

# INTRODUCTION TO STATA FOR GRADUATE STUDENTS

Kenn Chua

Summer, 2017

---

E-mail: [chuax025@umn.edu](mailto:chuax025@umn.edu)

Web: [http://github.com/kennchua/Workshop\\_Stata](http://github.com/kennchua/Workshop_Stata)

Office Hours: Mon - Fri 4:00 - 5:00pm

Class Hours: Mon - Fri 1:00-4:00 pm

Office: Ruttan Hall 249

Classroom: Coffey Hall 50

---

## Course Description

This course aims to introduce the statistical computing software Stata to incoming MS and PhD students. It aims to familiarize students with the various uses of Stata as a tool for data management and econometric analysis. The course will be taught at the beginner to advanced beginner levels and shall cover basic commands as well as impart efficient coding practices that will be essential for empirical research.

## Course Objectives

At the end of the course, students should be able to use Stata to:

1. Create, open, modify, and save data files
2. Generate descriptive statistics and perform basic econometric analyses
3. Produce tables and figures that can be easily incorporated in reports
4. Document all steps in data processing and analysis stages for later reproducibility
5. Write concise and efficient codes to run several commands and operations

## Instructional Format

The daily sessions will be divided into two parts. The first consists of an interactive lecture in which the instructor presenting the day's concepts and demonstrating their applications in Stata. Students are expected to follow the lecture by replicating the output on their computers. After the lecture, students will be asked to complete a series of exercises that will test their mastery of

the lesson. Copies of the lectures and answers to the exercises will be uploaded before the class begins. Students are highly encouraged to work together.

## **Class Requirements**

Apart from daily attendance and completion of in-class exercises, no assignments or quizzes will be required. Practice outside of class is highly encouraged but not mandatory. The class is non-credit and no grades will be given as part of the course.

## **Course Structure**

### **Day 1: Introduction**

1. Parts of the Stata window
2. Getting help
3. Do files and introduction to writing code
4. Loading and importing data (ASCII, .xls, .dta)
5. Data storage types
6. Arithmetic, logical, and relational operators
7. Examining the data using `describe`, `list`, `assert`, and `browse`
8. Summarizing data and generating frequency tables
9. Subsetting with `if` and `in`
10. Keeping or dropping observations and variables
11. Renaming variables and replacing values
12. Generating new variables with `gen`
13. Variable labels and value labels
14. Saving and exporting data

### **Day 2: Data Wrangling**

1. Handling missing values and outliers
2. Arranging columns and rows using `order`, `sort`, and `gsort`
3. Generating new variables with `egen`
4. Grouped data manipulation and summaries using the `by` and `bysort` prefix
5. Using `collapse`
6. Appending datasets
7. Merging datasets
8. Reshaping data from wide to long and long to wide

**Day 3: Data Visualization and Analysis**

1. Creating log files
2. Installing user-written commands
3. Creating Stata graphs
4. Exporting Stata graphs
5. Regression analysis in Stata
6. Working with indicator variables and categorical variables
7. Post-estimation commands
8. Exporting regressions results using `outreg2`
9. Setting up panel data for analysis

**Day 4: Scalars, Matrices, and Macros**

1. Scalars in Stata
2. Matrices and matrix operations in Stata
3. System variables
4. `r-class` and `e-class` objects
5. Macros: Globals vs. Locals
6. Temporary files

**Day 5: Loops and Automation**

1. `while` loops
2. `forvalues` and `foreach` loops
3. Conditional expressions
4. Nested loops

**Additional Material**

1. Writing your own program in Stata
2. Setting up a project directory and workflow
3. Exporting results to  $\text{\LaTeX}$
4. Exporting results to MS Excel
5. Other topics as suggested by the class