

UCT TRAINING PROGRAMME

DATA WRANGLING WITH STATA: HOW TO MANAGE YOUR EMPIRICAL RESEARCH PROJECT

Presented by SALDRU at UCT

14-18 July 2025

This one-week course aims to address a crucial aspect of empirical research that is often overlooked: data management. Graduate programmes often focus on mastering sophisticated estimation models. However, without the ability to effectively manage and transform data into a well-structured and analysis-ready format, these advanced methods can be challenging to implement and may yield unreliable results. This hands-on course will provide participants with the essential skills and knowledge to manage and organise their data for empirical research projects (including those incorporating primary data collection) using Stata. Through a series of practical examples and exercises, participants will manage, clean, and process data of various types, use do files, import, export, merge and reshape datasets, and document their work for reproducibility and transparency.

By the end of the course, participants will be equipped with the necessary tools and techniques to handle data efficiently, allowing them to focus on the analysis and interpretation of results rather than getting lost in data management challenges. This course is recommended for graduate students and researchers conducting empirical social science. It should not only enhance your research skills but also provide a solid foundation for reproducible and transparent research practices.

Preliminary course outline

- 1) An introduction to data management using Stata
- 2) Principles of reproducible workflows
- 3) Organizing and documenting your data and code
- 4) Managing data with do files and global file paths
- 5) Merging, appending and reshaping data
- 6) Importing, exporting and saving data
- 7) Working with duplicates
- 8) Working with different types of data (e.g. strings and dates)
- 9) Using extensions to generate (egen) in Stata
- 10) Considerations for storing and sharing data

Course requirements

Course participants should have experience with using Stata and have used Stata recently. They should specifically be comfortable with doing the following in Stata:

- Using do files to run Stata code.
- Exploring data (using the following commands: *use*, *browse*, *describe*, *codebook*, *lookfor*, *count*, *numlabel*).
- Creating and labelling variables (using *generate*, *replace* and *recode*).

- Selecting observations and variables using logical operators (e.g., using `!=`, `>=`, `>`, `+`, etc.).
- Creating one-way and two-way frequency tables (using *tabulate*).

Should attendees be unfamiliar with any of the above course requirements, or wish to take a Stata refresher course, they are encouraged to also apply to attend the SALDRU *Social Science Research Using Survey Data* course which will run between 7-11 July 2025.

There is no cost to attend the course and a light lunch will be provided daily.

To apply for the course, click on the following link:

[DATA WRANGLING WITH STATA](#)
[COURSE APPLICATION](#)

Application deadline: 23 June 2025

Should you have any queries, please email course.saldru@uct.ac.za.

Note:

- It is assumed that course participants are **already somewhat familiar** with Stata.
- As this workshop tends to be heavily over-subscribed, we cannot guarantee a place for every applicant. Successful applicants will be notified by **25 June 2025**. Should you not have heard from us by that date, please consider your application unsuccessful.