

Introduction to Stata provides participants with the necessary introductory toolset to enable them to carry out efficient data analysis and data management in Stata in a reproducible manner. The course covers everything from the Stata “getting started” basics, to an overview of the available Stata commands for preliminary data analysis, data management, importing and exporting data from/to other data formats and the creation of graphs in Stata.

In common with TStat’s course philosophy, each individual session 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. Throughout the course, theoretical sessions are reinforced by case study examples, in which the course tutor discusses and highlights potential pitfalls and the advantages of individual techniques.

**WHO SHOULD ATTEND?** This course is invaluable to both professionals and researchers (working in Biostatistics, Economics, Epidemiology, Social and Political Sciences and Statistics), wishing to use Stata for applied statistical analysis, data management or data visualization.

**COURSE REQUISITES** Familiarity with PC and a working knowledge of English.

## PROGRAM

### SESSION I: INTRODUCTION – GETTING STARTED

- *Stata’s* GUI
- *Stata* File formats
- Interacting with your Operating System from within *Stata*
- Working interactively in *Stata*
  - Organizing one’s work in *Stata*
  - *Help*
  - Web resources in *Stata* – downloading updates and new commands via internet
- Interrupting *Stata* Sessions
- Loading *Stata* databases
- Saving databases in *Stata*
- Saving output: the log file
- Use of Comments in *Stata*
- *Exiting the software*

### SESSION II: PRELIMINARY DATA ANALYSIS

- A preliminary look at the data
- Variable types
- *Stata’s* Syntax
- Using abbreviations in *Stata*
- Constrained commands
- Creating Labels: variables, data and value labels

- Prefix *by*
- Summary statistics: *summarize*
- Statistical Tables: *table*, *tabstat*, *tabulate* and *tabi* commands
- Basic Data Management Commands: -
  - *drop - keep*
  - *rename*
  - *count*
  - *sort*
- Creating transforming data and variables
  - *generate*, *egen*
  - *recode*, *replace*
- Working with Missing Values
- Command *assert*

### SESSION III: CREATING SIMPLE STATA ROUTINE: DO FILES

### SESSION IV: DATA MANAGEMENT

- Renaming variables
- *Selecting or eliminating variables*
- The *count* command
- *sort* command
- Creating sub-groups: the prefix *by*
- Creating new variables: *generate* and *egen*
- Operators in *Stata*
- The command *assert*

- Missing values in *Stata*
- Modifying variables: *replace*, *recode*
- Creating Labels: variable labels and value labels
- Creating dummy variables

## SESSION V: IMPORTING DATA FROM SPREADSHEETS

- *import excel* and *export excel* commands
- The *insheet* and *outsheet* commands
- *infile* command – importing .txt files in *Stata*
- Importing SPSS files
- Memory Management – A Note
- Issues to watch out for when importing data
  - Missing values – redefining them
  - String variables – *encode*, *destring* – *tostring* commands
  - Working with dates in *Stata*
- *assert* command
- Exporting data to Spreadsheets
  - Issues to watch out for

## USEFUL REFERENCES

- A Gentle Introduction to *Stata*, 5<sup>th</sup> Ed. StataPress 2016, (Alan Acock)
- Data Analysis Using *Stata*, 3<sup>rd</sup> Ed., StataPress 2012, (Ulrich Kohler, Frauke Kreuter)
- Data Management Using *Stata*: A Practical Handbook, StataPress 2010, (Michael N. Mitchell)

## LOGISTICS

The course will be held in Frankfurt am Main on the 2<sup>nd</sup> and 3<sup>rd</sup> of June 2016, from 9:00 am. to 5:00 pm.

## REGISTRATION FEES

Academic € 900,00 • Commercial € 1200,00 • Students € 600,00  
All fees are subject to VAT (applied at the current Italian rate of 22%).

The number of participation is limited to 10. Places, will be allocated on a first come, first serve basis.

The course fee covers: course materials and a temporary licence of *Stata* valid for 30 days from the beginning of the course; light lunch and coffee break. The course materials comprise the theoretical handouts, along with the *Stata* do-files, ado-files and data-sets required to replicate all the empirical demonstrations implemented in the course sessions under the guidance of the course leader.

### Please note:

- A 15% discount will be applied to all additional participants from the same company or institution enrolling on the same course.
- TStat's Training Packages: individuals attending more than one of our training courses during the course of 2016 are entitled to a 15% discount off subsequent course fees.

Individuals interested in attending the training course, must return their completed registration forms either by email ([training@tstat.eu](mailto:training@tstat.eu)) or by fax (+39 0864 206014) to TStat by the 20<sup>th</sup> of May 2016.

## SESSION VI: MERGING DATA BASES

- *append* command
- *merge* command
  - merging data per observation
  - merging data bases using a unique (or multiple) identifiers
- *merge* options
- Merge reports
- System Variables (*\_n* and *\_N*)
- *updates* command
- Using the commands *append* and *merge*
- Concepts of wide or long format

## SESSION VII: GRAPHICS – AN INTRODUCTION (Time Permitting)

- *Stata's* graphical syntax
- Saving and exporting graphs
- Useful graph commands
- Personalizing a graph
- *Stata's* Graph Editor
- Creating multiple graphs – merging graphs

# ITALIAN STATA DISTRIBUTOR