

Introduction to Stata

Alexis Grigorieff



What is Stata?

Stata is a **powerful statistical software**, which can be used for:

- **data manipulation**
- **data analysis**
- **creating graphics and tables**

How to access Stata: in the computer lab

Use the following details:

- Username: **client-**** where ** is the number of the monitor
- Password: **lab2015**

How to access Stata: from your computer

You will be able to download Stata via a web platform.

You will receive an email when the platform is ready.

The screenshot shows the Stata MP 13.0 application window. The title bar reads "Stata/Multi-Platform 13.0". The menu bar includes Open, Save, Print, Log, Viewer, Graph, Do-file Editor, Data Editor, Data Browser, More, Break, Search Help, Results, Variables, Properties, and Command. The main area displays the following text:

Statistics/Data Analysis

MP - Parallel Edition

**13.0 Copyright 1985–2013 StataCorp LP
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)**

3-user 8-core Stata network perpetual license:
Serial number:
Licensed to:

Notes:

1. (-set maxvar-) 5000 maximum variables

The right-hand sidebar contains three panels: Variables (with a search filter), Properties (showing details for selected variables), and Data (showing summary statistics like Observations = 0).

The Stata screen

Stata/MP 13.0

Open Save Print Log Viewer Graph Do-file Editor Data Editor Data Browser More Break Search Help

Results

```
___/___/___/___/___/ 13.0 Copyright 1985-2013 StataCorp LP
___/___/___/___/___/ Statistics/Data Analysis StataCorp
                        4905 Lakeway Drive
                        College Station, Texas 77845 USA
                        800-STATA-PC http://www.stata.com
                        979-696-4600 stata@stata.com
                        979-696-4601 (fax)

MP - Parallel Edition

3-user 8-core Stata network perpetual license:
  Serial number:
  Licensed to:

Notes:
  1. (-set maxvar-) 5000 maximum variables
```

Command

where you can type commands

Variables

Enter filter text here

Name	Label
------	-------

Properties

Variables

Name	
Label	
Type	
Format	
Value Label	
Notes	

Data

Filename	
Label	
Notes	
Variables	0
Observations	0
Size	0
Memory	64M
Sorted by	

crucial

The Stata screen

The screenshot displays the Stata/MP 13.0 software interface. The top menu bar includes icons for Open, Save, Print, Log, Viewer, Graph, Do-file Editor, Data Editor, Data Browser, More, and Break, along with a search bar. The main Results window (highlighted in pink) shows the Stata startup screen with the following text:

```
_____ / _____ / _____ / _____ / _____ /
Statistics/Data Analysis      13.0  Copyright 1985-2013 StataCorp LP
                                StataCorp
                                4905 Lakeway Drive
                                College Station, Texas 77845 USA
                                000...STATA...PC      www.stata.com
                                                www.stata.com

MP - Parallel Edition

3-user 8-core
Serial
Licensed to:

Notes:
1. (-set maxvar-) 5000 maximum variables
```

Below the Results window is the Command window, which is currently empty. On the right side of the interface, there are two panels: the Variables panel and the Properties panel. The Variables panel has a search bar and a table with columns for Name and Label. The Properties panel shows details for the current variable, including Name, Label, Type, Format, Value Label, Notes, and Data summary (Filename, Label, Notes, Variables, Observations, Size, Memory, Sorted by).

Name	Label
------	-------

Properties	
▼ Variables	
Name	
Label	
Type	
Format	
Value Label	
Notes	
▼ Data	
► Filename	
Label	
Notes	
Variables	0
Observations	0
Size	0
Memory	64M
Sorted by	

The Stata screen

Stata/MP 13.0

Open Save Print Log Viewer Graph Do-file Editor Data Editor Data Browser More Break Search Help

Results

```
_____ / _____ / _____ / _____ /  
_____ / _____ / _____ / _____ /  
Statistics/Data Analysis 13.0 Copyright 1985-2013 StataCorp LP  
StataCorp  
4905 Lakeway Drive  
College Station, Texas 77845 USA  
800-STATA-PC http://www.stata.com  
979-696-4600 stata@stata.com  
979-696-4601 (fax)  
  
3-user 8-core Stata network perpetual license:  
Serial number:  
Licensed to:  
  
Notes:  
1. (-set maxvar-) 5000 maximum variables
```

Command

Variables

Enter filter text here

Name	Label
------	-------

List of your variables

Properties

Variables

Name	
Label	
Type	
Format	
Value Label	
Notes	

Data

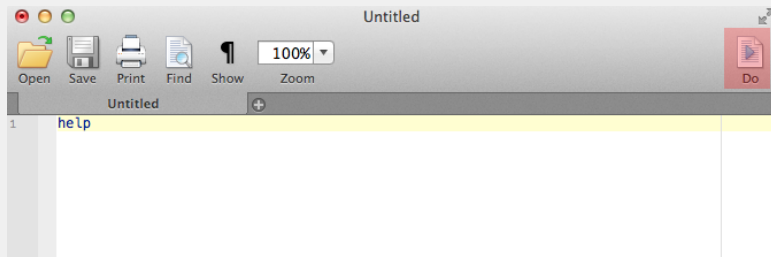
Filename	
Label	
Notes	
Variables	0
Observations	0
Size	0
Memory	64M
Sorted by	

crucial

Do-files

To execute the whole do-file, click on "Do"

To execute part of the do-file, select the commands you want to execute, then click on "Do"



Use comments in your do-files, which start by `//`:

`//This is a comment`

To save the output of your do-files, you can create a **log-file**:

```
. log using  
"/Users/alexisgrigorieff/Desktop/QM1/first-lecture.log"
```

At the end of the session, you can close the log-file:

```
. log close
```

Today's dataset

Records of 100,000 slaves who were brought to Louisiana in the 18th and 19th century.

Data on the **slaves' name, gender, age, occupation, prices paid by slave owners etc.**

For more information, see:

http:

[//www.ibiblio.org/laslave/introduction.php](http://www.ibiblio.org/laslave/introduction.php)

Opening a Stata dataset

Datasets in Stata are saved as **.dta files**

Download a Stata dataset **from the web**:

```
. use "https://github.com/alexisgrigorieff/QM1/  
raw/master/slaves-trade.dta", clear
```

To open a Stata dataset **saved on your computer**:

```
. use  
"Users/alexisgrigorieff/Desktop/QM1/slave-trade.dta",  
clear
```

Opening Excel and .csv datasets

To open an **Excel dataset**:

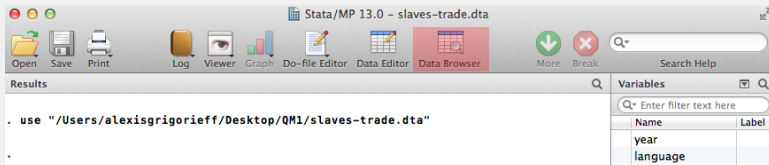
```
. import excel  
"/Users/alexisgrigorieff/Desktop/QM1/slave-trade.xlsx",  
sheet("Sheet1") firstrow clear
```

To open a **.csv dataset**:

```
. import delimited  
"/Users/alexisgrigorieff/Desktop/QM1/slave-trade.csv"
```

Exploring the dataset

It is very important to look at the data before doing any analysis



You can uncheck the variables that you want to hide

The screenshot shows the Stata Data Editor window. The title bar reads "Data Editor (Browse) - slaves-trade.dta". The toolbar includes icons for Edit, Browse, Filter, Variables, Properties, and Snapshots. The main window displays a table with the following columns: year, language, estatapo, seller, buyer, name, sex, race, and a checkbox column. The data is as follows:

	year	language	estatapo	seller	buyer	name	sex	race	
1	1726	1	0	Lebas de Courmont	Olivier (FREE BLACK	Cassova	1	3	<input type="checkbox"/>
2	1748	1	.	Gouvion	Huber dit LaCroix	Pocede or Jean Baptiste	0	3	<input type="checkbox"/>
3	1758	1	1		Arnoult	Apollon	0	3	<input type="checkbox"/>
4	1765	1	0		Forestal	Jannot	0	3	<input type="checkbox"/>
5	1765	1	0		Forestal	Pierre	0	3	<input type="checkbox"/>
6	1765	1	0		Beauregarde	Joseph	0	3	<input type="checkbox"/>

On the right, the Variables panel shows a search bar and a list of variables: "Name", "Label", "year", "language", "estatapo", "seller", and "buyer". The checkboxes for "year", "language", "estatapo", "seller", and "buyer" are checked.

Exploring the dataset

```
. summarize age
```

```
. sum age
```

Variable	Obs	Mean	Std. Dev.	Min	Max
age	3796	29.47321	10.71684	6	70

```
. inspect sex
```

```
. inspect sex
```

sex:

			Number of Observations		
			Total	Integers	Nonintegers
#		Negative	-	-	-
#		Zero	2445	2445	-
#		Positive	1351	1351	-
#	#				
#	#	Total	3796	3796	-
#	#	Missing	-		
			3796		
0		1			
(2 unique values)					

Exploring the dataset

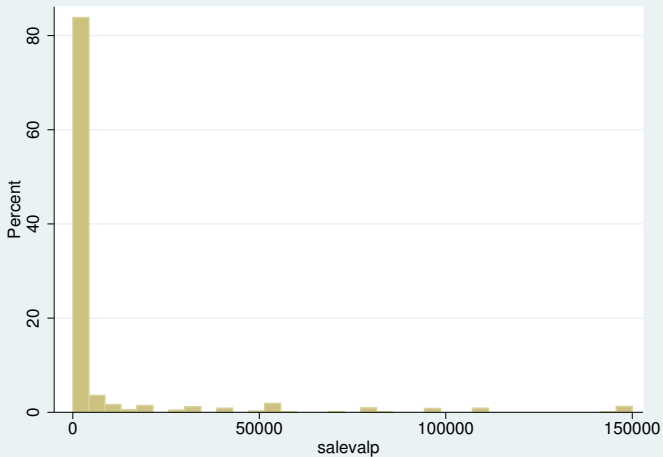
```
. tabulate race
```

```
. tab race
```

race	Freq.	Percent	Cum.
1	42	1.11	1.11
3	3,491	91.97	93.07
4	205	5.40	98.47
5	3	0.08	98.55
8	1	0.03	98.58
9	54	1.42	100.00
Total	3,796	100.00	

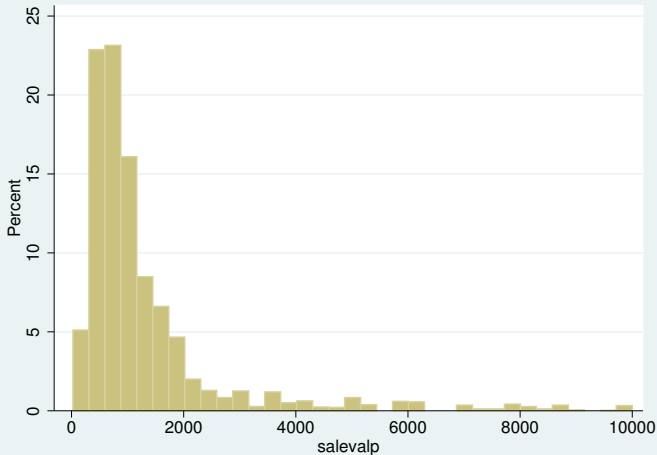
Exploring the dataset

```
. histogram salevalp, percent
```



Exploring the dataset

```
. histogram salevalp if salevalp <= 10000,  
percent
```



Before starting the analysis, it is important to **clean the data**:

- Replace a value with another one:
`. replace race=. if race==9`
- Create a new variable:
`. gen adult=1 if age>=18`
`. replace adult=0 if adult==.`
- Rename a variable:
`. rename sex female`

Before starting the analysis, it is important to **clean the data**:

- Label your variables:
 `. label variable female "female=1 male=0"`
- Drop some observations:
 `. drop if salevalp>10000`
- Drop a variable:
 `. drop brut`

Saving the new dataset

Once the dataset is cleaned, **save it in a different .dta file**:

```
. save "Users/alexisgrigorieff/Desktop/QM1/  
slave-trade-cleaned.dta", replace
```

Useful resources

- **Stata's help:**

`. help histogram`

- **Stata's website:**

<http://www.stata.com/features/documentation/>

- **UCLA website:**

<http://www.ats.ucla.edu/stat/stata/>

- **StackOverflow:**

<http://stackoverflow.com/questions/tagged/stata>

- **my email:**

alexis.grigorieff@economics.ox.ac.uk