Statistical Tools for Research: SPSS, R and STATA.

Introduction to Stata

Md Monowar Hossain

Institute of Statistical Research and Training (ISRT)

August 30, 2025

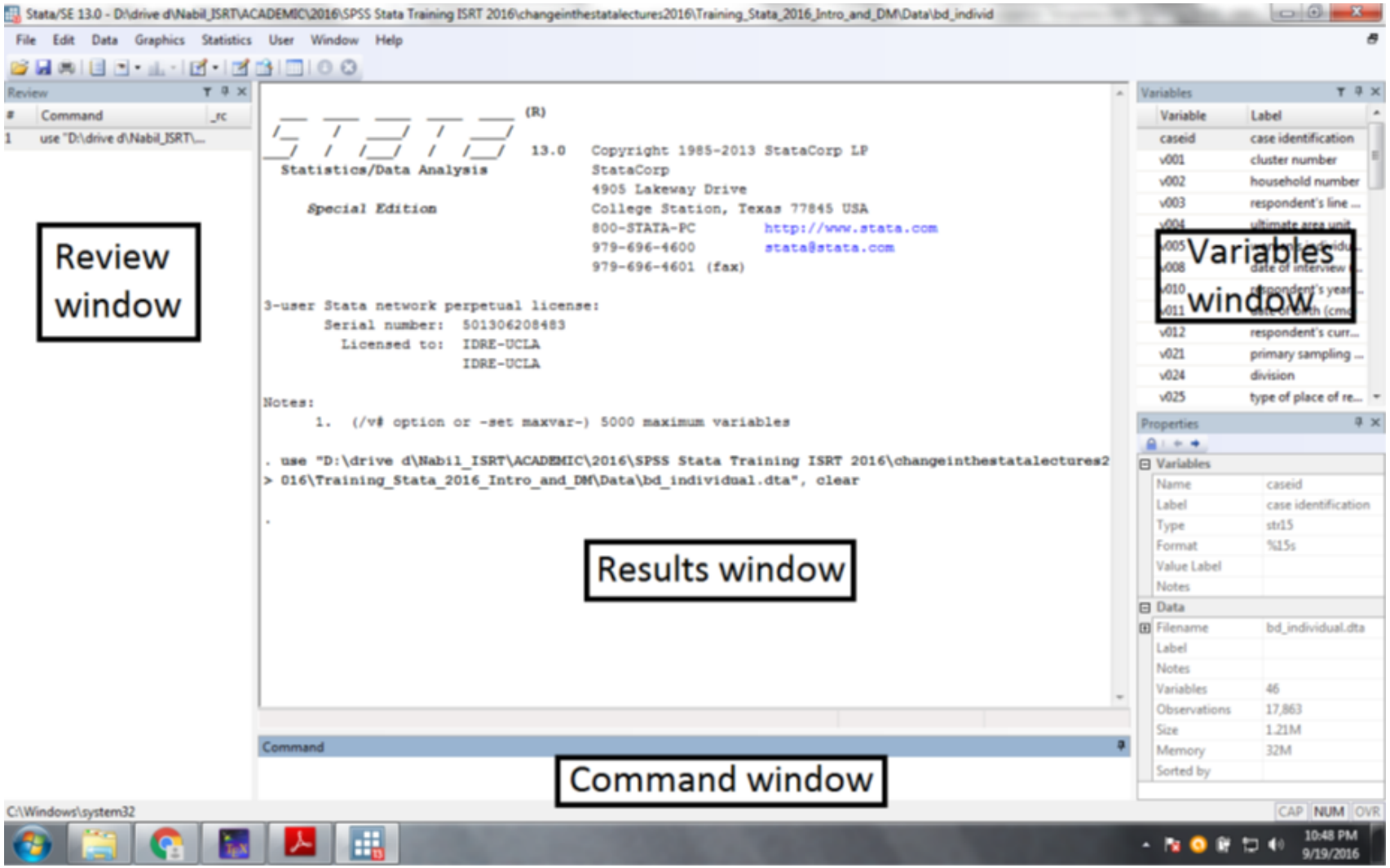
About Stata software

- Stata is a software package used for statistical analysis. The name Stata is a syllabic abbreviation of the words statistics and data.
- Stata was developed by StataCorp in 1985.
- Stata has always emphasized a command-line interface, which facilitates replicable analyses. Starting with version 8.0, however, Stata has included a graphical user interface which uses menus and dialog boxes.
- The latest release is Stata version 18, released in April, 2023.
- Purchase link: https://www.stata.com/order/dev/

Overview of Stata file extensions

- Stata data file: .dta, example: mydata.dta
- Stata syntax file: .do, example: mysyntax.do
- Stata output file: .smcl, example: myoutput.smcl, smcl stands for
- Stata Markup and Control Language

Stata windows



Stata windows

There is an additional window for graph that appears when a graph is produced in Stata.

- Command window: where commands are entered. Press the Enter key on your keyboard to execute any command.
- Results window: where results appear (i.e, are printed).
- Variables window: lists the variables in the current dataset.
- Review window: lists past (i.e, already executed) commands.
- Graph window: where graphs are displayed (appears when graphs are generated).

GUI and CLI

- You can either use the menu-based **Graphical User Interface** (GUI) or the command-driven **Command Line Interface** (CLI) to interact with Stata. But there are some advantages of using CLI:
  - **Command syntax** has the advantage of reproducibility.
  - Many of the commands are short and easy to execute, and as proficiency grows, it is often faster to type commands rather than use the menu.

We will mostly cover command syntax in this course. We will be providing commands so that you will become familiar with them. The commands will be useful for constructing longer and more complicated programs.

- The command executes if it is entered properly
- Issues an error message (in red) if it cannot figure out what to do, or if there is something which could inadvertently change the dataset.
- If Stata gives you an error, it will provide you with a return code, r(#). You can find out more about the error by typing: lookup rc #.
- To get help regarding any command, e.g. the regression command, type: help regression

Directory management

- To check the present working directory:

```
pwd
```

- To change the current working directory:

```
cd "E:\"
```

- To make a new directory inside the current working directory:

```
mkdir "E:\StataClass1"
```

- To make the newly created directory as working directory:

```
cd "E:\StataClass1"
```

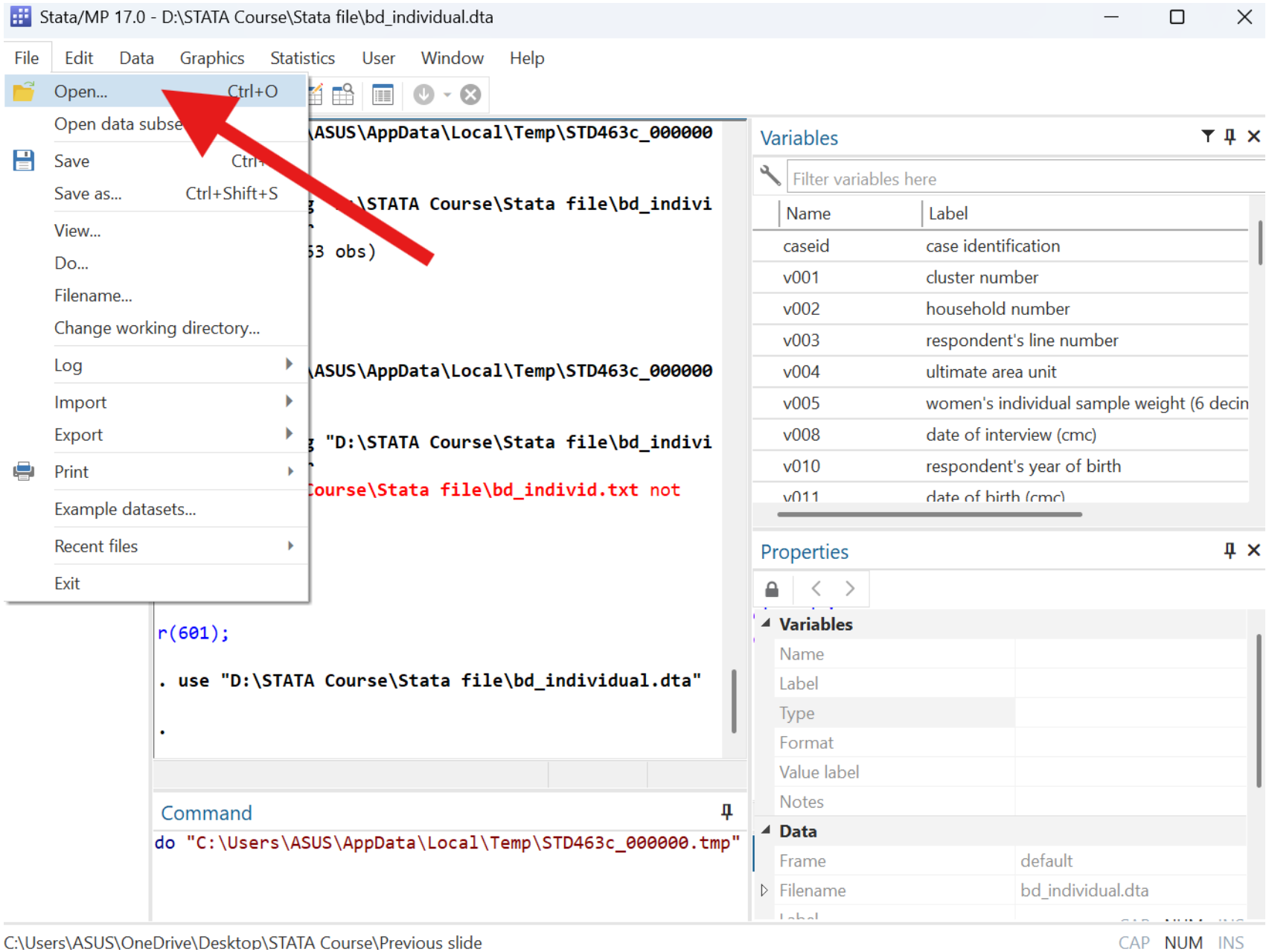
Reading a Stata data file

- Reading a data file from a particular location:

```
use "E:\bd_individual.dta", clear
```

- Reading a data file from a working directory:

```
use bd_individual.dta, clear
```


Stata/MP 17.0 - D:\STATA Course\Stata file\bd_individual.dta
File
Edit
Data
Graphics
Statistics
User
Window
Help
Open... Ctrl+O
Open data subset...
Save Ctrl+S
Save as... Ctrl+Shift+S
View...
Do...
Filename...
Change working directory...
Log
Import
Export
Print
Example datasets...
Recent files
Exit
\ASUS\AppData\Local\Temp\STD463c_000000
\STATA Course\Stata file\bd_indivi
53 obs)
\ASUS\AppData\Local\Temp\STD463c_000000
"D:\STATA Course\Stata file\bd_indivi
Course\Stata file\bd_individ.txt not
r(601);
. use "D:\STATA Course\Stata file\bd_individual.dta"
.
Command
do "C:\Users\ASUS\AppData\Local\Temp\STD463c_000000.tmp"
Variables
Filter variables here
Name
Label
caseid case identification
v001 cluster number
v002 household number
v003 respondent's line number
v004 ultimate area unit
v005 women's individual sample weight (6 decin
v008 date of interview (cmc)
v010 respondent's year of birth
v011 date of birth (cmc)
Properties
Variables
Name
Label
Type
Format
Value label
Notes
Data
Frame default
Filename bd_individual.dta
C:\Users\ASUS\OneDrive\Desktop\STATA Course\Previous slide
CAP NUM INS

Reading a Stata data file from a website

```
use http://www.stata-press.com/data/r14/systolic, clear

webuse query

webuse lifeexp

webuse apple

webuse set http://www.zzz.edu/users/~sue
```

Importing data files

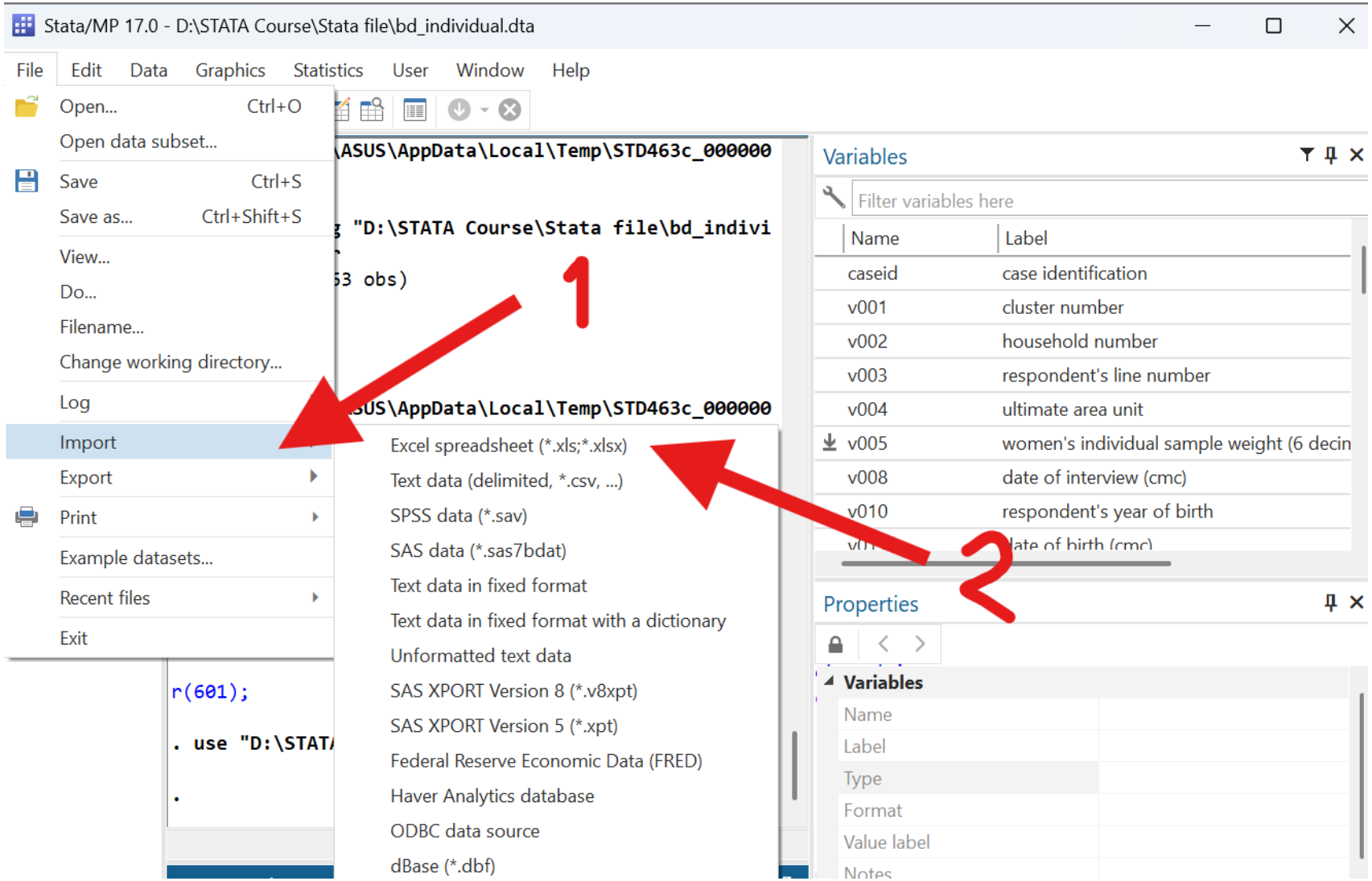
- We can read dataset in .csv or .txt format, by using the following commands:

```
insheet using bd_individ.csv, clear
insheet using bd_individ.txt, clear
```

- To read an excel file we will use the following command:

```
import excel bd_individ.xls, ///
sheet("Sheet1") firstrow clear
```

- Here, firstrow means we can use variable names contained in the first row of the excel file as the variable names of the Stata file.


Stata/MP 17.0 - D:\STATA Course\Stata file\bd_individual.dta
File
Edit
Data
Graphics
Statistics
User
Window
Help
Open... Ctrl+O
Open data subset...
Save Ctrl+S
Save as... Ctrl+Shift+S
View...
Do...
Filename...
Change working directory...
Log
Import
Export
Print
Example datasets...
Recent files
Exit
Excel spreadsheet (*.xls;*.xlsx)
Text data (delimited, *.csv, ...)
SPSS data (*.sav)
SAS data (*.sas7bdat)
Text data in fixed format
Text data in fixed format with a dictionary
Unformatted text data
SAS XPORT Version 8 (*.v8xpt)
SAS XPORT Version 5 (*.xpt)
Federal Reserve Economic Data (FRED)
Haver Analytics database
ODBC data source
dBase (*.dbf)
1
2
Variables
Filter variables here
Name
Label
caseid
case identification
v001
cluster number
v002
household number
v003
respondent's line number
v004
ultimate area unit
v005
women's individual sample weight (6 decin
v008
date of interview (cmc)
v010
respondent's year of birth
Properties
Variables
Name
Label
Type
Format
Value label
Notes

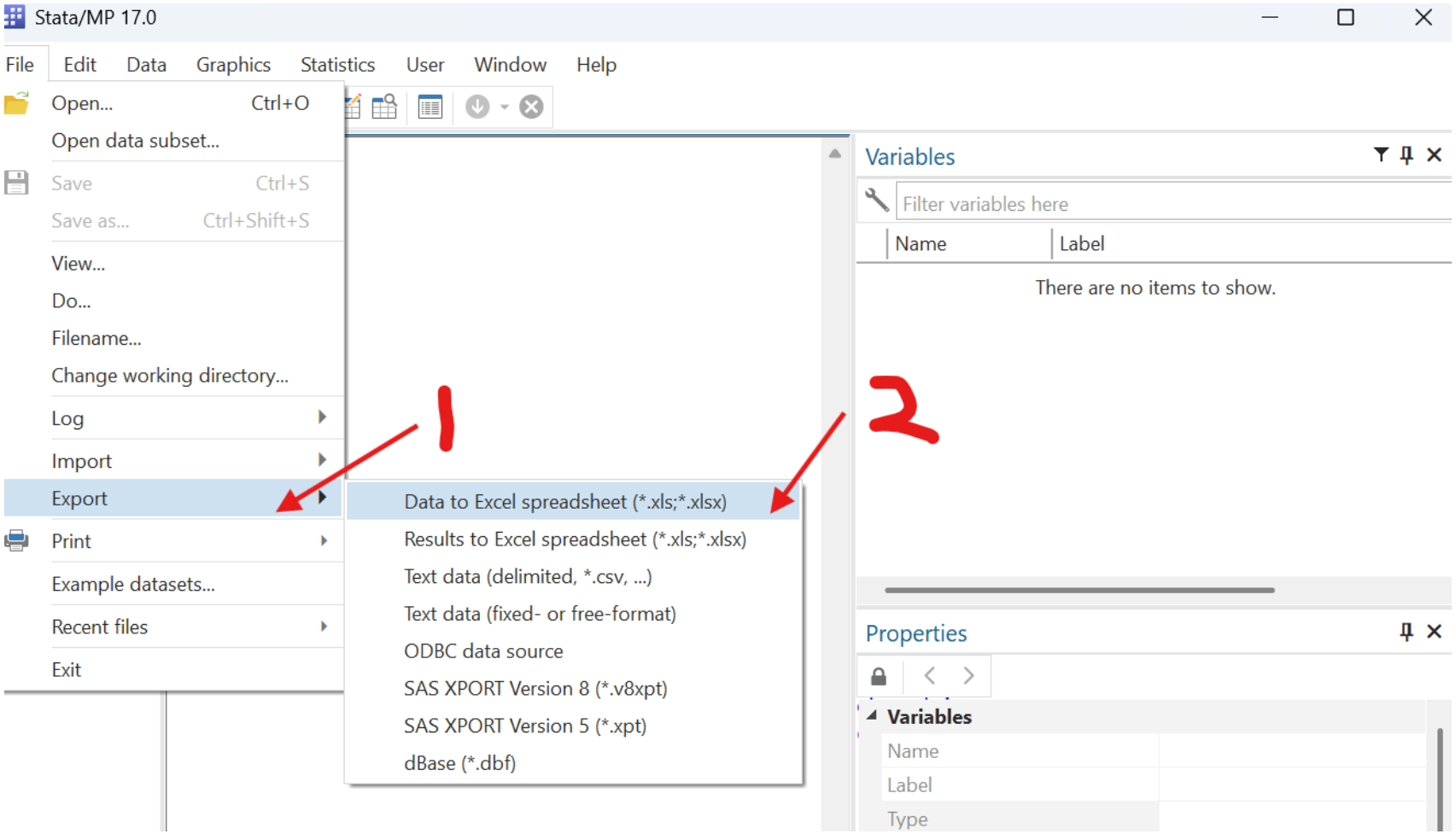
Exporting data files

- Often, we need to save the entered dataset as .csv or .txt format, so that it can be read in different software.
- To create such a .csv file, we have to export our data as follows:

```
outsheet using "E:\bd_individ.csv",comma nolabel replace
```

- Here, comma is the delimiter, nolabel means we want the values of the labeled variables as outputs, not the labels themselves, replace replaces any previous file in the same name.
- To export the data into .txt format we can use:

```
outsheet using "E:\bd_individ.txt", nolabel replace
```


Stata/MP 17.0
File
Edit
Data
Graphics
Statistics
User
Window
Help
Open... Ctrl+O
Open data subset...
Save Ctrl+S
Save as... Ctrl+Shift+S
View...
Do...
Filename...
Change working directory...
Log
Import
Export
Print
Example datasets...
Recent files
Exit
Data to Excel spreadsheet (*.xls;*.xlsx)
Results to Excel spreadsheet (*.xls;*.xlsx)
Text data (delimited, *.csv, ...)
Text data (fixed- or free-format)
ODBC data source
SAS XPORT Version 8 (*.v8xpt)
SAS XPORT Version 5 (*.xpt)
dBase (*.dbf)
1
2
Variables
Filter variables here
Name
Label
There are no items to show.
Properties
Variables
Name
Label
Type

Exporting data files

- If we want it to save the data in excel format, then we can write the following syntaxes:

```
export excel using bd_individ.xls, ///
firstrow(variables) nolabel replace
```

- Here, firstrow(variables) indicates we want to save the variable names into the first row of the excel file.

Entering data into Stata

- You can enter data into Stata using **Graphical User Interface** (GUI) by clicking the Edit button or by typing edit in the command window.
- Also, we can use a do file to enter data by writing the following syntaxes:

```
input hhid sex str10 location
10030 1 urban
10031 0 rural
10032 1 rural
end
```

where hhid and sex are numeric variables and location is string variable.

- For any string variable we need to specify the maximum number of characters the string will have using str10, where 10 indicates that location can have maximum characters 10.

Entering data using Data Editor

- Now enter the previous data.

Data editor

- You are allowed to enter, view, or edit your data in "data editor". It looks like a spreadsheet.
- Typically, variables are listed across the top, and cases are listed down the side.
- To just have a view of your data, you need to type the command browse in command windows; in the browse mode you cannot change your data.
- To enter or edit your data you need to type edit which will take you to the edit mode.

Do file

- A do file contains a list of Stata commands.
- The do file can be created using Stata's do file editor.

THANK YOU