

Course type: Workshop

## Stata for Empirical Research and Data Science – Advanced

### Organizational details

Instructor: *Dr Tobias Keller*  
Date: January 18 and 19, 2018, 9 – 17h  
Location: HS 30, PC-Pool, Licher Strasse 68, 35394 Giessen  
ECTS: 3 ECTS  
Participants: 15

### Objectives

Learn how to use Stata efficiently in your research project. Improve your skills in data preparation, data management and data analysis. Learn how to automate repetitive analyses using loops and see how useful and easy it is to write your own little Stata programs. Finally, improve the structure and reproducibility of your project by applying best practices from both Data Science projects and software development projects.

### Content and methods

Participants will learn by examples and exercises from the instructor's experience in research and practice.

Exercises will make up about 40% of the course time. Participants will complete those using the computers of the "PC Pool". However, participants will be expected to prepare certain analyses before the course, either using their own Stata copy or using the "PC Pool", and to bring the results stored on a USB stick. Please see the preparation requirements below. The exercises will be based on exemplary datasets that will be provided to the participants before the course.

### Structure

1. Best Practices: how to structure your Data Science project
  - Digression: The Data Scientist's Role
  - Examples of Data Science projects from Practice and Research
  - CRISP: an idealized Data Science process
  - Using a software repository to keep track of different versions of your code
2. Data Analysis
  - Some important user written programs in Stata
  - Detecting and treating outliers and influential cases in your data
  - Shortcuts for interaction terms, indicator variables ("Dummies")
  - Dealing with multicollinearity

3. Data Preparation and Data Management
  - Time series operations
  - Using egen / egenmore for more complex calculations
  - Groupwise calculations
  - Appending / Merging / Joining datasets
  - Reshaping: long and wide format
  - Detecting and dealing with duplicates
4. Stata Programming
  - Loops
  - global and local macros
  - Your first own Stata program

**To gain the ECTS credit points participants have to:**

- Download and extract the course material and exemplary datasets that will be provided in time before the course.
- Using the “PC Pool” or your own Stata copy, complete the pre-course assignments distributed in time before the course. Bring the results on a USB stick.
- Familiarize yourself with the main exemplary datasets from the course material. They have originally been downloaded from the World Bank Open Data Platform (<http://data.worldbank.org/data-catalog/world-development-indicators>).
- In order to better understand the main research questions we will be dealing with, take a look at the World Development Indicators Report from 2017, downloadable here: <https://openknowledge.worldbank.org/bitstream/handle/10986/26447/WDI-2017-web.pdf>
- In order to achieve the ECTS credits, participants will have to complete an additional assignment after the course, due on January 31, 2018. The assignment will be a repetition of the most important topics and should take approximately 10 to 20 hours.

**Target group**

Doctoral candidates or postdoctoral researchers doing empirical research or intending to work as data scientists. It is required that the participants have basic Stata knowledge and have made their first experiences in applying Stata in their research projects. In particular, the contents of the introductory Stata course offered by the GGS are required knowledge. (For more information, see <http://www.uni-giessen.de/fbz/zentren/ggs/veranstaltungen/curriculum/archiv/ss17/ss17/syllabus-stata>)

**Course language**

English (German, if only German participants)

Please note: As this is not an English language course proficiency in English at the C1 level of competency is required.

**Registration**

By **January 8, 2018** via e-mail at [info@ggs.uni-giessen.de](mailto:info@ggs.uni-giessen.de).