IV Residential SUMMER SCHOOL

ANALYSING MICRO DATA IN STATA

Florence, 28 August -3 September 2022

TStat's Analysing Micro Data in Stata Summer School offers participants a comprehensive introduction to the principle methodologies used in the analysis of micro data. Micro data contains information at the level of a specific unit (such as individuals, firms or entities), by its very nature micro data has become an increasingly important source of information offering researchers and policy makers an effective tool with which to obtain a more in-depth understanding of an array of political, socio-economic and Public Health phenomena. The collection and subsequent analysis of micro data has over recent years therefore proved to be the key to policy formulation, the targeting of interventions and the subsequent monitoring and measurement of the impact of such interventions and policies.

Although micro data analysis techniques were originally developed and applied in the field of economics, the increasing availability of micro data has resulted in a steady increase in the analysis of such data by researchers working in Political and Social Sciences, Biostatistics, Education, Epidemiology and Public Health.

Throughout the course of the Summer School, the course leaders will focus, from both a theoretical and applied point of view, on the principal methodologies implemented for the analysis of both cross-section and panel data: linear models, count models, binary dependent variable models, multinomial models, Tobit and Interval Regression models, models with Sample Selection, and estimation of Extended Regression Models (ERM), which implement Maximum Likelihood estimators capable of simultanously treating issues of Sample Selection and the presence of both endogenous regressors and treatment variables.

The school opens with an optional introductory one day course (Module A) to the statistical package Stata, during which participants will be provided with the necessary tools to enable them to use Stata independently and actively participate in the applied empirical Lab sessions during the course of the week.

In common with TStat's training philosophy, the summer school is composed of both a theoretical component (in which the techniques and underlying principles behind them are explained), and an applied (hands-on) segment, during which participants have the opportunity to implement the techniques using real data under the watchful eye of the course tutor. Theoretical sessions are reinforced by case study examples, in which the course tutor discusses and highlights potential pitfalls and the advantages of individual techniques. The intuition behind the choice and implementation of a specific technique is of the utmost importance.

SUMMER SCHOOL CODE

I-SS13

DATE AND LOCATION

Florence,
28 August - 3 September 2022

CISL Studium Center
Via Della Piazzola, 71
I-50123 Florence
http://www.centrostudi.cisl.it

In this manner, the course leader is able to bridge the “often difficult” gap between abstract theoretical methodologies, and the practical issues one encounters when dealing with real data. Special attention is also given to the interpretation and presentation of results.

At the end of the course, participants are expected to be able, with the aid of the Stata routines implemented during the sessions, to independently implement the methodologies and techniques acquired during the course by adopting the Stata routines to their own particular research needs.

PREREQUISITES

It is assumed that course participants have at some point followed a university basic course in econometrics or statistics and thus be comfortable with the arguments covered in chapters 1-9 in J.M. Wooldridge, Introductory Econometrics: A Modern Approach, South-Western College Pub, 2013, 5th edition.

Previous exposure to Stata would also be an advantage. Participants with no previous knowledge of Stata are however, strongly encouraged to follow the Introduction to Stata Course offered at the beginning of the School.

TARGET AUDIENCE

The Summer School program has been particularly developed for both doctoral students and young researchers working in biostatistics, business, economics, epidemiology, finance, public health, psychology, social and political sciences needing to acquire the necessary toolset to independently conduct empirical analysis using micro data, but who may not have access to a specific micro data analysis course in their home institution. It is however, also particularly useful to professionals working in one of these fields needing to either refresh their existing micro data skills or acquire new ones.

MODULE A | DAY 1 - STATA IN JUST ONE DAY!

SESSION I: INTRODUCTION
GETTING STARTED

1. Stata’s GUI
2. File types in Stata
3. Working interactively in Stata
4. Saving output: the log file
5. Interrupting Stata
6. Loading Stata databases
7. The Log Output File
8. Saving databases in Stata
9. Exiting the software

SESSION II: PRELIMINARY DATA ANALYSIS

1. A preliminary look at the data: *describe, summarize* commands
2. Abbreviations in Stata
3. Stata’s syntax
4. Summary statistics
5. Statistical Tables: *table, tabstat* and *tabulate* commands

ANALYSING MICRO DATA IN STATA

SESSION III: DATA MANAGEMENT

1. Renaming variables
2. Selecting or eliminating variables
3. The *count* command
4. *sort* command
5. Creating sub-groups: the prefix *by*
6. Creating new variables: generate
7. Operators in Stata
8. The command *assert*
9. Missing values in Stata
10. Modifying variables: *replace, recode*
11. Creating Labels: variable labels and value labels
12. Creating dummy variables

SESSION IV: IMPORTING DATA FROM SPREADSHEETS

1. *Import Excel* and *Export Excel* commands
2. The *insheet* and *outsheet* commands
3. Reading in Text Data Files
4. Issues to watch out for when importing data
   - Missing values
   - String variables
   - Date variables
5. Redefining missing values
6. *destring* command
7. *tostring* command
8. dealing wih "messy" strings

SESSION V: GRAPHICS - A BRIEF INTRODUCTION

1. Stata's syntax for two way graphs
2. Saving and exporting graphs
3. Useful *graph* commands
4. Personalizing a graph
5. Stata's Graph Editor

APPENDIX A

1. Merging data bases
2. *do* files

APPENDIX B: MORE ADVANCES ISSUE *(time permitting)*

1. *do files*
2. Merging data bases
3. *e-class* and *r-class* variables
4. *collapse* command
5. *preserve* command
6. *restore* command

MODULE B | DAY 2 - LINEAR REGRESSION MODELS

SESSION I: THE LINEAR MODEL WITH EXOGENOUS REGRESSORS

1. Identification
2. The Ordinary Least Squares (OLS) Estimator: regress
3. Specification tests and tests for robust inference: *estat imtest, estat hettest, estat bgodfrey, actest*

SESSION II: THE LINEAR MODEL WITH ENDOGENOUS REGRESSORS

1. Identification
2. IV e GMM Estimators: *ivregress, gmm*
3. Specification tests and tests for robust inference: *ivhettest, actest, estat overid, estat endogenous, estat firststage, weakivtest*

https://www.tstattraining.eu/training/analysing-micro-data_ss/

DAYS 3-4 - LINEAR PANEL DATA REGRESSION MODELS

SESSION I: PANEL DATA IN STATA SOME BASIC CONCEPTS

1. Panel Data structures in Stata
2. Time Series Operators in Stata
3. The advantages of Panel Data for applied micro data analysis

SESSION II: LINEAR PANEL DATA MODELS WITH EXOGENOUS VARIABLES

1. One-way and two-way fixed effect estimators: *xtreg, fe*
2. Random Effects Estimators: *xtreg, re; xtmixed*

SESSION III: LINEAR PANEL DATA MODELS WITH EXOGENOUS VARIABLES: ROBUST INFERENCE

1. Robust covariance estimators
2. The first-difference estimator
3. Testing for non *i.i.d.* errors
4. Testing Random Effects against Fixed Effects:
   - non-robust approach using Hausman
   - robust approach using Mondlak auxiliary regression (Wooldridge, 2010)

SESSION IV: LINEAR PANEL DATA MODELS WITH ENDOGENOUS VARIABLES

1. Fixed and Random Effect IV Estimators: *xtivreg*
2. Hausman and Taylor's estimator: *xthtaylor*

DAYS 5-7 NON-LINEAR REGRESSION MODELS

SESSION I: COUNT MODEL ESTIMATORS

1. The Poisson Model: *poisson, nl, gmm*
2. The Poisson Model with engodenous regressors: *ivpoisson, gmm*
3. Estimation and tests in the presence of *overdispersion* (the negative *binomial regression mode*l): *nbreg*
4. Estimation and interpretation of the marginal estimation effects using Stata's post estimation command *margins*
5. Fixed and Random Panel Data Estimators: *xtpoisson, xtnbreg*

SESSION II: DISCRETE DEPENDENT VARIABLE MODELS

1. Estimating linear models with binary dependent variables – Logit, Probit and the Linear Probability Model: *probit, logit, regress*
2. The Heteroskedastic Probit Model and tests of heteroskadicity: *hetprobit*
3. Measures of Goodness of Fit and Specification Tests: *estat classification, estat gof*
4. Estimating and interpreting marginal effects: *margins*
5. Fixed and Random Panel Data Estimators: *xtprobit, xtlogit, clogit*

SESSION III: PROBIT MODELS WITH ENDOGENOUS REGRESSORS

1. Maximum likelihood estimation in the presence of continuous endogenous regressors: *ivprobit*
2. Measures of Goodness of Fit: *tabulate, estat classification, estat correlation*
3. Estimating and interpretation of estimated marginal effects: *margins*

SESSION IV: MULTINOMIAL MODELS

1. Ordered categorical variable models (the Ordered Probit and Ordered Logit Estimators): *oprobit* and *ologit*
2. The Heteroskedastic Probit Model and tests of heteroskadicity: *hetoprobit*
3. Random Effect Ordered Panel Data Probit Models: *xtoprobit*
4. Models with unordered categorical variables - Multinomial Logit and Multinomial Probit estimators: *mlogit, mprobit*
5. MacFadden's Choice Model - categorical variable models with alternative specific regressors: *cmclogit, cmcprobit*

https://www.tstattraining.eu/training/analysing-micro-data-stata_ss/

6. Measures of Goodness of Fit and Specification Tests
7. Estimation and interpretation of marginal effects using the Stata post estimation command *margins*

SESSION V: THE TOBIT MODEL, INTERVAL REGRESSION AND SAMPLE SELECTION

1. The Tobit Model: *tobit*
2. Estimating the Tobit model with endogenous regressors: *ivtobit*
3. Interval Regression: a generalization of the Tobit Model: *intreg*
4. Fixed and Random Effects Panel Data Estimators: *xttobit, xtintreg*
5. Estimators for Sample Selection Models: *heckman*
6. Estimation and interpretation of marginal effects using the Stata post estimation command *margins*
7. Random Effect Panel Data Estimators: *xtheckman*

SESSION VI: EXTENDED REGRESSION MODELS WITH BOTH ENDOGENOUS REGRESSORS AND TREATMENT EFFECTS IN THE PRESENCE OF SAMPLE SELECTION

1. Extended Regression Models: eregress
2. Extended Regression Probit Models: eprobit
3. Ordered Extended Regression Probit Models: eoprobit
4. Extended Interval Regression Models: eintreg
5. Extended Regression Random Effect Panel Data models: xteregress, xteprobit, xteoprobit, xteintreg

COURSE REFERENCES

- A Gentle Introduction to Stata, 6th Ed., Alan Acock (2018) Stata Press
- Data Analysis Using Stata, 3rd Ed., Ulrich Kohler, Frauke Kreuter (2012) Stata Press
- Data Management Using Stata: A Practical Handbook, 2nd Ed., Michael N. Mitchell, (2020) Stata Press
- The Workflow of Data Analysis Using Stata, J. Scott Long (2009) Stata Press
- Mostly Harmless Econometrics: An Empiricist's Companion, Joshua D. Angrist e Jorn-Steffen Pischke (2008) Princeton University Press
- Microeconometrics Using Stata, Colin Cameron and Pravin K. Trivedi (2010) Stata Press

REGISTRATION FEES

ENTIRE WEEK (MODULES A plus B, 7 days)
Full-Time Students*: € 1890.00
Ph.D. Students: € 2770.00
Academic: € 3080.00
Commercial: € 4550.00

MODULE B (6 days)
Full-Time Students*: € 1620.00
Ph.D. Students: € 2380.00
Academic: € 2640.00
Commercial: € 3900.00

**To be eligible for full-time student prices, participants must provide proof of their full-time student status for the current academic year. Our standard policy is to provide all full-time students, be they Undergraduates or Masters, access to our student registration rates. Part-time master and doctoral students on the other hand, who are also currently employed will however, be assigned the standard academic registration fee. Residential costs for full-time students are completely covered TStat Training through our Investing in Young Researchers Programme. Participation is however restricted to a maximum of 3 students.

Fees are subject to VAT (applied at the current Italian rate of 22%). Under current EU fiscal regulations, VAT will not however applied to companies, Institutions or Universities providing a valid tax registration number.

Please note that a *non-refundable deposit* of €100.00 for full-time students and €250.00 for Academic and Commercial participants, is required to secure a place and is payable upon registration. The number of participants is limited to 15. Places will be allocated on a first come, first serve basis.

Course fees cover: i) teaching materials (copies of lecture slides, databases and Stata routines used during the summer school; ii) a temporary licence of Stata valid for 30 days from the day before the beginning of the school; iii) half board accommodation (breakfast, lunch and coffee breaks) in a single room at the CISL Studium Centre or equivalent (7 nights for entire school, 6 nights for Modules B). Participants requiring accommodation the night of the final day of the school, are requested to contact us as soon as possible.

To maximize the usefulness of this summer school, we strongly recommend that participants bring their own laptops with them, to enable them to actively participate in the empirical sessions.

Individuals interested in attending this summer school must return their completed registration forms by email (training@tstat.eu) to TStat by the 18th of August 2022.

COURSE LEADERS

Dr. Una-Louise BELL
TStat Training | TStat S.r.l.

Dr. Giovanni BRUNO
Bocconi University, Milan

CONTACTS

Monica Gianni
TStat Training | Kleebergstraße, 8
D-60322 Frankfurt am Main

TStat S.r.l. | Via Rettangolo, 12-14
I-67039 Sulmona (AQ)
T. +39 0864 210101

training@tstat.eu
www.tstattraining.eu
www.tstat.eu