![Stockholm University logo](Stockholm University)

Department of Sociology

Introduction to Quantitative
Data Management and Statistics (IQMDS)

7.5 credits, Fall 2019

Syllabus

Last Updated: November 19, 2019

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. The course is offered within the Multidisciplinary Master's program in Demography at Stockholm University and 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 survey data, basic programming, how to assess variables and write an analysis plan, as well as conduct descriptive and uni/bivariate analyses. 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? 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 encouraged to begin looking for a data source to use in their thesis and are required to assess the appropriateness of data sources in the course. The course will be given before the "Quantitative Methods" course to prepare students for learning regression analyses and more advanced statistical methods using STATA, as well as before the course "Research methods – Master's thesis proposal", which will give them ample information and time to consider research questions and suitable sources for their thesis.

The course is offered at full-time during the 4th period in the Fall term. Entrance qualifications: Course participants should have completed a BA in social science.

Winter term: Full-time studies. Language: English

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 Demography.

2. **Decision**

3. **Course code**

4. **Entrance qualifications**

   BA in social science and a plan to write their master's thesis with a quantitative orientation.

5. **Course organization**

   The course is offered at full-time over 5 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 MONDO 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 conducting descriptive and uni/bivariate analyses. The course will be given before the "Quantitative Methods" course to prepare students for learning regression analyses and more advanced statistical methods using STATA, as well as before the course "Research methods – Master's thesis proposal", which will give them ample information and time to consider research questions and suitable sources. Students are expected to develop abilities to independently plan, execute and document different stages of advanced quantitative data analysis.

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 basic statistics:
- Describe variables in terms of distribution, percentages, mean, median and variance
- Conduct bivariate analyses
- Make correct inferences from a sample to a population and understand confidence intervals

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 an analysis plan.

9. Criterion-referenced assessment

Examination: Students are expected to complete individual assignments independently, which accounts for 50% of their grade. Because each exercise builds on skills learned in previous exercises, timely completion of each exercise is imperative to ensure students are successfully and cumulatively broadening their skills. To accommodate different learning curves in computer programming and statistical proficiency, students are given full points on each assignment if they submit

the exercise on time and demonstrate that they put effort into answering the questions using the tools taught in the accompanying lecture. The remaining 50% of their grade is based on their final assignment, which includes an analysis plan and codebook that is based on a list of specific components.

The student's achievement is evaluated according to the following criterion-referenced assessment:

A = Excellent
B = Very good
C = Good
D = Satisfactory
E = Sufficient
Fx = Insufficient
F = Fail

A = Excellent. The student completed all individual exercises on time and demonstrated knowledge of the skills taught in the accompanying lecture. The student also demonstrated excellent practical and analytical skills in addressing each component of the analysis plan and codebook by including basic variable construction and data file management, clear documentation of variable construction and analysis procedures, analysis planning, and conducting and evaluating descriptive and uni/bivariate analyses.

B = Very good. The student completed all individual exercises mostly on time and managed to use the skills taught in the accompanying lecture. The student also demonstrated practical and analytical skills in addressing most components of the analysis plan and codebook.

C = Good. The student completed all individual exercises by the end of the course. The student also managed to successfully comply with at least half of the components in the analysis plan and codebook.

D = Satisfactory. The student completed all individual exercises by the end of the course. The student also managed to successfully comply with less than half but more than 25% of the required components in the analysis plan and codebook.

E = Sufficient. The student completed all individual exercises by the end of the course. The student demonstrated capacity to use skills taught in the course, but did not manage to successfully comply with more than 25% of the required components in the analysis plan and codebook.

Fx = Insufficient. The student has not successfully completed all exercises or the analysis plan and codebook.

F = Fail. The student has not completed most of the exercises

E is required to pass the course. Students with the grade Fx are offered the opportunity to upgrade his or her grade. Students with the grade F or Fx are entitled to another examination as long as the course is provided in order to achieve at least grade E.

10. Literature

Articles and handouts

Suggested reading

Long, J. S. (2009). *The Workflow of Data Analysis Using Stata*. Stata Press

Baum, C.F. (2009). *An introduction to STATA programming,* College Station, Tex.: STATA Press.

Treiman, D. (2009). Quantitative Data Analysis: Doing Social Research to Test Ideas. John Wiley and Sons.

STOCKHOLM UNIVERSITY
Department of Sociology

Schedule: **Introduction to Quantitative Data Management and Statistics, 7.5 ECTS-credits, Fall 2019**

Meeting	Date, Time, Room	Topic	Assignments
1	05/12/19, Thursday 09-12 in B389	Stata + Univariate Statistics (1)	Assignment 1
2	06/12/19, Friday 13-16 in B389	Stata + Univariate Statistics (2)	Assignment 2
3	09/12/19, Monday 13-16 in B389	Variable Operationalization	Assignment 3
4	11/12/19, Wednesday 13-16 in B389	Data storage and cleaning	
5	13/12/19, Friday 13-16 in B389	Manipulating the data set (1)	Assignment 4
6	16/12/19, Monday 13-16 in B389	Manipulating the data set (2)	Assignment 5
7	18/12/19, Wednesday 13-16 in B389	Bivariate statistics (1)	Assignment 6
	19/12/18	**Deadline 1: Submit Final Project Topic**	
8	20/12/19, Friday 09-12 in B389	Bivariate statistics (2)	Assignment 7
9	08/01/19, Wednesday 09-12 in B389	Survey weights and advanced commands	Assignment 8
10	10/01/19, Friday 13-16 in B389	More on advanced commands	
	15/01/19	**Deadline 1: Submit Final Project Topic**	
11	17/01/19, Friday 09-12 in F487	Peer-Review	

Note: Schedule is subject to change. Changes from the first published schedule are marked **red**.