"Basic programming for longitudinal data in Stata"

SIC organized and taught by Bram Hogendoorn (b.hogendoorn@uva.nl)

Objective

This short intensive course (SIC) aims to help students be more efficient in the preparation, estimation, and presentation of quantitative data. Its main focus is on longitudinal data, which is particularly useful for students who want to describe individual and social change or who try to estimate causal effects. Nonetheless, all that is learned can be applied to cross-sectional data and other types of nested data more generally. No prior programming experience is required and we will not be using Mata language, but students should be familiar with basic Stata commands and syntax writing. Special attention is paid to the importance of preparing clear syntax that is publicly accessible.

After completion of the course, the students should be able to:

- Understand and perform different ways of organizing data
- Understand and apply macros
- Write and debug loops
- Write and debug a small program
- Prepare and publish syntax

In preparation of the course, students are expected to read the prescribed literature. They are further expected to bring in a programming problems that they face in their current research, and if not applicable, to work out a concrete hypothetical problem that can be discussed during the course.

Description

To achieve the above-mentioned aims the PhD students will take part in three tutorials of two hours per tutorial. In preparation of the tutorials, participants will be asked to have read and practiced with the literature. In addition, participants will be asked to submit a half-page long summary of their research projects. The research summary should be supplemented by a half-page document about a particular programming problem that they face. This document should describe the objective, the data, the problem in achieving the objective, and what has been tried so far to solve the problem. Absent a particularly problem, participants are asked to work out a hypothetical problem that they might face in their research.

During the first tutorial, the instructor will discuss the submitted problems as a means of introducing the course. The remaining part of the tutorial will focus on how to prepare the data for programming: understanding the organization of the data, rearranging the organization, working with time order (or any other dimension), referring within a given syntax file. This tutorial will close with a preface to looping.

During the second tutorial, students will be invited to comment on each other's work. These comments will primarily address a reformulation of the problem, but if a simple programming solution can be thought of, then students can work this out in pairs. The remainder of the second tutorial forms the main part of the course. It will focus on loops: writing, debugging and executing a loop on the students' own data. Then it will continue with the programming basics: understanding program classes, using arguments, applying the syntax command to structure a program. This tutorial will close with students writing a short program to solve the problem that they submitted.

During the last tutorial, students will discuss the importance of open science with regard to their own research program. Following this discussion, the tutorial will focus on: exporting results, cleaning and organizing syntax, and using the Open Science Framework (OSF). Students are then asked to incorporate the exportation of results into the small program that they wrote during the previous tutorial. The tutorial will close with a brief group discussion in which we reflect on the course.

Schedule

REC B2.10	Tuesday 22nd October
15.00 – 15.15	Welcome and introduction
15.15 – 15.30	Discussion of submitted research problems and course outline
15.30 – 16.00	Establishing order: *by, sort, reshape, indexing*
16.00 – 16.10	Coffee break
16.10 – 16.30	Working with time: *tsfill, egen seq*
16.30 – 17.00	Cross referring: *local, global, foreach, forvalues*

REC B2.07	Tuesday 29th October
15.00 – 15.15	Welcome and summary of last week
15.15 – 15.35	More on looping: *set trace, capture, confirm, if-else, continue-exit, quietly-noisily*
15.35 – 15.50	Reformulation of research problems in pairs
15.50 – 16.00	Coffee break
16.00 – 16.35	Programming basics: *program, syntax, viewsource*
16.35 – 17.00	Writing own program

REC B2.06	Tuesday 5th November
15.00 – 15.15	Welcome and summary of last week
15.15 – 15.40	Discussion about open science and demonstration of OSF
15.40 – 16.00	Preparation of syntax for publication
16.00 – 16.10	Coffee break
16.10 – 16.25	Exporting results: *save, putexcel, matrix, asdoc, estout*
16.25 – 16.50	Integration and completion of program
16.50 – 17.00	Reflection and conclusion

Literature list

Tutorial	Resource	Sections
1	Kohler, U., & Kreuter, F. (2005). *Data analysis using Stata* (3rd ed.). College Station, United States: Stata Press. Available on https://www.dawsonera.com/abstract/9786000039189.	Chapters 12.2 and 12.3 (except 12.3.8 and 12.3.10).
	StataCorp (2017). Stata user's guide, release 15. Statistical Software. College Station, United States: StataCorp LLC. Available on https://www.stata.com/manuals/u.pdf.	Chapter 11 (except 11.4.3, 11.6, 11.7) Chapter 13.4 - 13.10 Chapter 18 (except 18.3.6 - 18.3.13, 18.4.2, 18.4.3, 18.11 - 18.14).
2	StataCorp (2017). Stata programming reference manual, release 15. Statistical Software. College Station, United States: StataCorp LLC. Available on https://www.stata.com/manuals/p.pdf	capture (pp. 13-16) confirm (pp. 66 - 70) continue (pp. 71 - 72) exit (pp. 230 - 231) foreach and forvalues (pp. 238 - 250) gettoken (pp. 255 - 256) if and else (pp. 259 - 261) matrix (pp. 339 - 341) quietly and noisily (pp. 443 - 446) syntax (pp. 540 – 549, 551 - 554) viewsource (p. 591)
3	Anderson, R. G., Greene, W. H., McCullough, B. D., & Vinod, H. D. (2008). The role of data/code archives in the future of economic research. *Journal of Economic Methodology*, *15*(1), 99-119.	Sections 1, 4, 5.
	Andreoli-Versbach, P., & Mueller-Langer, F. (2014). Open access to data: An ideal professed but not practised. *Research Policy*, *43*(9), 1621-1633.	Everything except sections 4.2.4 and 5.

List of participants

	Name	Discipline	Programme group
1	Laura Mulder	Political Science	Challenges to Democratic Representation
2	Sander Kunst	Political Science	Challenges to Democratic Representation
3	Twan Huijsmans	Political Science	Challenges to Democratic Representation
4	Andrea Forster	Sociology	Institutions, Inequalities and Life Courses
5	Christoph Janietz	Sociology	Institutions, Inequalities and Life Courses
6	Dieuwke Zwier	Sociology	Institutions, Inequalities and Life Courses
7	Gabriel Otero Cabrol	Sociology	Institutions, Inequalities and Life Courses
8	Katharina Stückradt	Sociology	Institutions, Inequalities and Life Courses
9	Katja Chkalova	Sociology	Institutions, Inequalities and Life Courses
10	Maria Eismann	Sociology	Institutions, Inequalities and Life Courses
11	Qiong He	Geography	Urban Geography
12	Rik Damhuis	Geography	Urban Geography
13	Lin Rouvroye	Medicine	Social Medicine