

# GESIS Summer School in Survey Methodology

## Syllabus for course:

### "Introduction to Stata for Data Management and Analysis"

Lecturers: Katharina Blinzler, M.A.  
E-mail: [katharina.blinzler@gesis.org](mailto:katharina.blinzler@gesis.org)  
Homepage: <https://www.gesis.org/en/institute/staff/person/Katharina.Blinzler>

Klara Dentler, M.A.  
[klara.dentler@gesis.org](mailto:klara.dentler@gesis.org)  
<https://www.gesis.org/en/institute/staff/person/Klara.Dentler>

Date: 03.-05. August 2022

Time: 10:00-12:00 + 13:00-15:00 CEST

Venue: Online via Zoom

## About the Lecturers:

*Katharina Blinzler* is a data processing specialist for the Comparative Study of Electoral Systems (CSES) at GESIS Köln. Katharina holds a bachelor's and master's degree in political science from the University of Mannheim, Germany, and joined CSES in April 2018. Her research interests are comparative politics and EU politics. Before joining the CSES Secretariat, Katharina worked as a data processing specialist for the German Longitudinal Election Study (GLES) at GESIS in Köln.

*Klara Dentler* is a doctoral researcher working for the Comparative Study of Electoral Systems (CSES) at GESIS Mannheim. She holds a bachelor's degree in political science and business administration and a master's degree in political science, both from the University of Mannheim. Klara is responsible for processing CSES micro data, data documentation, collecting district data, data harmonization, data quality checks, website content, and responding to user queries. She is pursuing a Ph.D. in political science at the Graduate School of Social Sciences at the University of Mannheim. Her research interests are comparative politics focusing on voting behavior, especially political leaders, economic voting, and political attitudes.

## Selected Publications:

- Dentler, K., Gschwend, T., & Hünlich, D. (2021). A swing vote from the ethnic backstage: The role of German American isolationist tradition for Trump's 2016 victory. *Electoral Studies*, 71. <https://doi.org/10.1016/j.electstud.2021.102309>
- Dentler, K., Bluemke, M., & Gabriel, O.W. (2020). German Satisfaction with the Political System Short Scale (SPS). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. <https://doi.org/10.6102/zis278>
- Blinzler, K., Blumenberg, M. S., & Bucher, H. (2019). "Die Regierungsbildung." In S. Roßteutscher, R. Schmitt-Beck, H. Schoen, B. Weißels & C. Wolf (Eds.), *Zwischen Polarisierung und Beharrung: Die Bundestagswahl 2017* (Wahlen in Deutschland, Vol. 3, pp. 357-374). Nomos. <https://doi.org/10.5771/9783845287607-357>

## 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 will focus on their practical application using

electoral research data in Stata. In the first part of the course, we will cover the program's interface and introduce its syntax structure and basic rules for writing 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. Depending on the participants' prior knowledge, we might provide further insight into the automatization of data wrangling and analysis procedures and the export of publication-ready results. Additionally, 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:

Using Stata effectively, data management, descriptive statistics, data analysis, visualization

### Course Prerequisites:

- Familiarity with quantitative (survey) data
- Basic knowledge of uni- and bivariate statistics (i.e., 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 in Stata 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 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 and support work on assignments.

### 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. On the first day, we will focus on the interface of Stata and the syntax language, how to open your data in Stata and how to get a first grasp of your data using descriptive statistics. On the second day, we will take a first look at how to efficiently prepare your data for analysis by focusing on data management and manipulation, and how to analyse your data to answer basic research questions using bi- and multivariate statistics. On the third day, we will continue with bi- and multivariate statistics and, finally, focus on 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 Comparative Study of Electoral Systems (CSES), the different possible applications of Stata will be demonstrated using CSES 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 – 12.00	<b>Introduction to the Stata interface and syntax language</b>
	12.00 – 13.00	<b>Lunch Break</b>
	13.00 – 15.00	<b>Getting to know your data using descriptive statistics</b>
	<u>Suggested reading (optional):</u> <ul style="list-style-type: none"> <li>▪ Acock, Alan C. 2018. Getting Started. In <i>A Gentle Introduction to Stata</i>, 6th Ed. College Station, TX: Stata Press, pp. 1-19.</li> <li>▪ Acock, Alan C. 2018. Working with commands, do-files, and results. In <i>A Gentle Introduction to Stata</i>, 6th Ed. College Station, TX: Stata Press, pp. 77-92.</li> <li>▪ Kohler, Ulrich, and Frauke Kreuter. 2012. The first time. In <i>Data Analysis Using Stata</i>, 3<sup>rd</sup> Ed. College Station, TX: Stata Press, pp. 1-24.</li> </ul>	
2	10.00 – 12.00	<b>Data management and data manipulation</b>
	12.00 – 13.00	<b>Lunch Break</b>
	13.00 – 15.00	<b>Bi- and multivariate statistics I</b>

	<u>Suggested reading (optional):</u> <ul style="list-style-type: none"> <li>▪ Acock, Alan C. 2018. Preparing data for analysis. In <i>A Gentle Introduction to Stata</i>, 6th Ed. College Station, TX: Stata Press, pp. 51-67.</li> <li>▪ Kohler, Ulrich, and Frauke Kreuter. 2012. Creating and changing variables. In <i>Data Analysis Using Stata</i>, 3<sup>rd</sup> ed. College Station, TX: Stata Press, pp. 77-114.</li> </ul>	
<b>3</b>	<b>10.00 – 12.00</b>	<b>Bi- and multivariate statistics II</b>
	<b>12.00 – 13.00</b>	<b>Lunch Break</b>
	<b>13.00 – 15.00</b>	<b>Presentation and visualization of results</b>
	<u>Suggested reading (optional):</u> <ul style="list-style-type: none"> <li>▪ Kohler, Ulrich, and Frauke Kreuter. 2012. Introduction to linear regression. In <i>Data Analysis Using Stata</i>, 3<sup>rd</sup> Ed. College Station, TX: Stata Press, pp. 253-340.</li> <li>▪ Kohler, Ulrich, and Frauke Kreuter. 2012. Creating and changing graphs. In <i>Data Analysis Using Stata</i>, 3<sup>rd</sup> Ed. College Station, TX: Stata Press, pp. 115-156.</li> </ul>	

### 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 Literature:

- Baum, Christopher F. 2016. *An Introduction to Stata Programming*. 2<sup>nd</sup> Ed. College Station, TX: Stata Press.
- Cameron, A. Colin, and Pravin K. Trivedi. 2010. *Microeconometrics Using Stata*. College Station, TX: Stata Press.
- Cox, Nicholas, and Joseph Newton. 2014. *One Hundred Nineteen Stata Tips*. 3<sup>rd</sup> Ed. College Station, TX: Stata Press.
- Kohler, Ulrich, and Frauke Kreuter. 2012. *Data Analysis Using Stata*. 3<sup>rd</sup> Ed. College Station, TX: Stata Press.
- Mehmetoglu, Mehmet, and Tor Jakobsen. 2017. *Applied Statistics Using Stata. A Guide for the Social Sciences*. London: Sage.
- Mitchell, Michael N. 2020. *Data Management Using Stata: A Practical Handbook*. 2<sup>nd</sup> 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/](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/>