

10th GESIS Summer School in Survey Methodology
[2nd Virtual GESIS Summer School]
28 July – 20 August 2021

Syllabus for Short Course A:
Introduction to Stata for Data Management and Analysis

Instructors:	Nils Jungmann, M.A.	Anne-Kathrin Stroppe, M.A.
E-mail:	nil.jungmann@gesis.org	anne-kathrin.stroppe@gesis.org
Homepage:	https://www.gesis.org/	https://www.gesis.org/

Date: 28-30 July 2021
Time: 10:00-12:30 and 13:30-15:00
Time zone: CEST/CEDT, course starts on Wednesday at 10:00 am
Venue: Online via Zoom

About the Instructors:

Nils Jungmann is a doctoral researcher at GESIS where he is working as a data curator for the German Longitudinal Election Study (GLES). Trained in political science and political communication at the Universities of Mannheim and Düsseldorf, he is interested in understanding the cognitive processes underlying voter decision making and aims to do so by employing experimental methods.

Anne-Kathrin Stroppe is a doctoral researcher at GESIS where she is working as a data curator for the German Longitudinal Election Study (GLES). She is a political scientist who graduated with a M.A. degree from the University of Konstanz. As part of her doctoral thesis, she studies the influence of accessibility to public infrastructure on political attitudes and behaviour. Her methodological focus lies thereby on linking survey and geodata to answer social science questions.

Selected Publications:

- Jungmann, Nils, Ina Bieber, Manuela Blumenberg, and Konstantin Glinitzer. (forthcoming). „The Push and Pull of Political Leaders: Changing Candidate Evaluations and Vote Switching between the 2013 and 2017 Federal Elections.” In *The Changing German Voter*, edited by Rüdiger Schmitt-Beck. Oxford: Oxford University Press.
- Glinitzer, Konstantin, and Nils Jungmann. 2019. "Spitzenkandidaten." In *Zwischen Polarisierung und Beharrung: Die Bundestagswahl 2017*, edited by Sigrid Roßteutscher, Rüdiger Schmitt-Beck, Harald Schoen, Bernhard Weßels, and Christof Wolf, *Wahlen in Deutschland* 3, 247-262. Baden-Baden: Nomos.
- Gummer, Tobias, and Anne-Kathrin Stroppe. 2019. "Regierungs- und Parteileistung." In *Zwischen Polarisierung und Beharrung: Die Bundestagswahl 2017*, edited by Sigrid Roßteutscher, Rüdiger Schmitt-Beck, Harald Schoen, Bernhard Weßels, and Christof Wolf, *Wahlen in Deutschland* 3, 279-294. Baden-Baden: Nomos

Short Course Description:

This course will give a thorough introduction to the 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. The course will not cover basic statistical methods and their underlying mathematics but how to apply these methods using Stata.

In the first part of the course, we will cover the program's interface and introduce its syntax structure and basic rules to write clean and reproducible Stata code. Subsequently, we will provide you with skills in hands-on data management, common data analyses, and the visualization of results. Also, we will review available help and

support features (online and offline) to equip participants with the necessary knowledge to further develop their skills and solve occurring problems.

Keywords:

Stata, data management, data analysis, visualization

Course Prerequisites:

- Familiarity with quantitative data
- Basic knowledge of uni- and bivariate statistics (e.g. descriptive statistics, basics of regression analysis)
- Knowledge of other syntax-based software is helpful but not required

Target Group:

Participants will find the course useful if:

- they are new to statistical computing (with Stata);
- they are familiar with other statistical software but want to get to know Stata;
- they have already worked with Stata before but want to refresh basic 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 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;
- know how to proceed from here and how to get additional support if needed.

Organizational Structure of the Course:

The course is offered as an expanded short course consisting of 12 hours of interactive group instruction, spread over three days. The course will alternate between short lectures and hands-on exercises. The two lecturers will be available to the students at all time and support work on assignments in breakout rooms.

Software and Hardware Requirements:

Participants need a laptop/desktop computer that enables them to access the internet and smoothly work with Stata. Participants will be provided with access to Stata licenses by GESIS but must install the software prior to the course on their own devices.

Long Course Description:

This introductory course aims to enable participants to efficiently employ the statistics software Stata to read, modify, and analyse data. This includes learning how to open your data in Stata, how to get a first grasp of your data using descriptive statistics, how to efficiently prepare your data for analysis using data management commands, how to analyse your data to answer basic research questions using bi- and multivariate statistics, and, finally, how to visualise your results and get them ready for publication using tables and graphics. We will accomplish all of this by writing code in Stata that documents all our steps and makes them comprehensible and reproducible.

The secret to finding your way around a program like Stata is practice and writing your own code. The focus of this course will therefore be to provide the participants with a hands-on experience in working with Stata. The lecturers will give introductory lectures on all topics which will be alternated with practical exercises.

As the lecturers are political scientists and work for the German Longitudinal Election Study (GLES), the different possible applications of Stata will be demonstrated using GLES survey data. However, it will also be possible for participants to apply the acquired knowledge to their own projects and questions.

This course is no introduction to statistical methods, and we will therefore not spend time on the statistical foundations of the methods we will use but, rather, we will "get our hands dirty" and apply these methods directly in Stata. To be able to easily follow this course's content, knowledge of uni- and bivariate statistics is required. Furthermore, the course will not cover the broad array of statistical methods (e.g., SEM, CFA/ECA, ...) but concentrate on bivariate statistics and regression modelling.

Day-to-day Schedule and Literature:

Day		Topic(s)
1	10.00 – 12.30	Introduction to the Stata interface and syntax language
	12.30 – 13.30	Lunch Break
	13.30 – 15.00	Getting to know your data using descriptive statistics
	<u>Suggested reading (optional):</u> <ul style="list-style-type: none"> ▪ Acock, Alan C. 2018. Getting Started. In A Gentle Introduction to Stata, 6th Ed. College Station, TX: Stata Press, pp. 1-19. ▪ Acock, Alan C. 2018. Working with commands, do-files, and results. In A Gentle Introduction to Stata, 6th Ed. College Station, TX: Stata Press, pp. 77-92. ▪ Kohler, Ulrich, and Frauke Kreuter. 2012. The first time. In Data Analysis Using Stata, 3rd Ed. College Station, TX: Stata Press, pp. 1-24. 	
2	10.00 – 12.30	Data management and data manipulation
	12.30 – 13.30	Lunch Break
	13.30 – 15.00	Data management and data manipulation
	<u>Suggested reading:</u> <ul style="list-style-type: none"> ▪ Acock, Alan C. 2018. Preparing data for analysis. In A Gentle Introduction to Stata, 6th Ed. College Station, TX: Stata Press, pp. 51-67. ▪ Kohler, Ulrich, and Frauke Kreuter. 2012. Creating and changing variables. In Data Analysis Using Stata, 3rd ed. College Station, TX: Stata Press, pp. 77-114. 	
3	10.00 – 12.30	Bi- and multivariate statistics
	12.30 – 13.30	Lunch Break
	13.30 – 15.00	Visualization of results
	<u>Suggested reading:</u> <ul style="list-style-type: none"> ▪ Kohler, Ulrich, and Frauke Kreuter. 2012. Introduction to linear regression. In Data Analysis Using Stata, 3rd Ed. College Station, TX: Stata Press, pp. 253-340. ▪ Kohler, Ulrich, and Frauke Kreuter. 2012. Creating and changing graphs. In Data Analysis Using Stata, 3rd Ed. College Station, TX: Stata Press, pp. 115-156. 	

Preparatory Reading:

- Agresti, Alan. 2018. Statistical Methods for the Social Sciences. 5th Ed. Harlow: Pearson.
- 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: <https://www.youtube.com/user/statacorp>

Additional Recommended Resources:

- Cameron, A. Colin, and Pravin K. Trivedi. 2010. *Microeconometrics Using Stata*. College Station, TX: Stata Press.
- Kohler, Ulrich, and Frauke Kreuter. 2012. *Data Analysis Using Stata*. 3rd Ed. College Station, TX: Stata Press.
- Mehmotoglu, Mehmet, and Tor Jakobsen. 2017. *Applied Statistics Using Stata. A Guide for the Social Sciences*. London: Sage.
- Cox, Nicholas, and Joseph Newton. 2014. *One Hundred Nineteen Stata Tips*. 3rd Ed. College Station, TX: Stata Press.
- Cheat Sheets from the official Stata website:
- <https://www.stata.com/bookstore/statacheatsheets.pdf>
- Cheat Sheets by Tim Essam and Laura Hughes:
- https://geocenter.github.io/StataTraining/portfolio/01_resource/
- Free online workshops, analysis walk-throughs, annotated syntax etc. from the Institute for Digital Research and Education at UCLA: <https://stats.idre.ucla.edu/stata/>