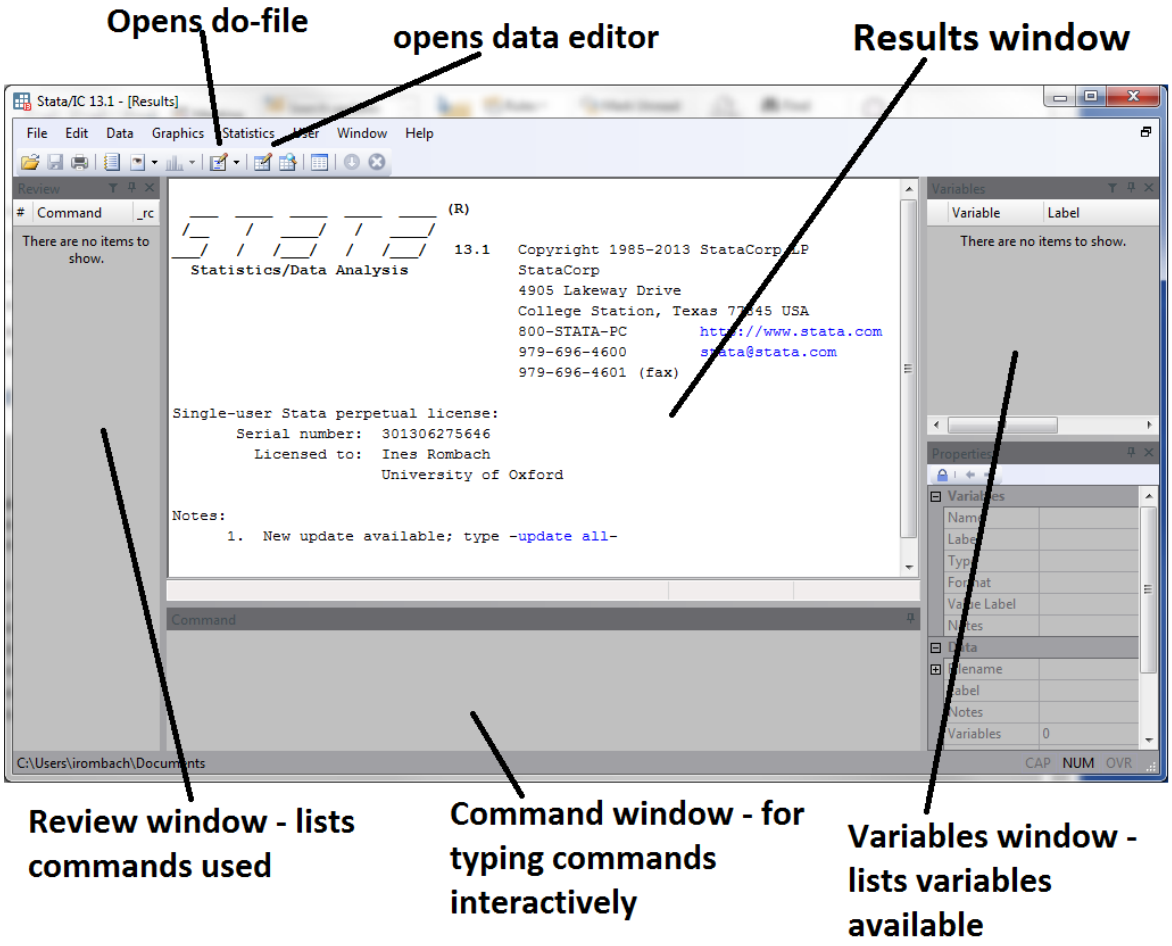
Data analysis: A beginner's tour of Stata

Stata is a statistical software package used for data management and statistical analysis.

SPSS, SAS and R are alternative statistical software packages available to statisticians.

Opening Stata: You can open Stata by clicking on Start → All Programmes → Stata 13



Stata can be used interactively, either by using to menu (point and click) to perform the required tasks. However, in this session we will focus on using commands/ programming to do this. We will use the do-file for all programming. This allows up to

- Keep a record of all commands run. This is particularly important for complex projects.
- Re-run commands easily if needed.
- Adapt previous code to new datasets/ additional variables.

Stata do-files

Stata do-files are opened by clicking on the relevant icon:

This can be found at the top menu of the Stata window. 

Stata is case sensitive. When you come to typing your own commands, Stata will only accept commands in lower case. If you define new variables and capitalise them, you need to refer to them in exactly the same way subsequently.

Stata data editor

The Stata data editor can be opened by clicking on the relevant icon: 
Once data has been read into Stata, all observations and variables can be seen in this window.

Exercises:

1) Open Stata
2) Open a do-file and the data editor, as shown above. Arrange all three windows so that you can see them at the same time on your screen.
   Remind yourself of the purpose of the different sections on the main Stata window.
3) In the do-file window, click on file → Open → File and open the following file:
   "H:\Stata_beginners\StataProgs\Exercise.do"
   Comments, which provide information on the purpose of each command, are shown in green. These cannot be executed (i.e. run) and have no effect on the results window or data editor.
4) The Stata commands are coloured blue, black and red. These can be run by highlighting at least one character (letter) in the relevant line and clicking on the 'execute button':
   
   This button can be found at the far right in the line of icons at the top of the do-file.
   Run each command in turn. For each command, consider the questions below:

   a) After you have run the 'use' command, notice how the data is now shown in the Data Editor window.
      Variable names and variable labels are shown in the variables window at the right hand site of the main Stata window.
      Find out what the data in the variable named 'pop65p' represent.
      What information does the variable 'medage' contain?
      Below the variables window is the 'properties' window. Use the information displayed there to identify how many variables and observations are included in the 'census' dataset.
   b) The 'drop' command deletes variables. Check the Data Editor and the 'properties' window to confirm that the variable has been deleted.
      The command is repeated in the results window.
      Try to run the same command again. Note how now a red error message is displayed in the results window. This indicates that the command could not be run successfully. The error message (in red) tells you that this is because the relevant variable could not be found, as it had been deleted previously.
   c) The 'tabulate' command can be used to categorise categorical data.
      Observed how results are displayed in the results window. The Data Editor has not changed.
      How many of the states included in the dataset are classified as "South"?
   d) The 'summarize' command is used to summarise continuous variables.
      Observe how results are displayed in the results window.
      What is the mean number of marriages reported across all 10 states?
   e) The 'help' command can be useful to find out more about how to use a specific command. Have a look at the information it provides.
      Try to use the 'help' command to find out more about the 'use' command. You will need to type this command line yourself.
   f) The 'save' command saves the dataset in its current state (including data manipulations) in the specified location.
      Go to the specified folder to verify that the file has been saved successfully.

Data analysis:
A beginner's tour of Stata

Pradeep Virdee
courses@it.ox.ac.uk

iT Centre Learning

Your comfort is important

- The toilets are along the corridor outside the lecture rooms.
- The rest area is where you registered; it has vending machines and a water cooler.
- The seats at the computers are adjustable.
- You can adjust the monitors for height, tilt and brightness.

Today's arrangements

- Your teacher is: Pradeep Virdee
- Course duration: 1 hour
- You should have: Course notes

Today's objectives

- Understanding how Stata compares to other statistical packages
- Become familiar with the Stata interface, .do-files and data editor
- Understand the concept of writing code, and run code from the Stata .do-file
- Save .do-files

Your safety is important

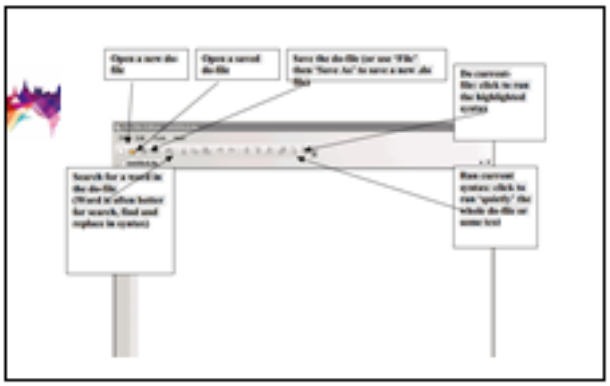
- Where is the fire exit?
- Beware of hazards:
  Tripping over bags and coats
- Please report any equipment faults to us
- Let us know if you have any other concerns

What can Stata do for me?

- Data management and analysis
- Carry out repetitive tasks easily
- Records your methods
- Create graphs
- Statistical techniques e.g. regression
- Simulations

Is Stata the right choice?

- Advantages:
  - Logical syntax
  - Large range of statistical analyses possible
  - Large user community
- Disadvantages
  - Expensive - check for student or group discounts in your department
  - Easiest to use with syntax



Alternatives

- R - Free
- SPSS - Point and click
- SAS - For very large datasets

Stata commands

- command *variable(s)* [if] [in], options

Programming facts and figures

- Stata has several hundred commands
- It is unnecessary to know all of Stata's commands to master it.
- Across these three courses we focus on the essentials

Troubleshooting

Today:

During each practical today, we will be able to help you with any problems you are experiencing.

After class

1. Make sure you correctly typed the command and variable name(s) - command names should appear blue in the .do file
2. If you know the exact name of command that you are having difficulties with try using Stata's help function.

Resources 1

- Try searching Stata's website which has more detailed information about the various commands. http://www.stata.com/
- Consult Stata's official documentation which can be accessed through the program (help menu) or download it free of charge in PDF format on the official website. http://www.stata.com/support/documentation/
- Search Stata's official mailing list archive. This resource is extremely useful. http://www.stata.com/statalist/archive/

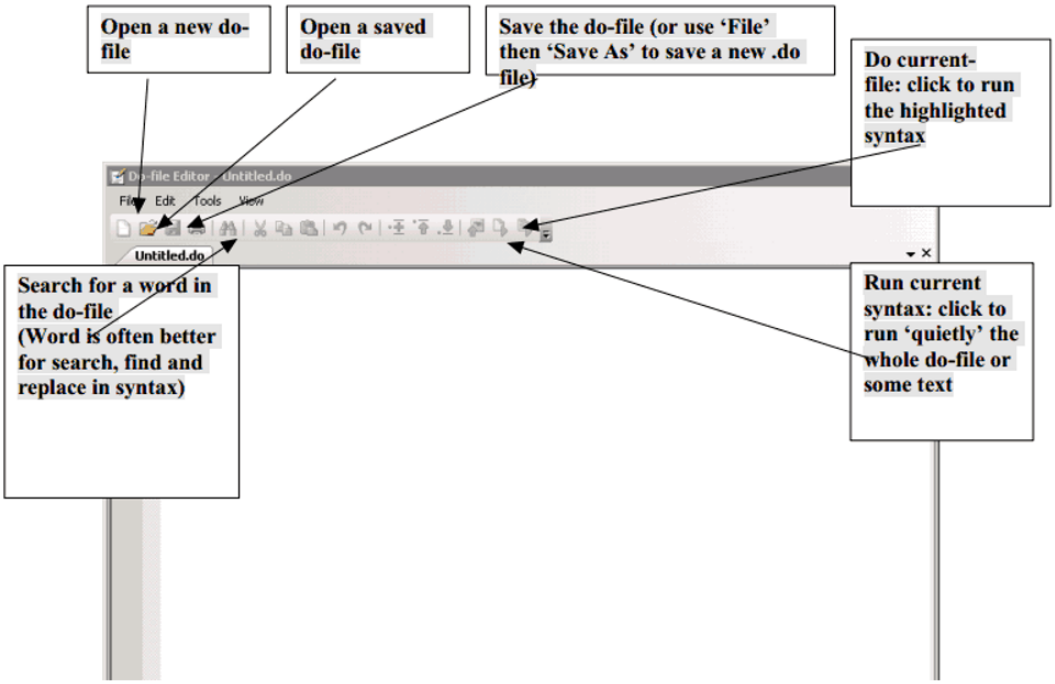
Resources 2

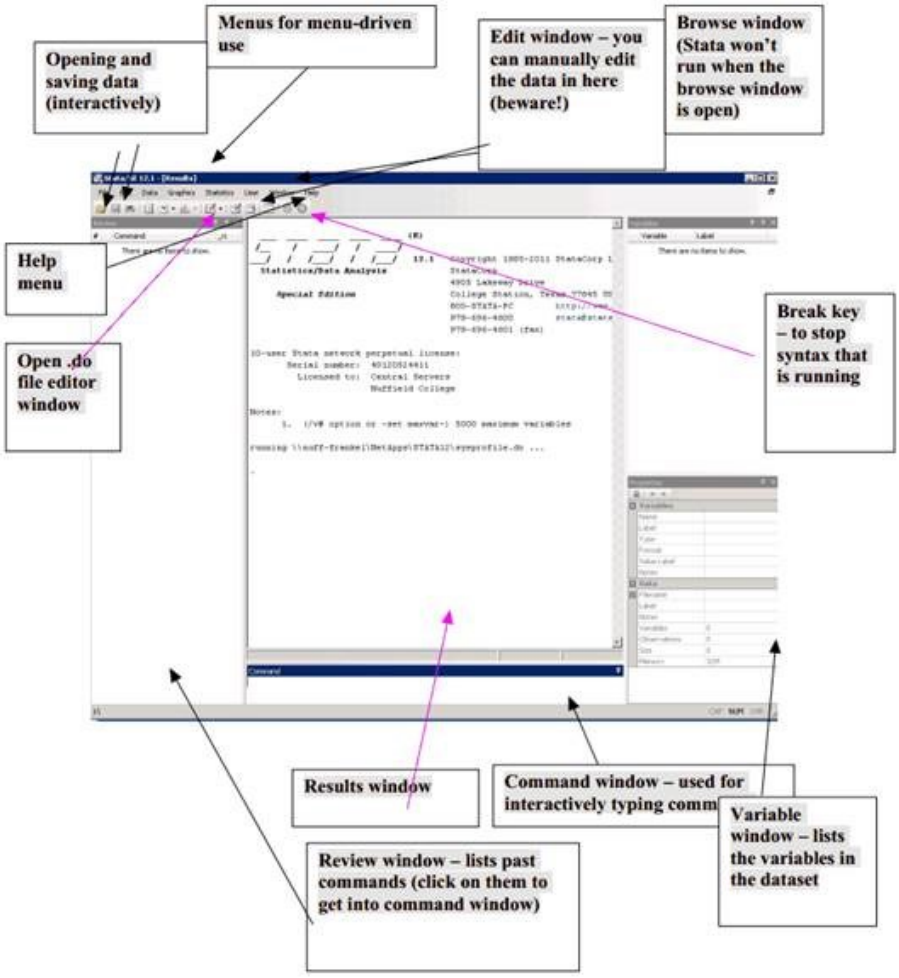
- Sign up for Stata's official mailing list and ask a question there. http://www.stata.com/statalist/
- Consult one of the many Stata textbooks, many of which are available at our libraries. http://www.stata.com/bookstore/
- Search other internet forums or just google it! https://stackoverflow.com/questions/tagged/stata

IT Learning Centre courses

Today:	Data management & analysis in Stata
Further:	Data manipulation & analysis Graphics & analysis

Other relevant courses: See IT learning centre course list for courses on R, SPSS or statistics.


Open a new do-file
Open a saved do-file
Save the do-file (or use 'File' then 'Save As' to save a new .do file)
Do current-file: click to run the highlighted syntax
Do-file Editor - Untitled.do
File Edit Tools View
Untitled.do
Search for a word in the do-file (Word is often better for search, find and replace in syntax)
Run current syntax: click to run 'quietly' the whole do-file or some text


Menus for menu-driven use
Opening and saving data (interactively)
Edit window – you can manually edit the data in here (beware!)
Browse window (Stata won't run when the browse window is open)
Help menu
Open .do file editor window
Break key – to stop syntax that is running
Results window
Command window – used for interactively typing comm
Variable window – lists the variables in the dataset
Review window – lists past commands (click on them to get into command window)