8th GESIS Summer School in Survey Methodology
Cologne, August 2019

Syllabus for the Short Course C: "Introduction to Data Analysis Using Stata."

Instructors:	Dr. Kathrin Busch	Julia Klinger
E-mail:	Kathrin.Busch@gesis.org	Klinger@wiso.uni-koeln.de
Homepage:	www.gesis.org	www.iss-wiso.uni-koeln.de

Date: 01.-02. August 2019
Time: 09:00-13:00, 14:00-16:00
The course starts on Thursday morning at 09:00

About the Instructors:

Kathrin Busch is the scientific coordinator of the GESIS Method Seminar. She owned her Ph.D. in political science with studies on the question of how voting rationally is influenced by voters' knowledge about parties' positions. Currently, she is a research assistant in the BMBF funded project SOLIKRIS, in which she explores the change of political attitudes and voting behavior of the European Youth during the years of the economic and financial crises.

Julia Klinger is a scientific associate at the Institute of Sociology and Social Psychology (ISS) at the University of Cologne. She is responsible for data collection coordination and data management of a migrant survey. Before, she was a scientific associate at the Research Data Center ALLBUS at GESIS where she was preparing regional and small-scale data. Her research interests contain migration, integration and social inequality.

Selected Publications:

- Busch, Kathrin B. (coming forward): No Shortcut to Voting. The Limited Influence of Parties' Left-Right Positions on Voting Behavior. (Dissertation, online publication, 2018).
- Busch, Kathrin B. 2016. "Estimating parties' left-right positions: Determinants of voters' perceptions' proximity to party ideology." Electoral Studies 41 159-178. doi: dx.doi.org/10.1016/j.electstud.2016.01.003.
- Klinger, Julia, Stefan Müller, and Merlin Schaeffer. 2017. "Der Halo-Effekt in einheimisch-homogenen Nachbarschaften: Steigert die ethnische Diversität angrenzender Nachbarschaften die Xenophobie in Deutschland?" Zeitschrift für Soziologie 46 (6): 402–419. doi: dx.doi.org/10.1515/zfsoz-2017-102.

Short Course Description:

This course will give a thorough introduction to the Statistics Software Stata. It is tailored to the needs of academics and other research practitioners who are new to Stata or who wish to refresh their skills.

In the first part of the course, we will cover the program's interface and give an introductory to its basic syntax structure. Also, we will review available help and support features (online and offline). Subsequently, we will provide you with skills in hands-on data management, common data analyses, and the visualizing of results. Special attention will be devoted on how to use Stata within an integrated workflow and how to obtain publication-ready results (tables and graphs) from Stata. If time permits and if participants are familiar with basic procedures, we will advance to special topics, such as understanding and interpreting interactions in linear and non-linear models (e.g., logit models), or managing of longitudinal or hierarchical data. These topics will be selected according to the participants' main interests and needs.

The course will be interactive and case-based and use survey data from the field of Social Sciences. All necessary materials will be provided. By the end of this course, participants will be able to effectively use Stata for common statistical procedures and know how to go about more advanced applications.

Keywords:

Stata, data management, univariate statistics, bivariate statistics, regression

Course Prerequisites:

A basic understanding of statistics.

Target Group:

Participants will find the course useful if they are:

- users who are new to statistical computing with Stata
- advanced academics and research professionals who are familiar with other software
- former Stata users who want to refresh their knowledge

Course and Learning Objectives:

By the end of the course participants will:

- be familiar with Stata's interface and facilities;
- understand how to integrate Stata into their research process to efficiently create reproducible and publication-ready results;
- know how to solve common data management problems and how to document all modifications of the data;
- be able to perform typical descriptive and inferential statistical procedures and use graphs to communicate their results effectively;
- have transferred the covered topics to their research problems;
- know how to proceed from here and how to get additional support if needed.

Organizational Structure of the Course:

The course is a full-time course, consisting of 7 hours of interactive group instruction per day (including a one-hour lunch break). The course will alternate between short lectures and hands-on exercises on the PC. The two lecturers will be available to the students at all time.

Software and Hardware Requirements:

None. Participants will be provided with access to Stata via laptops/desktop PCs by GESIS.

Long Course Description:

Stata has become one of the most popular and flexible statistical software packages for quantitative data analysis in the social sciences, due to its combination of an extensive and growing set of core procedures and capabilities maintained by StataCorp with an open and adaptable architecture that is continuously expanded by a very active user community.

The goal of this two-day course is to familiarize users with the fundamental makeup and mechanics of Stata. After completing the course, participants will be proficient in standard statistical procedures (including uni-, bi- and multivariate statistics). The emphasis of this course is on software training rather than statistical instruction.

On the first day, we will focus on getting to know the Stata software, including its interface, basic syntax, and help features. These include Stata's native help as well as web-resources. We will then explore different ways to import data into Stata, learn how to prepare data for subsequent analyses, conduct uni- and bivariate analysis, and use Stata facilities to produce publication-ready tables, and graphs.

On the second day, we will proceed with multivariate analyses. We will focus on linear and logistic regression models and how to interpret and (graphically) display interactions. Again, we will show how to obtain publication-ready tables and graphs from Stata. If time permits, we will also introduce the participants to advanced topics, such as management and analysis of hierarchical or longitudinal data.

At the end of the second day, participants will have time to apply the covered topics to their research and pursue individual projects. Instructors will be present for support.

Day-to-day Schedule and Literature:

Day	Topic(s)
1	▪ **Getting started with Stata** ▪ **Importing data & data management** ▪ **Uni- and bivariate statistics** ▪ **Exporting tables to Word and Excel** ▪ **Graphs**
	Suggested reading (suggested, yet do not have to be read before the session): ▪ Acock, Alan C. (2016). A Gentle Introduction to Stata. 5th ed. Stata Press. Chapter 1-7.
2	▪ **Multivariate statistics** ▪ **Linear regression** ▪ **Logistic regression** ▪ **Individual projects & free study time**
	Suggested reading: ▪ Acock, Alan C. (2016). A Gentle Introduction to Stata. 5th ed. Stata Press. Chapter 8, 10-11, 15.

Preparatory Reading:

The StataCorp YouTube channel provides some excellent preparatory material. Participants are especially encouraged to browse the "Tour of the Stata 15 interface", "Quick help in Stata," and any other of the available material according to personal interest.

Also, the UCLA's website is a great resource both for learning Stata as well as a reference later on.

Additional Recommended Literature:

"The Workflow of Data Analysis Using Stata" by Long (2009) takes a holistic perspective on the research process and provides profound recommendations on organizing and conducting your investigation, accessible for the novice in quantitative research.