



Department of Sociology

## **Introduction to Quantitative Data Management and Statistics (IQDMS)**

7.5 credits

Syllabus

The course “Introduction to quantitative data management and statistics” is at the advanced level for students who wish to acquire practical skills in the management of quantitative data and detailed planning of statistical analyses, as well as a working knowledge of basic statistics and data visualization. The course is particularly aimed at students who have little or no previous experience working with quantitative data and methods. Among the topics covered are: differences between types of data, responsible data management and the ethics of using data, basic programming, how to assess variables and write an analysis plan, as well as conduct descriptive analyses and present results in an intuitive manner. We use STATA software in the course, which offers students an important opportunity to master a key skill for quantitative research. The course will increase students’ ability to independently plan, execute and document data analyses: Are the data source and variables suitable for the research question? How can data management and documentation increase efficiency and accuracy? How can results be conveyed in the clearest manner? Throughout the course, students will have access to a pre-existing dataset. This dataset will be used during most of the course, but students are required to assess the appropriateness of data sources in the course.

The course is offered at half-time during the 1<sup>st</sup> and 2nd period in the Spring term. Entrance qualifications: Course participants should have completed a BA in social science.

Spring term: Half-time studies. Language: English



Department of Sociology

# **Course Syllabus:**

## **Introduction to Quantitative Data Management and Statistics**

### **1. General information**

The course consists of 7,5 ECTS credits and is offered at the advanced level in Sociology and Sociological Demography.

### **2. Decision**

### **3. Course code**

### **4. Entrance qualifications**

BA in social science.

### **5. Course organization**

The course is offered at half-time over 10 weeks. Course participants and instructors meet approximately twice a week for lectures, computer-based exercises and/or seminars. Course participants submit exercises and comment on peers' exercises in Athena and/or during seminars. The course is offered in English.

### **6. Course contents**

The course provides an introduction to quantitative data, management and basic statistics in terms of basic variable construction and data file management, documentation of variable construction and analysis procedures, analysis planning, and visual representation of analyses. Students are expected to develop abilities to independently plan, execute and document different stages of advanced quantitative data analysis as well as visually represent results.

### **7. Learning outcomes**

After having completed the course, students are expected to be able to:

*In terms of understanding types of data:*

- Identify the different constructions and purposes of micro and macro-level data, cross-sectional and longitudinal data, survey and register data
- Understand how to link questionnaires to data sets and identify filters
- Assess the limitations of data sources in relation to the timing of the information available

*In terms of documentation and ethics:*

- Document variable constructions, data file management, and analysis procedures in such a way that replication is facilitated
- Understand the significance of research ethics in quantitative data management

*In terms of data management:*

- Use STATA to manipulate and analyze data
- Construct variables based on one and/or several pre-existing variables
- Manipulate the shape and size of data sets by reshaping or merging and appending other data sources
- Clean the data of errors and inconsistencies
- Use survey weights for descriptive analyses

*In terms of analysis planning:*

- Judge whether a variable is suitable or not for a specific research question and understand the different types of variables
- Write an analysis plan containing operationalizations, basic variable construction and descriptive statistics, and a discussion of relevant methodological problems

*In terms of descriptive statistics and visualization:*

- Learn to reduce data complexity through descriptive visualization and statistics
- Make correct inferences from a sample to a population and understand confidence intervals
- Understand and use different visualization techniques that accurately represent the data and findings

## **8. Instruction and examination**

Course work consists of computer-based exercises, written and oral presentations, and lectures. Computer exercises will be individually completed. Examination consists of individually assembled assignments and a take-home, final assignment.

## **9. Criterion-referenced assessment**

Examination: Students are expected to complete all assignments independently.

The individual assignments following each lecture/lab and the peer review part of the final assignment are assessed as pass or fail. All examinations must receive at least a passing grade to achieve a final grade for the course.

The take-home final exam consists of three parts. The first part requires a description of the data used in the assignment, demonstrating understanding of data and documentation. The following descriptions are graded: sample selection (1), how the data was collected (2), the response rate (3), and the survey weights that are provided (4).

The second part consists of working with a numeric variable as an outcome, whereas the third requires working with a categorical outcome. The following aspects are graded: variable construction and recoding (5); data cleaning (6); descriptive statistics (7); inferential statistics (8); clear analysis plan that includes research question (9), description of sample selection (10), and hypotheses (11); interpretation of results in terms of statistical and substantive significance (12); visual representation of the relationship (13); and critical consideration of variables and analyses (14). The final exam will be accompanied by a do-file with Stata syntax that is clear and easy to follow (15).

The final grade is based on the 15 dimensions listed above, and each of these is graded as Good (2 points) points, Pass (1 point) and Fail (0 points). In addition, the timely submission of complete individual homework assignments and timely submission of a complete peer-review assignment also give 5 points each. The total number of points a student can receive is 40. The final grade is based on the following criterion:

- A (Excellent): 38-40 points
- B (very good): 35-37 points
- C (good): 32-34 points
- D (sufficient): 29-31 points
- E (pass): 20-28 points
- Fx (insufficient): failed one to four dimensions above
- F (fail): failed more than four dimensions above.

Students who have received a grade of Fx or F in an examination are entitled to take further examinations as long as the course is given to achieve at least a grade E. Students who have received at least a grade E in an examination may not take another examination for a higher grade. Students who have received a grade of Fx or F on an examination twice by an examiner have the right to request that another examiner determine the grade of the examination. This request must be made to the Director of Studies.

Examination takes place continuously during the course and the final take-home exam, which must be submitted by the specified deadline. Any late assignments must be submitted no later than one week after the end of the course in order to be examined during the course. If a student fails to comply with this deadline or submits at least one incorrect assignment, a new examination will not take place until the next examination session.

Plagiarism, cheating, and unauthorized collaboration or use of AI

Part of your responsibility as a student is to know the rules for examination. Detailed information can be found on Stockholm University's website [here](#). Teachers are obliged to report suspected cheating and plagiarism to the Vice-Chancellor and the Disciplinary Board. Plagiarism and cheating is always a disciplinary matter and can lead to suspension. An example of plagiarism is copying a text verbatim or almost

verbatim (including single sentences or lines of programming) and not indicating where it comes from. Always make sure that an acknowledgement and quotation marks are used when you submit text you did not write yourself. This even applies to texts you have written before (self-plagiarism). Cheating also includes, for example, the use of unauthorized aids, such as a mobile phone or generative AI, on a test. Having study groups together is stimulating and time-saving, but when it comes to examination assignments, you must be careful to work on your own (unless clearly stated otherwise), so as not to risk being counted as unauthorized collaboration. Generative AI, including ChatGPT, can be useful, for example in editing your own writing. However, the use of ChatGPT or similar tools in examination assignments without the examiner's explicit permission and without acknowledgement is considered cheating.

## 10. Literature

Articles and handouts

*Suggested reading*

Baum, C.F. (2016). *An introduction to STATA programming, 2<sup>nd</sup> Edition*. College Station, Tex.: STATA Press. (Available on reserve at the university library)

Long, J. S. (2009). *The Workflow of Data Analysis Using Stata*. Stata Press. (Available on reserve at the university library)

Miller, J. (2013). *The Chicago Guide to Writing about Multivariate Statistics, 2<sup>nd</sup> Edition*. Chicago: University of Chicago Press.

Mitchell, M. (2022). *A Visual Guide to Stata Graphics*. College Station, Tex.: STATA Press. (Available on reserve at the university library)

Salkind, N, & Frey, B. (2019) *Statistics for People Who (Think They) Hate Statistics, 7<sup>th</sup> Edition*. Los Angeles: Sage Publications.

Treiman, D. (2009). *Quantitative Data Analysis: Doing Social Research to Test Ideas*. John Wiley and Sons. (Available as an e-book from the university library.)

## 11. Schedule

B307					
9-12:00 unless otherwise noted					
					Lecturer
Thurs	18-Jan		L1		SB
Tues	23-Jan		L2		KK
Thurs	25-Jan		L3		SB
Thurs	1-Feb		L4		KK
Tues	6-Feb		L5		SB
Thurs	8-Feb		L6		KK
Thurs	15-Feb		L7		MW
Thurs	22-Feb		L8		KK
Thurs	7-Mar		lab	10am to 12pm	SB
Thurs	14-Mar		Final assignment presentation		SB

SB: Sunnee Billingsley

KK: Konstantin Kazenin

MW: Matthew Wallace