BAK4.2:
Quantitative Methoden der empirischen Sozialforschung

Winter 2020

Instructor:	Daniel Weitzel	**Time:**	Thu 13:15 – 14:45
Email:	daniel.weitzel@univie.ac.at	**Place:**	Online

Office Hours: After class, or by appointment, or post your questions in the forum provided for this purpose on Moodle.

Main References: This is a restricted list of various interesting and useful books that will be used during the course. You need to consult them. The two books are available online for free but can also be purchased as paperbacks.

- James Long and Paul Teetor. *R cookbook* (2nd edition), O'Reilly Media, 2019. Available as paperback and ebook or for free at `https://rc2e.com/`.
- Hadley Wickham and Garrett Grolemund, *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*, O'Reilly Media, 2017. Available as paperback and ebook or for free at `https://r4ds.had.co.nz/`
- Additional chapters and articles listed in the tentative course schedule will be made available in Moodle.

Objectives:

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. The aim of the course is to equip students with the basic applied skills for easy data projects. 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. Students are expected to attend the 210014 VO lecture component, which will cover theoretical concepts and more abstract ideas. Students will learn the basic tools to conduct quantitative data analysis using the programming language R. By the end of the course, students should be able to describe and manipulate a dataset and conduct basic inferential analysis using R in R Studio.

The examination will focus on different statistical concepts covered in class and will include basic data analysis using the programming language R. Detailed instructions about the homework assignments, midterm, and the final assignment will be posted on Moodle in due time.

Prerequisites: An undergraduate-level understanding of probability and statistics, as taught in "210014 VO BAK 4 Quantitative methods in the empirical social sciences", is assumed.

Grading Policy: Grading is based on four components. In order to complete the course with a positive grade students have to pass all four grading components with a positive assessment. In cases of suspected plagiarism you may be called upon to reasonably demonstrate that any work you have submitted is your own.

- Participation (10%): Participation will be assessed based on the submission of functioning and well-annotated R Scripts. These scripts will be following exercises in my lecture with occasional additional questions. Scripts for a given lecture must be submitted before the next lecture starts. Late submissions will not be accepted. A maximum of two participation exercises can be missed. For classes without R code brief explanations of core concepts will be required.
- Three Homework assignments (25%): The assignments are based on materials in the course. Students are encouraged to form study groups but assignments must be completed individually. Students must submit their R code and their text.
- Midterm (25%): The test will concern theoretical questions and/or interpretation of R output. Duration: max 45 minutes. Joint work is *NOT* allowed.
- Final (40%): 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: 31 March 2021.

Important Dates:

Homework #1	05.11.2020
Homework #2	03.12.2020
Midterm	10.12.2020
Homework #3	14.01.2021
Final Exam	31.03.2021

Note: Depending on progress of the class these dates are subject to change. Updated deadlines will be announced in the class lecture and on Moodle.

Software:

- We will be using R and R Studio in this class. R is a very powerful and capable free and open source programing language for statistical computing. R Studio is an excellent integrated development environment (IDE) for R that makes writing code and analyzing data a lot easier. Please follow the instructions here to install both R and R Studio.

Class Policy:

- Regular attendance is essential and expected. You can only miss two participation assignments.
- We will use the statistical programing software R and its graphical user interfaceR Studio in this course. Students have to submit their homework as a pdf and always include all their code. I encourage students to submit quantitative work based on Rmarkdown files.
- All homework and exam submissions must be through *turnitin* and you must submit your write up and your code.
- The material in the class can be challenging and difficult at first. It is hence of utmost importance that you follow the seminar closely, prepare for the class, do the homework, and always contribute to an open and engaging class environment. I strongly encourage students to ask questions. If you don't understand something or are stuck on a problem you are usually not alone.
- The Moodle page has a Forum for students to ask questions, help each other, and interact. I expect professional and courteous behavior that adheres to the rules of academic honesty.
- Due to the Coronavirus I expect that issues will arise over the course of the semester. The course is already designed to give students maximum flexibility. They have one week to submit their participation grades and sufficient time will be given for homeworks, midterm, and the final assignment. Please reach out to me as soon as possible (i.e. before the issue arises) if you encounter problems that might put your academic success in this class at risk.

Academic Honesty: Lack of knowledge of the academic honesty policy is not a reasonable explanation for a violation. Exercises and class papers will be examined from violations of the university's academic honesty policy.

Tentative Course Outline:

Session 1: Introduction to the course 08.10.2020
General overview, description of assignments, software, and readings.

Session 2: Basics of quantitative data analysis 15.10.2020
Agresti and Finlay (2009), Chapter 1, Chapter 2 pages 11-14

Session 3: Introduction to R and R Studio 22.10.2020
Nagler (1995) and blog post here
Monogan (2015), Chapter 1-2
Long and Teetor (2019), Chapter 1-3

Session 4: Data manipulation and descriptive statistics 29.10.2020
Agresti and Finlay (2009), Chapter 3 pages 31-55
Kellstedt and Whitten (2013), Chapter 5 pages 109-126
Long and Teetor (2019). Chapter 4
Wickham and Grolemund (2017), Chapter 5

Session 5: Descriptive statistics (univariate) 05.11.2020
Agresti and Finlay (2009), Chapter 3 pages 31-55
Monogan (2015), Chapter 4
Kellstedt and Whitten (2013), Chapter 5 pages 109-126
Long and Teetor (2019), Chapter 9

Session 6: Descriptive statistics (bi- and multivariate) 12.11.2020
Agresti and Finlay (2009), Chapter 3 pages 55-59
Monogan (2015), Chapter 4
Wickham and Grolemund (2017), Chapter 7

Session 7: Visualizing data 19.11.2020
Monogan (2015), Chapter 3
Long and Teetor (2019), Chapter 10
Wickham and Grolemund (2017), Chapter 3

Session 8: Populations and sampling 26.11.2020
Kellstedt and Whitten (2013), Chapter 6

Session 9: Confidence intervals and hypothesis testing 03.12.2020
Kellstedt and Whitten (2013), Chapter 6, and Chapter 7 (pages 145-156)
Wackerly et al 2008, Chapter 10

Session 10: Hypothesis testing and control variables 10.12.2020
Wackerly et al 2008, Chapter 10

Session 11: Midterm 17.12.2020

Session 12: Comparing means in two groups 07.01.2021
Kellstedt and Whitten (2013), Chapter 7 pages 145-161

Session 13: Covariance and correlation 14.01.2021
Agresti and Finlay (2009), Chapter 8

Session 14: Introduction to linear regression 21.01.2021
Agresti and Finlay (2009), Chapter 9
Monogan (2015), Chapter 6
Long and Teetor (2019) Chapter 11

Session 15: Summary, optional topics, & student questions ... 28.01.2021

Additional Literature: In case you want to go further.

1. More advanced textbooks about statistics
    - Imai, Kosuke, *Quantitative Social Science: An Introduction*, Princeton University Press, 2018.
    - Angrist, Joshua D. and Jörn-Steffen Pischke , *Mostly Harmless Econometrics: An Empiricist's Companion.*, Princeton University Press, 2009.
    - Morgan, Stephen L. and Christopher Winship , *Counterfactuals and Causal Inference*, Cambridge University Press, 2007.
    - Wackerly, Denis, William Mendenhall, Richard L. Scheaffer, *Mathematical Statistics with Applications*, Thomson Brooks/Cole, 2008.
    - Casella, George and Roger L. Berger, *Statistical Inference*, Cengage Learning, 2001. *(very advanced!)*
    - Healy, Kieran, *Data Visualization: A Practical Introduction.*, Princeton: Princeton University Press, 2019.
2. Fun reads
    - Mcgrayne, Sharon Bertsch, *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*, Yale University Press, 2012.
    - Silver, Nate *The Signal and the Noise: Why So Many Predictions Fail - but Some Don't*, Penguin Books, 2015.
    - Pearl, Judea and Dana Mackenzie, *The Book of Why: The New Science of Cause and Effect*, Basic Books, 2018.
    - Salsburg, David, *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century*, Holt Paperbacks, 2002.
3. Movies
    - *Moneyball*, a movie about the role of statistics in baseball.
    - *Behind the curve*, a documentary about the Flat Earth movement that highlights the importance of a rigorous scientific approach with falsifiable hypotheses.