

# Quantitative Methods of Empirical Social Research

210015 BAK 4

Winter 2021, Wednesday 15:00-16:30

## Overview

The aim of this course is to equip students with the basic applied skills for conducting easy data projects using the programming language R. The content of the course includes basic descriptive and inferential statistics as well as the graphic representation of results. The core focus of this course will be hands-on and practical.

## Prerequisites

This course is intended for bachelor level students. This course is complementary to the theoretical course “210014 VO BAK 4 Quantitative methods in the empirical social sciences (2020W)” taught by Professor Markus Wagner. Students are expected to attend the 210014 VO lecture component, which will cover theoretical concepts and more abstract ideas. Aside from that, no pre-existing knowledge of mathematics or programming is required.

## Learning objectives

- Students can conduct basic quantitative data analysis using the programming language R
- Students can produce elegant and informative graphs using the ggplot2 environment in R
- Students have acquired the mathematical literacy to proceed to more advanced courses in quantitative methods.
- Students are familiar and comfortable with R to an extent where they can learn to apply more advanced methods on their own.

- Students can critically evaluate strengths and weaknesses of theoretical arguments.
- Students can formulate methodological strategies for a viable independent research project e.g. a bachelor or master's thesis.

## Course assessment

The learning objectives of conducting basic quantitative data analysis using R are reflected in the graded assignments. Please upload all homework assignments and the final assignment in moodle by the due date. I strongly encourage putting considerable effort into homework assignments, since this is the best and most efficient way of mastering the material and will prepare you to write an excellent final paper. Do not miss the opportunity to get feedback before the final exam!

Homework assignments are uploaded on Moodle two weeks before the due date. Assignments should be completed using R markdown and uploaded as **one pdf file** on the due date (before midnight). Failure to submit the assignments on time will result in a one-unit deduction in your grade per 24 hours.

All assignments must be completed individually. The software turnitin will be used to check plagiarism.

The final assessment will be based on the following components:

1. **Participation** (10% of final grade) Regular attendance in class (maximum 2 classes can be missed). Participation will be assessed using the online quiz.
2. **Four homework assignments** (25% of final grade) based on materials in the course texts. Students are encouraged to form study groups but assignments must be completed individually.
3. **One mid-term exam** (25% of final grade). The test will concern theoretical questions and/or interpretation of R output. Duration: max 45 minutes.
4. **Final assignment** (40% of final grade). At the end of the course, you will be required to write a final paper of 2000-2500 words, focusing mostly on methods with applications in R. Joint work is NOT allowed for the final assignment. Deadline for handing in the final assignment: 16th February 2022.

## Weekly Responsibilities

Most importantly - students should download R Studio to their personal computer. Mastering this programming language is essential for passing the class. Downloading R is a two-step process: First download R for free at <https://cran.r-project.org>, then download the integrated development environment R Studio Desktop for free at <https://rstudio.com/products/rstudio/download/>.

Students should follow the online video instructions and complete the online exercises in R every week. Since no physical participation is strictly required, participation will be evaluated based on online quizzes. **Please complete the survey on Moodle every week after class, but before midnight.** Failure to reply by the deadline will be counted as a missed class.

## Trouble-shooting and office hours

Due to Covid-19 restrictions, this entire class will not be able to fit in the physical classroom assigned to us. Hence, the class will be taught online using Zoom. Please find the link at the top of the Moodle page.

However, I wish to offer the students, who believe they could benefit from it and opportunity to follow the online instruction from the classroom 3 ZID UniCampus Hof 7 (2H-O1-25). There are 10 seats available. **Please sign on in advance on Moodle (first-come-first-serve) and bring your own laptop.** I might be better able to help trouble-shoot any technical issues if students are physically present and we can look at their laptops together. Remember to follow the 3G rules (vaccinated, tested, recovered) and wear an FFP2-mask

I encourage students to ask any questions related to the course material or statistical programming in general in the discussion forums on Moodle. I have created discussion forums for each weekly topics as well as discussion forums for each homework. I realize that it might feel intimidating to post your question publicly, but I assure you **there are no stupid questions**. If you are wondering about something, there is a really good chance another student in the class has the exact same problem. As you continue to work with R, you will inevitably run into problems which we have not covered in class and where you cannot find the answer on google. Hence, learning to ask questions to the R community is necessary. I highly encourage students to look at each other questions and help out whenever possible. Please make sure to include screen shots or copies of your code.

For any questions related to course organization, grading, or sick leave, please feel free to email me at [ida.hjerimitslev@univie.ac.at](mailto:ida.hjerimitslev@univie.ac.at)

## Weekly Readings

The following readings are required

- Alan Agresti (2018). Statistical methods for the social sciences (5th edition). New Jersey: Pearson Education International
- Garrett Golemund and Hadley Wickham. R for Data Science. Available online here: <https://r4ds.had.co.nz/>
- James Long and Paul Teetor. R cookbook (2nd edition) Available online here: <https://rc2e.com/>

Suggested optional readings:

- Paul M. Kellstedt, and Guy D. Whitten. (2018). The fundamentals of political science research. Cambridge: Cambridge University Press
- Imai, K. (2018). Quantitative social science: an introduction. Princeton University Press.

**Week 1 (Oct. 6th): Introduction**

- R Cookbook "Getting Started and Getting Help". Section 1.2-1.3

**Week 2 (Oct. 13th): Basics of R**

- R Cookbook "Some Basics". Section 2.2-2.5 and 2.7-2.14 (i.e. skip section 2.6)
- R for Data Science "Workflow Basics". Section 4.1-4.3

**Week 3 (Oct. 20th): Locating and loading data - HW1 is due**

- Agresti (2018) "Chapter 1: Introduction" pp. 13-20
- R for Data Science "Tibbles". Section 10.1-10.4
- R for Data Science "Data import". Section 11.1-11.2 and 11.5-11.6

**Week 4 (Oct. 27th): Operationalization and measurement**

- Agresti (2018) "Chapter 2: Sampling and measurement" pp. 23-33 (section 2.1-2.3)
- R for Data Science "Data transformations". Section 5.1-

**Week 5 (Nov. 3rd): Descriptive statistics**

- Agresti (2018) "Chapter 3: Descriptive statistics" pp. 41-66 (section 3.1-3.5)
- R Cookbook "Computing Basic Statistics". Section 2.6
- Treiman, Donald (2009) "Cross Tabulations" in *Quantitative Data Analysis: Doing Social Research to Test Ideas* pp. 8-12 (Available in online view)

**Week 6 (Nov. 10th): Graphs - HW2 is due**

- R for Data Science "Data Visualisations". Section 3.1-3.5 and 3.7-3.8 (i.e. skip section 3.6)

**Week 7 (Nov. 17th): Sampling and populations**

- Agresti (2018) "Chapter 4: Probability Distributions" pp. 79-107 (Do not get confused by the subsection on bivariate probability distributions – we will come back to that in week 12).
- R Cookbook "Probability". Section 8.3-8.6 and 8.8-8.11 (i.e. skip section 8.7)
- R Cookbook "General Statistics". Section 9.5-9.7

**Week 8 (Nov. 24th): Confidence intervals**

- Agresti (2018) "Chapter 5: Statistical Inference: Estimation" pp. 115-132 (section 5.1-5.3)
- R Cookbook "General Statistics". Section 9.8-9.12 AND 9.15

**Week 9 (Dec. 1st): Hypothesis testing - HW3 is due**

- Agresti (2018) "Chapter 6: Statistical Inference: Significance Test" pp. 151-167 (section 6.1-6.3)
- Agresti (2018) "Chapter 7: Comparison of Two Groups" pp. 191-202 (section 7.1-7.3)

**Week 10 (Dec. 15th): Midterm Exam****Week 11 (Jan. 12th): Statistics for contingency tables**

- Agresti (2018) "Chapter 8: Analyzing Associations between Categorical Variables" pp. 227-237 and 239-251 (section 8.1-8.2 and 8.4-8.6)

**Week 12 (Jan. 19th): Linear regression and correlation - HW4 is due**

- Agresti (2018) "Chapter 9: Linear Regression and Correlation" pp. 259-284 (section 9.1-9.5)
- R Cookbook "Computing Basic Statistics". Section 2.6
- R Cookbook "Testing a Correlation for Significant". Section 9.17
- R Cookbook "Linear Regression and ANOVA". Section 11-11.4

**Week 13 (Jan. 26th): Summary and Q&A**