

Data Management, Analysis and Visualization using Stata - 5 days at \$ 625 / delegate

Start date¹: 29th October, 2018

End date: 3rd November, 2018

Venue: Nairobi, Kenya

Get 16% when you register as a group (2 or more delegates).

Target clients / organizations:

Data managers, research officers, project managers, statisticians, data analysts, students (undergraduate and postgraduate), trainers/facilitators, research organizations, non-governmental organizations, policy makers, e.t.c.

¹ *This training happens every month starting on the first Monday and runs for one weeks. The same training can also be offered in SPSS, R or Python on request. training@drmathematics.com*

Contents

Introduction	1
1 Data management & graphics in Stata	1
1.1 Introduction to Stata	1
1.2 Creating and changing variables	2
1.3 Transforming variables and data sets	2
1.4 Introduction to Stata graphics	2
2 Statistics	2
2.1 Introduction to statistical concepts	2
2.2 Hypothesis testing	2
2.3 Tests of differences in population means	2
2.4 Analysis of contingency tables	3
2.5 Non-parametric methods	3
2.6 Linear regression and correlation	3
Miscellaneous	3

Introduction

Participants will learn the principles of statistics and gain skills in using statistical tools to describe, study and investigate the various variables in survey data sets using Stata. The statistical background required to conduct research, describe, summarize, develop hypothesis, assess associations, analyze data, interpret and communicate results will be studied comprehensively.

The course targets professionals in research (or research related) organizations / institutions who wish to acquire or increase their computational skills in the Stata software. The course is developed to benefit data managers, research officers,

project managers, statisticians, data analysts, students (undergraduate and postgraduate), trainers/facilitators, research organizations, non-governmental organizations, policy makers, among others.



1. Data management & graphics in Stata

This module will introduce participants to the Stata software version 15. All procedure that are needed to create data sets, transform variables, clean data, e.t.c will be studied using real survey data sets, The module will form the basis for all the computations that will be applied in various topics to be discussed in the workshop. Areas to be discussed include:

1.1 Introduction to Stata

Areas to be covered include the following:

1. Starting Stata
2. Setting layout
3. Directory management commands
4. Data types in Stata
5. Using Stata as a calculator
6. Stata command and options
7. Stata do-files
8. Creating data sets directly in Stata

9. Rename of variables
10. Managing variables and/or variable properties
11. Importing data from other software
12. Exporting data to other software
13. Loading data into the memory
14. The **in** and **if** qualifiers
15. The **by** prefix
16. Create subsets (keep and drop)
17. Create random variables (from distributions)
18. Random sampling
19. Sort variables
20. Change order of variables
21. Count number of observations
22. Generate sequential numbers
23. Working with dates
24. Viewing data sets
25. Interrupting computations
26. Help

1.2 Creating and changing variables

Areas to be covered include the following:

1. Create new variables
2. Extended generate command
3. Duplicate an existing variable
4. Replace contents of a variable
5. Convert numeric to string
6. Convert string numbers to numeric
7. Convert numeric values to missing and vice versa
8. Recode string variables
9. Decode numerically coded variables
10. Transforming a continuous variable to categorical
11. Reduce number of categories of a categorical variable
12. Managing duplicates

1.3 Transforming variables and data sets

Areas to be covered include the following:

1. Split variables
2. Extract parts of variables
3. Standardize variables
4. Create dummy variables
5. Create separate variables
6. Transpose variables
7. Stack variables
8. Unstack variables
9. Appending data sets
10. Combining data sets by a common variable
11. Convert datasets from wide to long
12. Convert datasets from long to wide
13. Some application to data cleaning

1.4 Introduction to Stata graphics

Areas to be covered include the following:

1. The graphics dialog windows

2. Graph elements (x and y labels, titles, legends)
3. Graph appearance (marker symbol, color, size, line width, pattern, e.t.c)
4. Multiple graphs (by option)
5. Graphics syntax
6. Adding text and annotations to graphs
7. Saving and printing graphs
8. Combining active graphs into one figure
9. Graphics window (interactive plotting)
10. Common graphs and charts

2. Statistics

There is currently a tremendous increase in the need for knowledge and information from available data. The intention of this section is to **a)** provide sufficient background information and detail in statistical techniques and **b)** tackle both descriptive and inferential techniques in Stata which will enable participants carry out analysis of real survey data sets on their own.

All the subsection provide a good starting point, both for statisticians who want to begin work on problems in the various application areas and data analysts who want to increase their working/computational knowledge of the Stata software.

2.1 Introduction to statistical concepts

Areas to be covered include the following:

1. Review of research process
2. Research designs
3. Sampling techniques
4. Types of data
5. Descriptive statistics
6. Graphs for descriptive statistics

2.2 Hypothesis testing

Areas to be covered include the following:

1. Definitions
2. Statistical inference
3. Generalizability
4. Confidence intervals in clinical research
5. P-values in clinical research
6. Hypothesis testing
7. Interpreting hypothesis test results

2.3 Tests of differences in population means

Areas to be covered include the following:

1. One sample t tests
2. Two sample independent t tests
3. Two sample paired t test
4. One way analysis of variance
5. Two way analysis of variance

2.4 Analysis of contingency tables

Areas to be covered include the following:

1. Introduction
2. Two by two tables: Proportion test
3. Two by two tables: Fisher's exact test
4. McNemar matched pairs for binary response
5. Other measures of association

2.5 Non-parametric methods

Areas to be covered include the following:

1. Sign test
2. Wilcoxon signed-rank test
3. Median test
4. Wilcoxon signed-sum (Mann-Whitney) test
5. Kolmogorov-Smirnov goodness-of-fit test
6. Kruskal-Wallis one way analysis of variance
7. Friedman two-way analysis of variance
8. Spearman rank correlation
9. Nonparametric regression analysis

2.6 Linear regression and correlation

Areas to be covered include the following:

1. Overview
2. Pearson correlation analysis
3. Simple linear regression
4. Multiple linear regression
5. Interpret results from linear regression
6. Regression diagnostics

Miscellaneous

The following sections give some information that may be helpful before, during or after the workshop. Please take a moment to read through.

Charges and venue

Charges: \$ 625 + 16% VAT

Venue: Nairobi, Kenya.

You can count on us that the training will not be canceled even if you were the only delegate except if you yourself decided so.

NOTE : The charges cater for snacks, food, drinks, writing materials and certificates but **DOES NOT** include any other expenses e.g. accommodation, travel, e.t.c.

Refund policy

If you have made payment for the training, you can cancel it at any time before the training starts as follows.

- **More than 7 days before start date:** 100% refund
- **Less than 7 days before start date:** 90% refund

- **After start date:** No refund

The official communication must be made by the delegate or sponsoring organization in writing (send in PDF format) to training@drmathematics.com. The refund process will be initiated the following **working day** of receipt of such a communication.

Onsite training

We can customize this content and have it offered at your own premises at your own pace.

Webinar

Are you willing to be trained on this course but are unable to travel (or want to save on travel expenses)? Sign up for a webinar and let us offer the same content online at \$ 150 + VAT.

Recommended literature

Participants could find the following books helpful as they prepare for the workshop.

1. Joseph, H. N. (2003) *Seventy-six Stata Tips, 2nd edition*. Stata Press.
2. Hamilton, L. C. (2012). *Statistics with Stata: version 12*. Cengage Learning.
3. Michael, N. M. (2004) *A Visual Guide to Stata Graphics*. Stata Press.

Other available training

The following training are also available. For any inquiries, please contact us at: training@drmathematics.com

1. Biostatistics and Epidemiology for Public Health Research using Stata (10 days at \$ 1,250)
2. Data Management and Analysis using SPSS (5 days at \$ 625)
3. Machine Learning Algorithms using Python (10 days at \$ 1,250)

All the above charges attract 16% VAT