operations • Clearing the console, history, and other information Getting help Updating the base R Installing and updating the packages
Week 6	Basic data structures • Vectors of various types • Basic operations with vectors • Types, conversion and coercion • Special values and constants • Logical operations Vector indexing • Numerical indexing • Logical indexing Organising objects • Creating, listing, copying, removing objects	Vectors • Creating basic vectors • Vector creating functions • Checking types, classes, structure • Simple arithmetic operations • Type conversion on request and by coercion • Simple and complex logical statements Indexing • Vector indexing exercises • The powerful tool of logical indexing
Week 7	Augmented vectors, useful functions and operations • Object attributes • Names Sorting and ordering Unordered and ordered factors	Verifying and setting attributes • Displaying attributes • Setting attributes Rearranging vector elements, sorting, ordering, and ranking • Various sorting methods • Differences between sorting, ordering and ranking Creating and modifying factors • Defining unordered and ordered factors

Week	Course (Lecture)	Lab
		• Redefining, collapsing and dropping levels
Week 8	Matrices • Creating matrices • Matrix indexing • Basic operations and functions Lists • Homogenous and heterogenous data structures • List creation • Indexing and recalling list elements • Extending and simplifying Dataframes • Creating dataframes • Consulting the dataframe structure • Indexing • Renaming columns and rows • Changing, removing, and adding elements	Matrices • Creating numeric, character and logical matrices • Arithmetic and other operation • Transposing and combining Lists • Creating lists • Exercices with list indexing (simple and double brackets) • Flattening and combining Dataframes • Create dataframes from basic structures • Lists vs dataframes • Indexing • Dataframe modification • Simple operations
Week 9	Entering, importing and exporting data • Console input • Using built-in example data • Save and load objects • Importing structured text files • Importing SPSS SAV files Simple descriptive statistics • Frequencies for categorical data • Creating classes • Measures of central tendency and dispersion	Importing and exporting data • Listing and using the example dataframes • Saving and loading objects and the environment • Import delimited text files (comma separated, semicolon separated, tab separated) • Importing SPSS files, preventing and correcting conversion problems Simple descriptive statistics • Calculate summary statistics for various dataframes
Week 10	Scripting in R • The role of scripting in reproducible research • Creating and saving scripts • RStudio helpers Functions • An overview of the R functions • Function structure • Defining custom functions • Parameters, scoping, return values • Messages and error handling Conditionals • Conditionals as flow control mechanisms • Conditional statements in R	Scripting • Create, edit, and save a short script • Run end partially run the script • Comments • Best practices Functions • Experimenting with builtin functions • Create and run custom functions • Exercises helping the understanding of scoping Conditionals • Various exercises with conditional statements • Using conditionals to retrieve and change data frame content
Week	Loops	Loops

Week	Course (Lecture)	Lab
11	• The significance of loops • *For* loops • *While* loops • *Repeat* loops Higher order functions • About the higher order functions • The *apply* family • Other higher order functions • Loops or higher order functions?	• Understanding the difference between the different types of loops • Creating and running simple loops • Setting the conditions correctly • Exiting loops with special keywords Higher order functions • Understanding the difference between the *apply* variants through exercises • Concise dataframe operations with the *apply* functions • Using higher level functions for simulations
Week 12	Operations with strings • Clarifications: strings, characters, character vectors • String length • Combining and displaying strings • Conversion • Extraction, cleaning, trimming, splitting • Regular expressions • String search and replace Date and time operations • Date and time formats and objects • Simple operations	Strings • Calculating string length • String creation and modification • Combining, pasting strings • Useful string cleaning operations • Tips for humans on creating regular expressions • Using string matching for recalling data frame elements • Search and replace exercises Date and time • Converting date and time formats • Importing date and time columns • Calculating time difference
Week 13	Advanced data frame operations • Handling missing data • Subsetting, selecting, filtering, splitting • Combining and joining data frames • Aggregate functions Simple bivariate statistics in R • Correlation • Simple linear regression • Crosstabs and chi-square test	Advanced data frame operations • Exercised relate to missing data • Selecting columns and rows • Extracting parts of a data frame based on various conditions • Combining and joining • Summarising data frame through simple and aggregate functions Simple bivariate statistics • Calculate correlation and regression using builtin datasets • Implement chi-square tests on imported questionnaire dataset
Week 14	Basic data visualisation in base R • The R graphical system • Simple univariate dot and line plots • Creating simple pie charts • Creating simple bar charts • Visualising numerical data: histograms and boxplots • Exporting plots Recommended R packages • The *tidyverse* collection	Basic data visualisation • Understanding the R graphical system through exercises • Creating simple plots using builtin and imported data • Modifying plot title and other parameters • Exporting plots in various formats

Bibliography

- Navarro, D. J., & Foxcroft, D. R. (2022). Learning Statistics with Jamovi (Version 0.75). https://doi.org/10.24384/hgc3-7p15
- https://datalab.cc/jamovi/
- User guide, Jamovi. (n.d.). Retrieved 29 September 2024, from https://www.jamovi.org/user-manual.html
- Venables, W. N., Smith, D. M., & R Core Team. (n.d.). An Introduction to R. Retrieved 29 September 2024, from https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf
- Grolemund, G. (n.d.). Hands-On Programming with R. Retrieved 29 September 2024, from https://rstudio-education.github.io/hopr/

Learning Outcomes

Upon completion of the course, students will have acquired a fundamental understanding of two open-source software solutions employed in the analysis of social science data. The students will gain an understanding of the fundamental principles of data analysis and will be able to perform basic data import and cleaning operations in the aforementioned software. By the end of the course, students will be able to perform descriptive statistics, calculate and interpret measures of central tendency and dispersion, create frequency tables, histograms and other basic statistical charts. They will also have a grasp of the basics of hypothesis testing and be able to perform some hypothesis testing in both Jamovi and R. Overall, the course will enhance the students' ability to apply data analysis techniques to social research questions and to organise a data analysis and reporting process in a reproducible manner.

Assessment

The final grade consists of the following:

- Mid term assignment in Jamovi (15%)
- Mid term assignment in R (15%)
- Final assignment: 5 pages descriptive report on a dataset with data analysis done in Jamovi or R, with mandatory data visualisation (40%)
- Moodle quiz (30%)

Requirements to Pass

As the course has an applied character, a minimum 50% presence at the lectures, and a 75% presence at the lab is required. For obtaining the passing grade students must complete and pass 2 mid term assignments, and complete and pass the final assignment (report), and also pass a Moodle quiz. The passing grade for both the assignment and the quiz is 5.