Course title: Quantitative Methods 1.
Instructor: Michal Kotnarowski, Ph.D. IFiS PAN & GSSR
Academic year: 2024/2025
Planned course timetable:
Class: Mondays 14.00-16.00,
Office hours: Thursdays 14.00-16.00 (by prior appointment).
Teaching period: October 14, 2024 – January 27, 2025.

Format of the course: onsite class, room 268.

Course description

The course will cover basic and intermediate statistical techniques essential for data analysis across various fields. The content is designed to provide a firm grounding in the fundamentals of statistics while introducing more complex methodologies that will enable students to analyze diverse data sets effectively and critically. In addition, the course will cover topics related to causality, an essential aspect of statistical analysis that aims to determine not only associations between variables but also cause-and-effect relationships. The scope of the course includes the following key areas:

1. Descriptive Statistics:
Students will learn to summarize and describe the main features of a data set. The course will cover measures of central tendency (mean, median, mode), measures of variability (range, variance, standard deviation), and the construction of graphs and charts to visualize data distributions. Descriptive techniques allow for the initial exploration of data, which is a crucial first step in understanding trends, patterns, and potential areas of interest.

2. Statistical Inference:
The course will explore statistical inference, focusing on how we can make generalizations about a population based on sample data. Topics will include formulating and testing hypotheses, confidence intervals, p-values, and proper use of probability distributions (such as the normal distribution) to understand variability and uncertainty in sample data. Students will be trained to interpret sample results and make inferences about the broader population with a sound understanding of statistical significance and estimation techniques.

3. Univariate, Bivariate, and Multivariate Analysis:
- Univariate Analysis: The analysis of a single variable, where students will focus on summarizing and modelling individual data points to detect patterns within a single data set.
- Bivariate Analysis: This section will focus on the relationships between two variables. Students will explore methods like correlation and simple regression analysis to quantify the association and predict relationships.

4. Measurement:

Accurate and reliable measurement is critical in statistical analysis. This section will cover defining, collecting, and measuring variables appropriately in various contexts. Topics will include:

- Scale Types: The course will introduce different levels of measurement, including nominal, ordinal, interval, and ratio scales, and how they influence the choice of statistical methods.
- Measurement Error: Students will explore the implications of measurement error and biases, such as systematic error, and learn methods to minimize and account for errors in data collection.
- Validity and Reliability: The importance of ensuring validity (the degree to which a tool measures what it is supposed to measure) and reliability (the consistency of a measure across different conditions or points in time) will be emphasized.

5. Prediction:

Predictive analytics will be another component of the course, where students will learn how statistical models can be used to predict future outcomes based on existing data. It will involve:

- Regression Analysis: From simple linear regression to more complex multivariate models, students will learn how to predict the values of one variable based on one or more other variables.
- Model Validation: Methods to evaluate the accuracy and robustness of predictive models, such as cross-validation, residual analysis, and goodness-of-fit measures, will be discussed. Students will also learn to assess model overfitting and underfitting, ensuring that their predictions are both accurate and generalizable to unseen data.

6. Uncertainty in Statistical Reasoning:

A key feature of statistical analysis is dealing with uncertainty. Students will learn how to quantify, interpret, and make decisions in the face of uncertainty:

- Probability Theory: Basic principles of probability will be covered, enabling students to understand how likely an event is to occur and how to use probability in statistical models.
- Confidence Intervals and Margins of Error: The course will explain how to construct confidence intervals around estimates to express uncertainty, along with how to interpret the margin of error in survey and experimental data.
- Statistical Significance: The concept of statistical significance will be taught, including the interpretation of p-values, Type I and Type II errors, and the role of sample size in determining confidence in results.

7. Causality:

Understanding causality—distinguishing correlation from causation—is essential to statistical analysis. This course will include key concepts and methodologies for identifying and testing causal relationships in data:

- Correlation vs. Causation: Students will learn the distinction between simple correlations (relationships between variables that do not imply causation) and actual causal effects (where one variable directly influences another).
- Causal Inference: Techniques for establishing causal relationships will be explored. These include randomized controlled trials (RCTs), where random assignment is used to ensure the cause-and-effect relationship, and quasi-experimental designs (such as regression discontinuity designs) that attempt to mimic the rigour of experimental methods in observational studies.
- Counterfactual Reasoning: A key concept in causal analysis, students will be introduced to counterfactuals (what would have happened in the absence of a certain event) and how they are used to infer causality.

8. Data Types:

A key component of the course is the application of these statistical techniques to various types of data, including:

- Survey Data: Students will learn to analyze data collected through questionnaires, including techniques to handle sampling errors.
- Administrative Data: The course will also cover the analysis of data derived from administrative records, such as education or financial data, often involving larger data sets and requiring attention to issues like missing data or outliers.
- Experimental Data: Statistical methods applicable to data from controlled experiments, including randomized control trials, will be covered. Students will learn how to interpret results from experimental designs and handle challenges like treatment effects and experimental biases.

Course Objectives.

By the end of the course, participants will have mastered a broad range of statistical methods and their appropriate applications. They will develop the ability to critically evaluate data, conduct rigorous analyses, and present findings effectively. The course will foster the ability to:

- Select the correct statistical method based on the data type and research question.
- Conduct appropriate data transformations and clean data sets to ensure reliable results.
- Interpret outputs from statistical software and translate them into meaningful insights.
- Address measurement challenges, deal with prediction models, reason through uncertainty, and assess causal relationships with statistical rigor.

The course is also designed as a foundation for those who may wish to pursue advanced statistical techniques in the future. Whether the participant plans to engage in more complex machine learning methods, big data analytics, or specialized fields like time series analysis, econometrics, or causal inference, this course provides the essential knowledge needed to progress confidently. Emphasis is placed on building a grounded understanding of statistical logic and reasoning rather than just memorizing formulas or methods.

The teaching approach will blend theoretical understanding with practical application, ensuring that participants not only comprehend statistical concepts but can also implement them in real-world scenarios, especially in understanding and assessing cause-and-effect relationships.

Course Requirements

1. Class Attendance: Regular attendance is required. Participation in class discussions and activities is essential to understanding the course content. If a student is unable to attend a class, they must inform the instructor in advance. Consistent attendance and active participation will contribute to the final grade.

2. Assigned Readings: Before each class, students will be assigned specific readings that are crucial to the upcoming lecture and discussions. Students are expected to thoroughly read and understand the material before attending class. During class, students should be prepared to:
    - Discuss key concepts from the readings.
    - Summarize what they have learned.
    - Engage in a conversation about challenges, problems, or questions related to the assigned texts.

3. Practical Assignments in R: After almost every class, students will be given practical assignments to complete in R. These assignments are designed to reinforce the statistical techniques and concepts covered in the lectures. Completed assignments must be submitted the day before the next class. Timely submission of these tasks is crucial, as they will be reviewed and discussed during the following session.

Grading Criteria

1. Class Attendance and Active Participation – 20% of the final grade: It includes regular attendance, engagement in class discussions, and demonstrated understanding of the assigned readings. Active participation in discussions on the readings will be key to this part of the evaluation.

2. Timely Submission of Practical Assignments in R – 80% of the final grade: Almost every session will include a practical assignment in R. Completing and submitting these assignments on time is essential. These tasks are designed to apply the statistical techniques learned in class, and their proper execution will form the majority of the overall grade.

Table below provides a grading scheme:

A 94-100 5		**B+** 87-89 4+	**C+** 77-79 3+	**D+** 65-69 2+	**F** 59-0 1
A- 90-93 5-		**B** 84-86 4	**C** 74-76 3	**D** 60-64 2	
		B- 80-83 4-	**C-** 70-73 3-		

Policies on Attendance, Late Materials, and Make-Up

Attendance is mandatory, and counts toward your classroom participation and attendance grade. Please notify the instructor in advance if you are unable to attend a given class session.
As a rule, this course meets in person. Under exceptional circumstances, such as documented participation in a short-term study/research visit or as a presenter at a scientific conference, and after priorly discussing your situation with the instructors, it may be possible to attend up to three of the fifteen class sessions online.
All assignments must be turned in electronically via the Google Classroom portal by the specified deadlines. The instructor will accept late materials at his discretion and *only if* notified 24 hours before the deadline. Except for documented reasons (e.g., doctor's note), late assignments will incur a 10% penalty per day.

Course Information Platform

The primary platform for communication and information exchange between the instructor and students will be **Google Classroom**. All course-related materials will be posted there, including readings, presentations, links to external resources, and any additional materials.
Google Classroom will also be used for assigning homework. Students are required to submit their assignments via Google Classroom; the instructor will not accept homework submissions sent via email. It is important to regularly check the platform for updates and new materials to stay informed and prepared for the course.

Software Requirement: R and RStudio

Throughout the course, students will be using R and RStudio, which are powerful and free tools for statistical computing and data analysis. While R and RStudio offer extensive capabilities, they can present challenges, especially for those unfamiliar with programming. Students are required to install R and RStudio on their computers before the course begins to ensure they are ready for the practical assignments.

- R can be downloaded from: https://cran.r-project.org/
- RStudio can be downloaded from: https://posit.co/download/rstudio-desktop/

Please ensure both programs are installed and functioning properly before the first class. If you encounter difficulties during installation, reach out for support before the course starts.

Preliminary Course Outline

Day	**Topic**	**Student's own work after the class**
Oct 14,	1. Course Introduction, Introduction to R and RStudio	QSS Chapter 1, PSR Chapter 1, installation of R and RStudio
Oct 21	2. Crash R Course. Data import, data export, basic operations	LSR Chapter 3 -4.
Oct 28	3. Causality I. Casual effects and Counterfactuals, Randomized Controlled Trials	PSR Chapter 3, QSS Chapter 2. Sections 2.1-2.4
Nov 4	4. Causality II. Observational studies	PSR Chapter 4. QSS Chapter 2. Sections 2.5-2.6
Nov 18	5. Measurement I. Concepts of measuring.	PSR Chapter 5. QSS Chapter 3. Sections 3.1-3.5
Nov 25	6. Measurement II. Describing variables, measuring relationships	PSR Chapter 6. QSS Chapter 3. Sections 3.6-3.8
Dec 2	7. Prediction. Linear regression I.	QSS Chapter 4. Sections 4.1 – 4.2
Dec 9	8. Prediction. Linear regression II.	QSS Chapter 4. Sections 4.3 – 4.4
Dec 16	8. Probability 1. Samples and Sampling Distribution	PSR Chapter 7. QSS Chapter 6. Sections 6.1-6.2
Jan 7	9. Probability 2. Central Limit Theorem	QSS Chapter 6. Sections 6.3-6.5
Jan 13	10. Uncertainty 1.	PSR Chapter 8. QSS Chapter 7. Section 7.1 – 7.2
Jan 20	11.Prediction and uncertainty.	PSR Chapter 9. QSS Chapter 7. Section 7.3-7.4
Jan 27	12. Free week to catch-up	

References:

PSR = Kellstedt, Paul M., and Guy D. Whitten. 2018. *The Fundamentals of Political Science Research*. 3rd ed. Cambridge University Press.

QSS = Imai, Kosuke. 2017. *Quantitative Social Science: An Introduction*. Princeton Oxford: Princeton University Press.

LSR: Navarro, Danielle. 2019. *Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.6.1). https://learningstatisticswithr.com/book/*