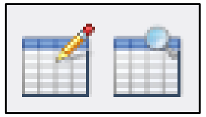
Crash Program on Econometrics Application

Session 1: Introduction to Stata

Farhin Islam
Research Associate, SANEM

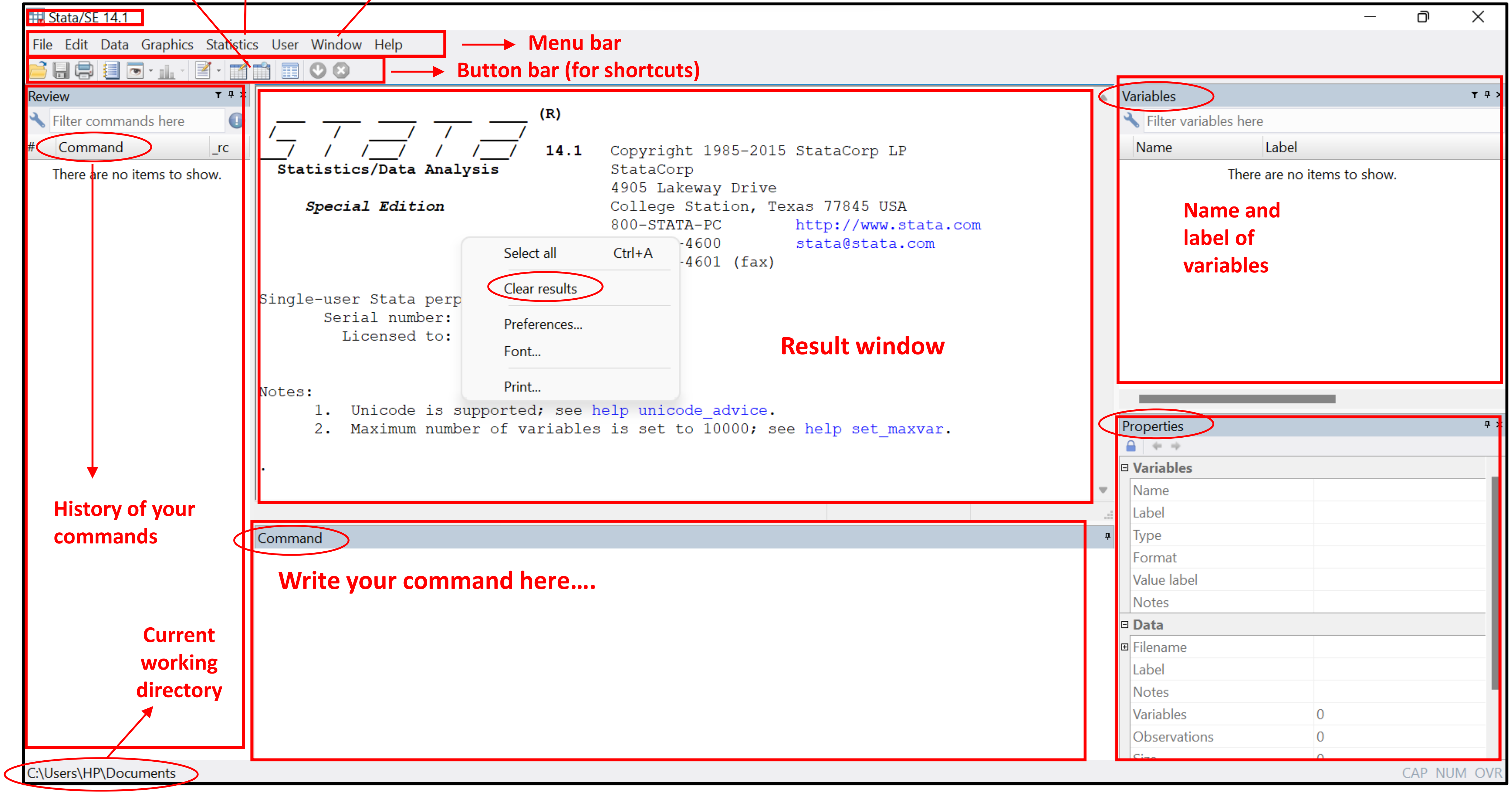
29 August 2022

Stata Interface



For analysis

Find missing window from here



```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   14.1   Copyright 1985-2015 StataCorp LP
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
     Special Edition                  College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      -4600               stata@stata.com
                                      -4601 (fax)

Single-user Stata perp
       Serial number:
         Licensed to:

Notes:
      1.  Unicode is supported; see help unicode_advice.
      2.  Maximum number of variables is set to 10000; see help set_maxvar.

.
```

Menu bar

Button bar (for shortcuts)

Result window

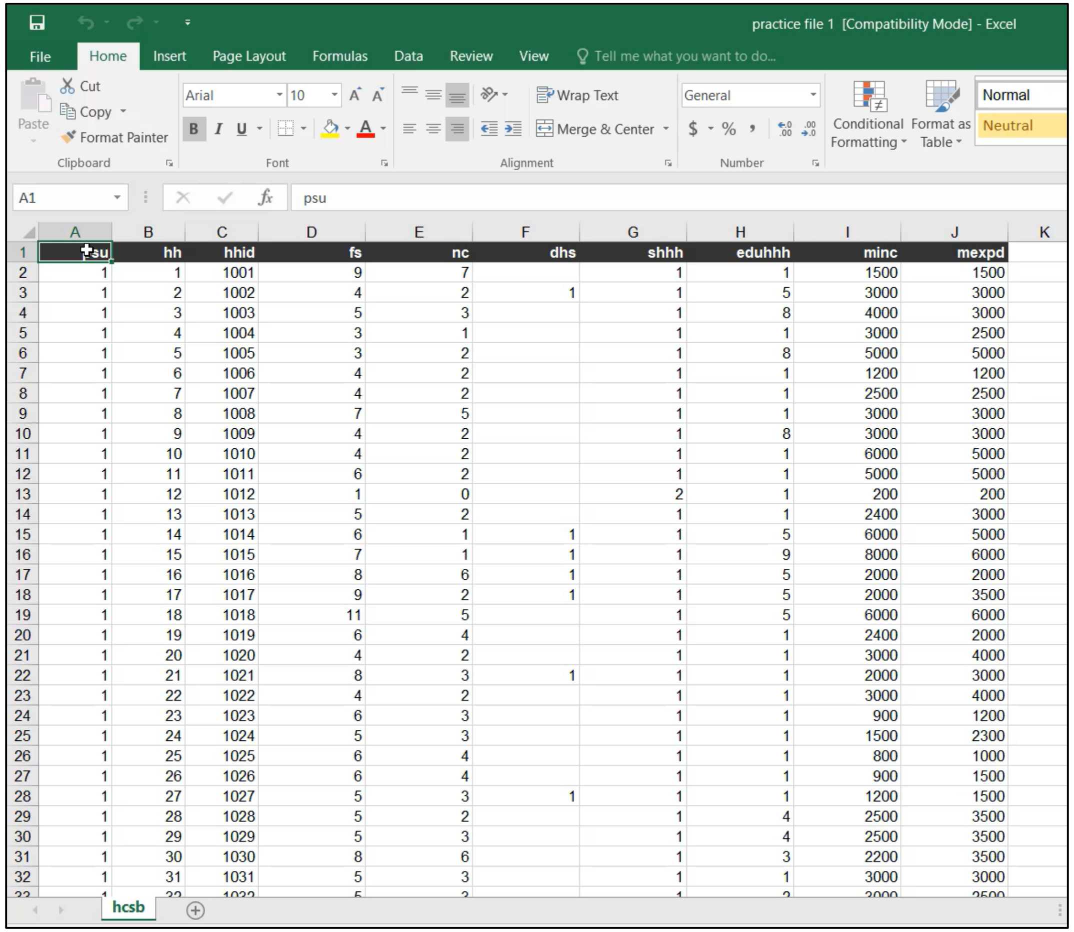
Name and label of variables

History of your commands

Write your command here....

Current working directory

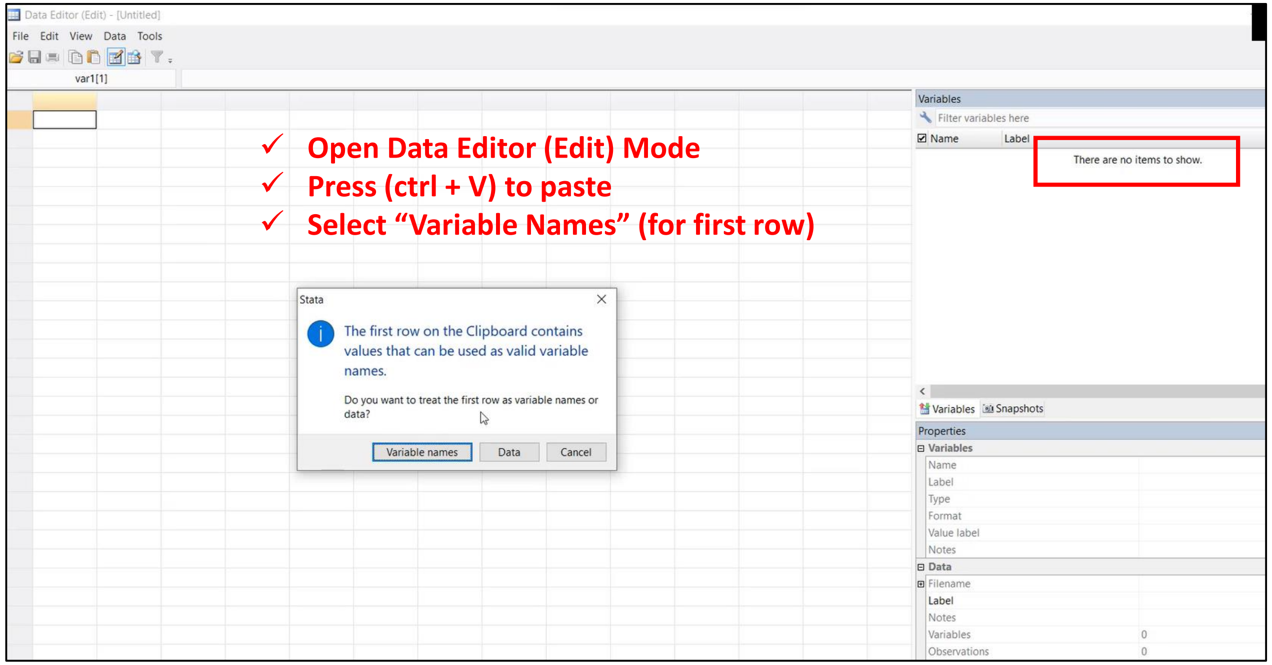
Import data from Excel



A1 psu

	psu	hh	hhid	fs	nc	dhs	shhh	eduhhh	minc	mexpd
2	1	1	1001	9	7		1	1	1500	1500
3	1	2	1002	4	2	1	1	5	3000	3000
4	1	3	1003	5	3		1	8	4000	3000
5	1	4	1004	3	1		1	1	3000	2500
6	1	5	1005	3	2		1	8	5000	5000
7	1	6	1006	4	2		1	1	1200	1200
8	1	7	1007	4	2		1	1	2500	2500
9	1	8	1008	7	5		1	1	3000	3000
10	1	9	1009	4	2		1	8	3000	3000
11	1	10	1010	4	2		1	1	6000	5000
12	1	11	1011	6	2		1	1	5000	5000
13	1	12	1012	1	0		2	1	200	200
14	1	13	1013	5	2		1	1	2400	3000
15	1	14	1014	6	1	1	1	5	6000	5000
16	1	15	1015	7	1	1	1	9	8000	6000
17	1	16	1016	8	6	1	1	5	2000	2000
18	1	17	1017	9	2	1	1	5	2000	3500
19	1	18	1018	11	5		1	5	6000	6000
20	1	19	1019	6	4		1	1	2400	2000
21	1	20	1020	4	2		1	1	3000	4000
22	1	21	1021	8	3	1	1	1	2000	3000
23	1	22	1022	4	2		1	1	3000	4000
24	1	23	1023	6	3		1	1	900	1200
25	1	24	1024	5	3		1	1	1500	2300
26	1	25	1025	6	4		1	1	800	1000
27	1	26	1026	6	4		1	1	900	1500
28	1	27	1027	5	3	1	1	1	1200	1500
29	1	28	1028	5	2		1	4	2500	3500
30	1	29	1029	5	3		1	4	2500	3500
31	1	30	1030	8	6		1	3	2200	3500
32	1	31	1031	5	3		1	1	3000	3000

- ✓ Select data (ctrl + A)
- ✓ Copy data (ctrl + C)



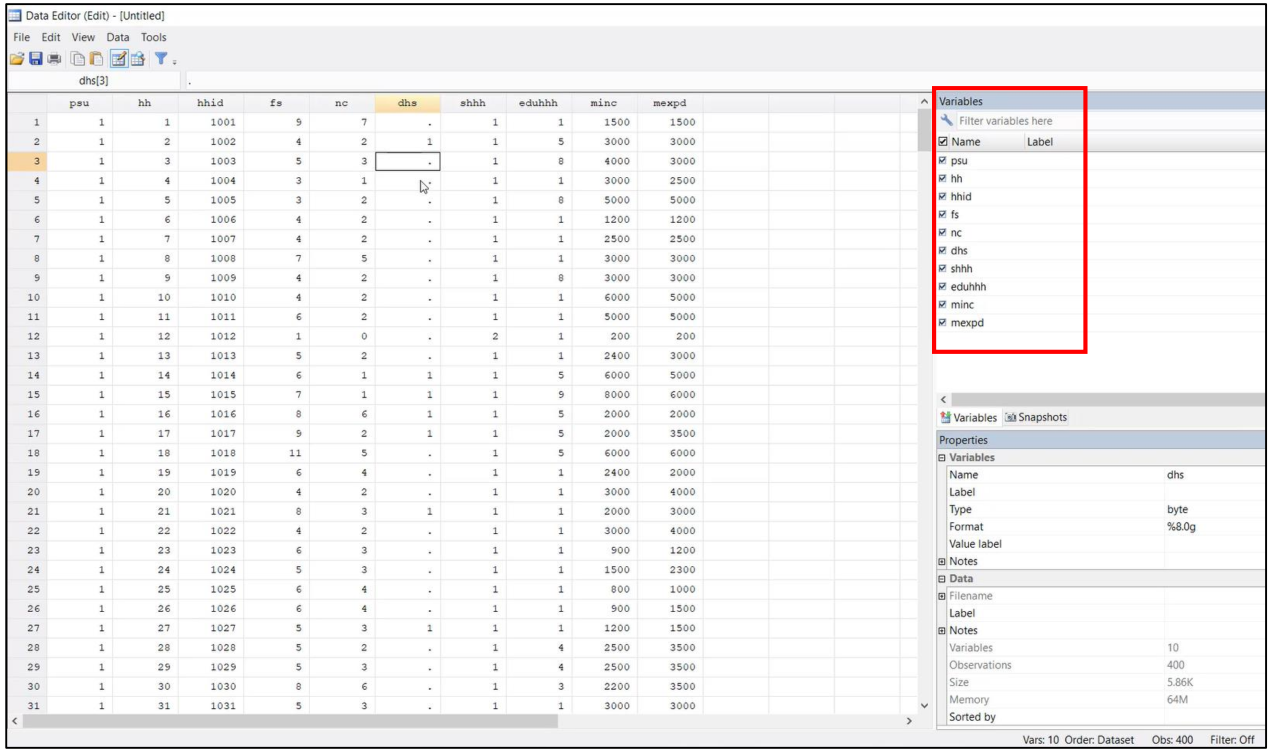
There are no items to show.

Stata

The first row on the Clipboard contains values that can be used as valid variable names.

Do you want to treat the first row as variable names or data?

- ✓ Open Data Editor (Edit) Mode
- ✓ Press (ctrl + V) to paste
- ✓ Select “Variable Names” (for first row)



dhs[3] .

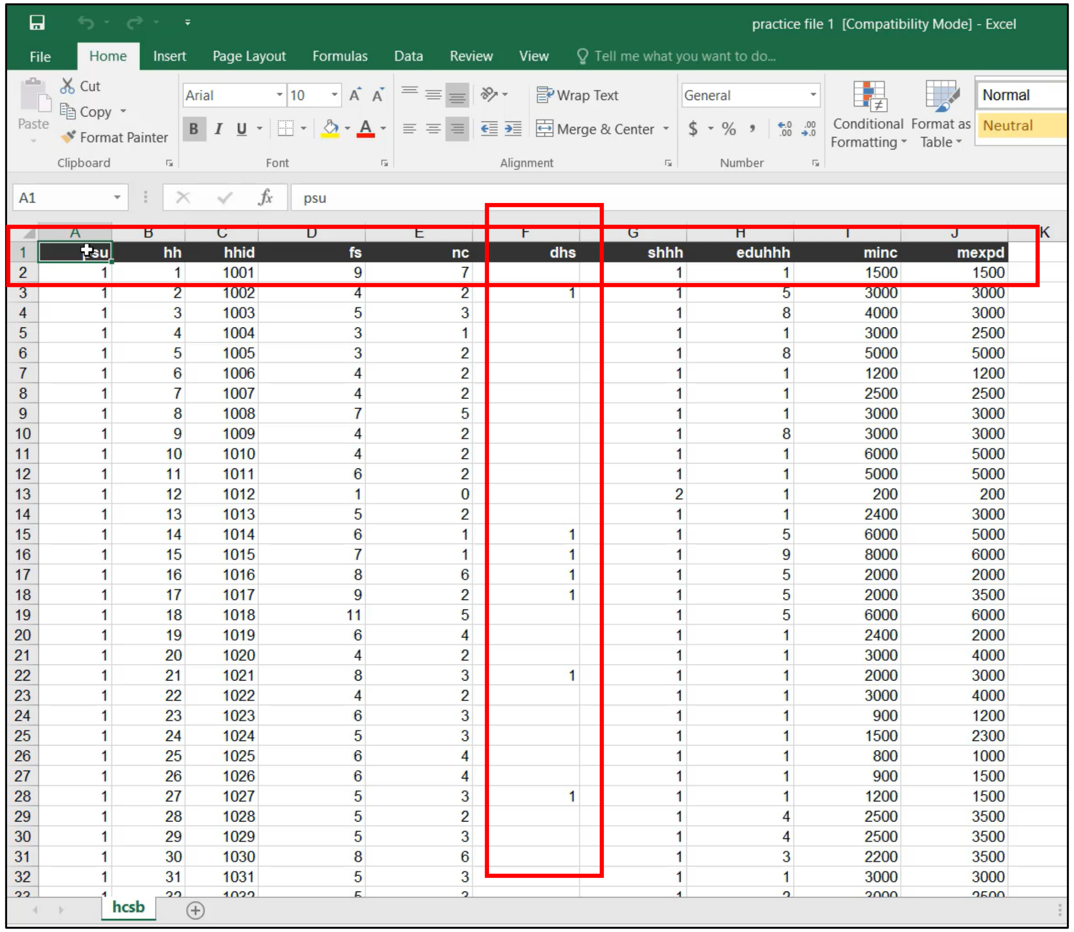
	psu	hh	hhid	fs	nc	dhs	shhh	eduhhh	minc	mexpd
1	1	1	1001	9	7	.	1	1	1500	1500
2	1	2	1002	4	2	1	1	5	3000	3000
3	1	3	1003	5	3	.	1	8	4000	3000
4	1	4	1004	3	1	.	1	1	3000	2500
5	1	5	1005	3	2	.	1	8	5000	5000
6	1	6	1006	4	2	.	1	1	1200	1200
7	1	7	1007	4	2	.	1	1	2500	2500
8	1	8	1008	7	5	.	1	1	3000	3000
9	1	9	1009	4	2	.	1	8	3000	3000
10	1	10	1010	4	2	.	1	1	6000	5000
11	1	11	1011	6	2	.	1	1	5000	5000
12	1	12	1012	1	0	.	2	1	200	200
13	1	13	1013	5	2	.	1	1	2400	3000
14	1	14	1014	6	1	1	1	5	6000	5000
15	1	15	1015	7	1	1	1	9	8000	6000
16	1	16	1016	8	6	1	1	5	2000	2000
17	1	17	1017	9	2	1	1	5	2000	3500
18	1	18	1018	11	5	.	1	5	6000	6000
19	1	19	1019	6	4	.	1	1	2400	2000
20	1	20	1020	4	2	.	1	1	3000	4000
21	1	21	1021	8	3	1	1	1	2000	3000
22	1	22	1022	4	2	.	1	1	3000	4000
23	1	23	1023	6	3	.	1	1	900	1200
24	1	24	1024	5	3	.	1	1	1500	2300
25	1	25	1025	6	4	.	1	1	800	1000
26	1	26	1026	6	4	.	1	1	900	1500
27	1	27	1027	5	3	1	1	1	1200	1500
28	1	28	1028	5	2	.	1	4	2500	3500
29	1	29	1029	5	3	.	1	4	2500	3500
30	1	30	1030	8	6	.	1	3	2200	3500
31	1	31	1031	5	3	.	1	1	3000	3000

Properties

Variables	
Name	dhs
Label	
Type	byte
Format	%8.0g
Value label	
Notes	
Data	
Filename	
Label	
Notes	
Variables	10
Observations	400
Size	5.86K
Memory	64M
Sorted by	

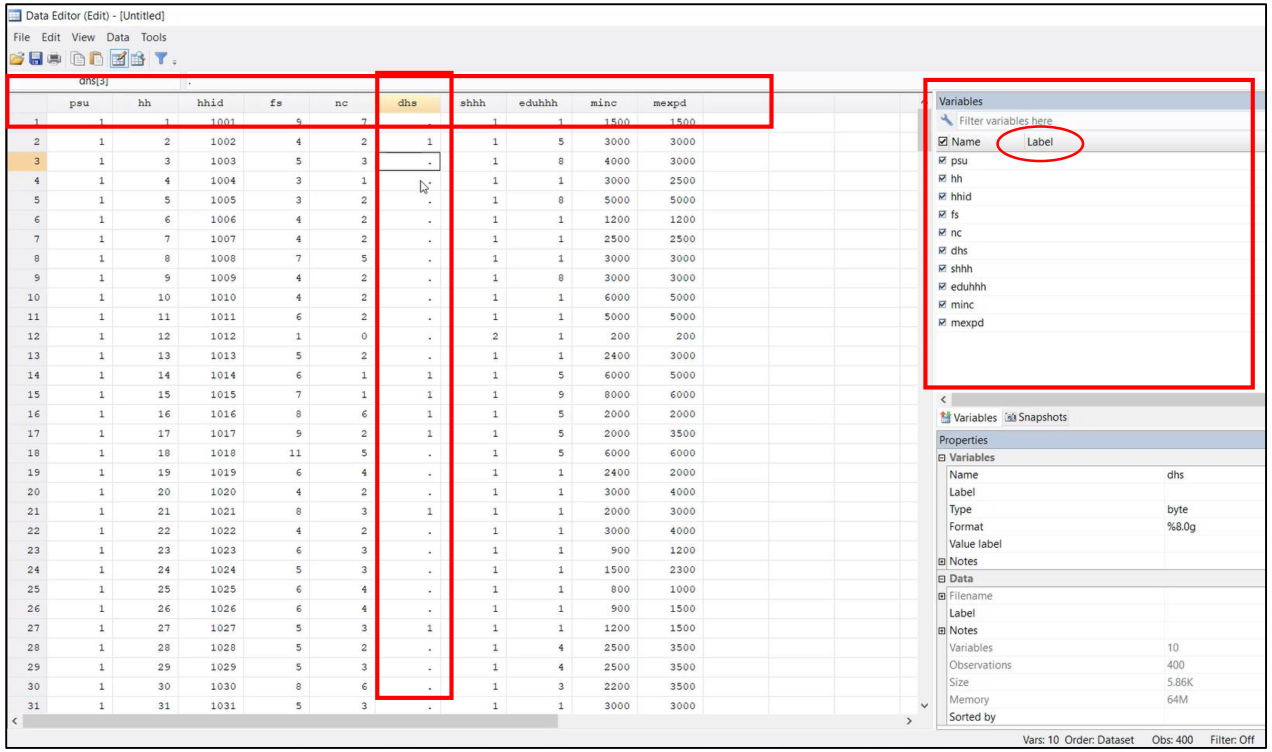
Vars: 10 Order: Dataset Obs: 400 Filter: Off

Import data from Excel



A1 psu

	A	B	C	D	E	F	G	H	I	J
1	psu	hh	hhid	fs	nc	dhs	shhh	eduhhh	minc	mexpd
2	1	1	1001	9	7		1	1	1500	1500
3	1	2	1002	4	2	1	1	5	3000	3000
4	1	3	1003	5	3		1	8	4000	3000
5	1	4	1004	3	1		1	1	3000	2500
6	1	5	1005	3	2		1	8	5000	5000
7	1	6	1006	4	2		1	1	1200	1200
8	1	7	1007	4	2		1	1	2500	2500
9	1	8	1008	7	5		1	1	3000	3000
10	1	9	1009	4	2		1	8	3000	3000
11	1	10	1010	4	2		1	1	6000	5000
12	1	11	1011	6	2		1	1	5000	5000
13	1	12	1012	1	0		2	1	200	200
14	1	13	1013	5	2		1	1	2400	3000
15	1	14	1014	6	1	1	1	5	6000	5000
16	1	15	1015	7	1	1	1	9	8000	6000
17	1	16	1016	8	6	1	1	5	2000	2000
18	1	17	1017	9	2	1	1	5	2000	3500
19	1	18	1018	11	5		1	5	6000	6000
20	1	19	1019	6	4		1	1	2400	2000
21	1	20	1020	4	2		1	1	3000	4000
22	1	21	1021	8	3	1	1	1	2000	3000
23	1	22	1022	4	2		1	1	3000	4000
24	1	23	1023	6	3		1	1	900	1200
25	1	24	1024	5	3		1	1	1500	2300
26	1	25	1025	6	4		1	1	800	1000
27	1	26	1026	6	4		1	1	900	1500
28	1	27	1027	5	3	1	1	1	1200	1500
29	1	28	1028	5	2		1	4	2500	3500
30	1	29	1029	5	3		1	4	2500	3500
31	1	30	1030	8	6		1	3	2200	3500
32	1	31	1031	5	3		1	1	3000	3000



dhs[3] .

	psu	hh	hhid	fs	nc	dhs	shhh	eduhhh	minc	mexpd
1	1	1	1001	9	7	.	1	1	1500	1500
2	1	2	1002	4	2	1	1	5	3000	3000
3	1	3	1003	5	3	.	1	8	4000	3000
4	1	4	1004	3	1	.	1	1	3000	2500
5	1	5	1005	3	2	.	1	8	5000	5000
6	1	6	1006	4	2	.	1	1	1200	1200
7	1	7	1007	4	2	.	1	1	2500	2500
8	1	8	1008	7	5	.	1	1	3000	3000
9	1	9	1009	4	2	.	1	8	3000	3000
10	1	10	1010	4	2	.	1	1	6000	5000
11	1	11	1011	6	2	.	1	1	5000	5000
12	1	12	1012	1	0	.	2	1	200	200
13	1	13	1013	5	2	.	1	1	2400	3000
14	1	14	1014	6	1	1	1	5	6000	5000
15	1	15	1015	7	1	1	1	9	8000	6000
16	1	16	1016	8	6	1	1	5	2000	2000
17	1	17	1017	9	2	1	1	5	2000	3500
18	1	18	1018	11	5	.	1	5	6000	6000
19	1	19	1019	6	4	.	1	1	2400	2000
20	1	20	1020	4	2	.	1	1	3000	4000
21	1	21	1021	8	3	1	1	1	2000	3000
22	1	22	1022	4	2	.	1	1	3000	4000
23	1	23	1023	6	3	.	1	1	900	1200
24	1	24	1024	5	3	.	1	1	1500	2300
25	1	25	1025	6	4	.	1	1	800	1000
26	1	26	1026	6	4	.	1	1	900	1500
27	1	27	1027	5	3	1	1	1	1200	1500
28	1	28	1028	5	2	.	1	4	2500	3500
29	1	29	1029	5	3	.	1	4	2500	3500
30	1	30	1030	8	6	.	1	3	2200	3500
31	1	31	1031	5	3	.	1	1	3000	3000

☑ Name	Label
☑ psu	
☑ hh	
☑ hhid	
☑ fs	
☑ nc	
☑ dhs	
☑ shhh	
☑ eduhhh	
☑ minc	
☑ mexpd	

Properties

⊟ Variables	
Name	dhs
Label	
Type	byte
Format	%8.0g
Value label	
⊞ Notes	
⊟ Data	
⊞ Filename	
Label	
⊞ Notes	
Variables	10
Observations	400
Size	5.86K
Memory	64M
Sorted by	

Vars: 10 Order: Dataset Obs: 400 Filter: Off

Keep in mind that

- ✓ **Missing data should be blank in excel**
- ✓ **Missing data will be automatically dot (.) in Stata**
- ✓ **There should be no space in variable names**
- ✓ **Label variables after importing from excel**

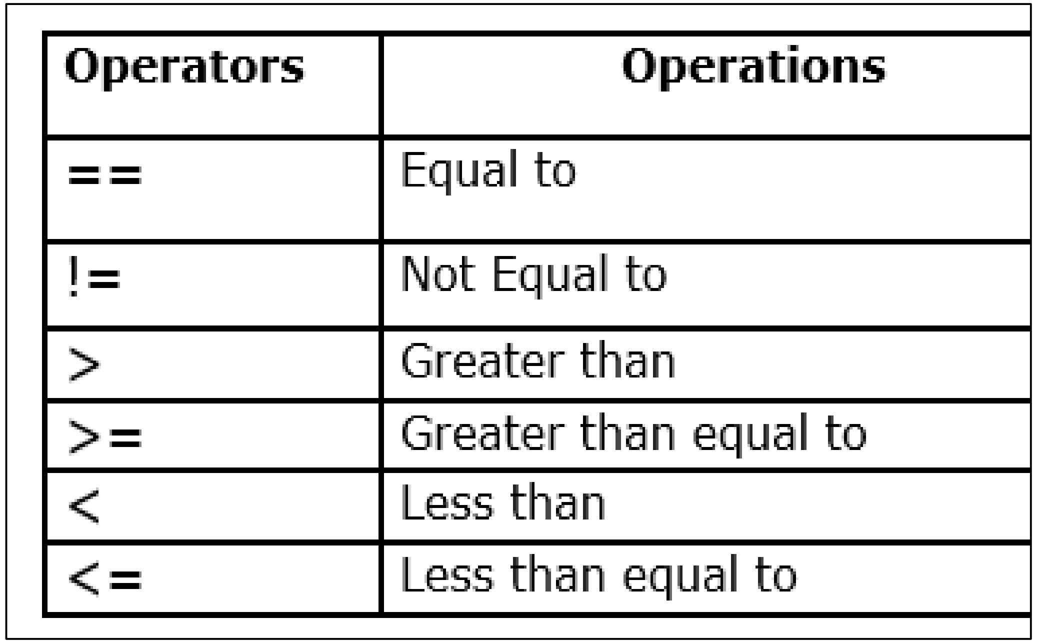
Examining data

- **browse: *br*** *variable_names*
- **edit: *ed*** *variable_names*
- **describe: *des*** *variable_names*
- **inspect: *ins*** *variable_names*
- **codebook: *codebook*** *variable_names*
- **List in result window: *list*** *variable_names*
- **Count observations: *count***

Using “in”

- ***br*** *variable_names* ***in 1/10***
- ***ins*** *variable_names* ***in f/5***
- ***codebook*** *variable_names* ***in -5/l***
- ***list*** *variable_names* ***in 11/20***

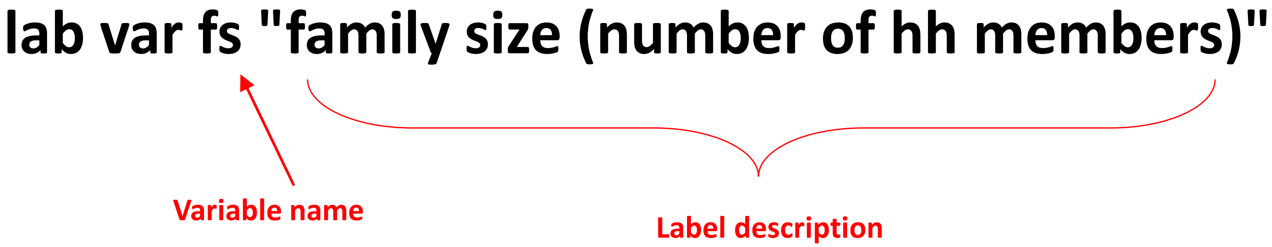
Relational operators



Operators	Operations
==	Equal to
!=	Not Equal to
>	Greater than
>=	Greater than equal to
<	Less than
<=	Less than equal to

- ***br variable_names if dhs==.***
- ***br variable_names if dhs!=.***
- ***count if minc<5000***
- ***count if mexpd>=5000***

Labelling variables

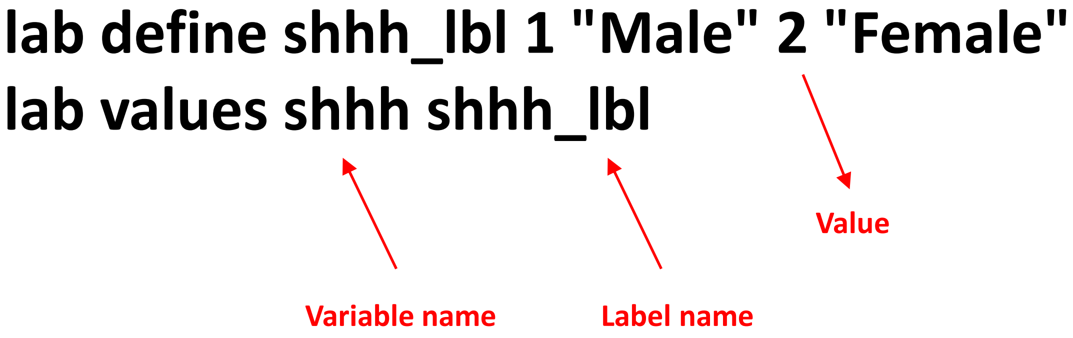


```
lab var fs "family size (number of hh members)"
```

Variable name

Label description

Labelling values



```
lab define shhh_lbl 1 "Male" 2 "Female"
lab values shhh shhh_lbl
```

Value

Variable name

Label name

rename variable, replace values

Command: *rename old_name new_name*
Example: *rename shhh gender*

Get average minc: *tabstat minc*
Replace missing value: *replace minc=2704.125 if minc==.*
Undo manually: *replace minc=. if minc==2704.125*

keep / drop variables or observations

Keeping variables: *keep psu hh hhid fs nc gender eduhhh minc mexpd*
Or, dropping variables: *drop dhs*
Keeping observations: *keep if minc!=.*
Or, dropping observations: *drop if minc==.*

Generating variable

….using arithmetic operator

- ***gen msav=minc-mexpd***
- ***gen msavsq=msav^2***
- ***gen pc_minc=minc/fs***
- ***gen msav_rate=msav/minc if minc!=0***

- ***gen fdummy=.***
- ***replace fdummy=1 if gender==2***
- ***replace fdummy=0 if gender==1***

can use “recode” instead

Symbol	Operator
+	Addition
-	Subtraction
*	Multiplication
/	Division
^	Raise to the power

Generating variable

….using function

Function for gen command	Use
log(x) or ln(x)	(Natural) logarithm of x
abs(x)	Absolute value of x
exp(x)	Antilog of x
int(x)	Truncation to integer value
round(x)	Rounds to the nearest integer
round(x,y)	x rounded in y decimal place
sqrt(x)	Square root of x

- ***gen lminc=log(minc)***
- ***gen amsav=abs(msav)***

"egen"= extended generation

- ***egen avgminc=mean(minc)***
- ***gen gavgminc=mean(minc), by(gender)***

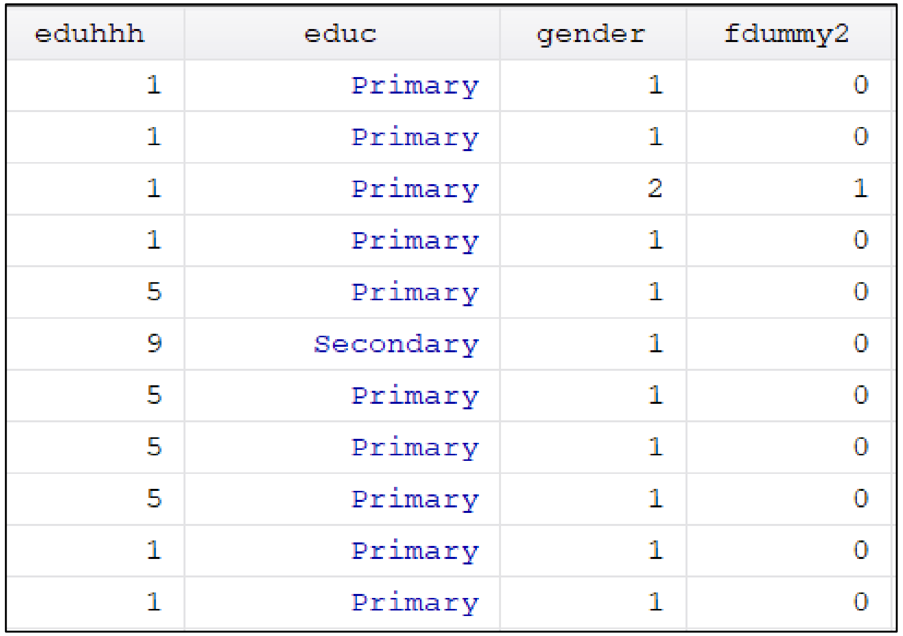
………………………………………………………………………………..………………………………more details later with household survey data

Recode variable

....to categorize

recode eduhhh (min/5=1 “Primary”) (6/10=2 “Secondary”) (11/12=3 “Highersecondary”) (13/max=4 “Graduation”), gen(educ)

recode gender (1=0) (2=1), gen(fdummy2)



eduhhh	educ	gender	fdummy2
1	Primary	1	0
1	Primary	1	0
1	Primary	2	1
1	Primary	1	0
5	Primary	1	0
9	Secondary	1	0
5	Primary	1	0
5	Primary	1	0
5	Primary	1	0
1	Primary	1	0
1	Primary	1	0

Logical operators

….after “if”

Operator	Symbol
Or	|
And	&

```
br if gender==1 & minc >5000
```

```
gen dummy=0
replace dummy=1 if minc>5000 | mexpd>5000
```

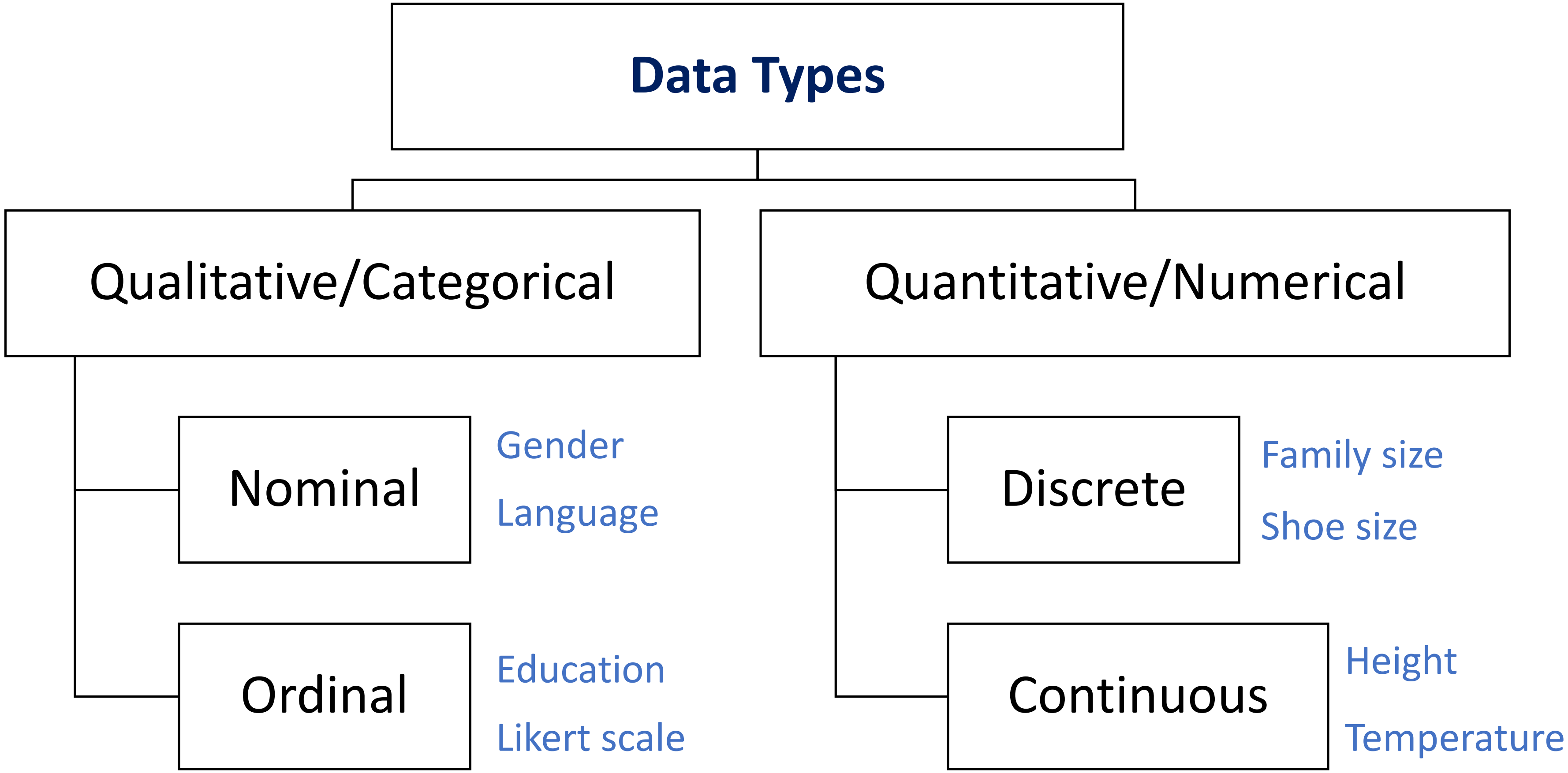
Order & Sort

```
order psu hh hhid minc mexpd
```

order variables at the beginning of the dataset

```
sort gender fs
sort psu hh
```

sort observations by values



Data Types

Qualitative/Categorical

Nominal

Gender

Language

Ordinal

Education

Likert scale

Quantitative/Numerical

Discrete

Family size

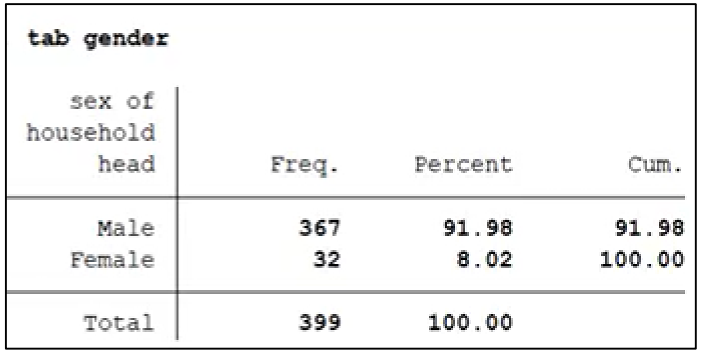
Shoe size

Continuous

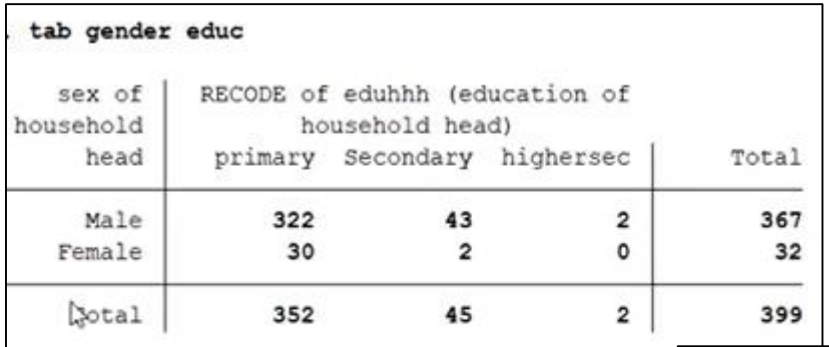
Height

Temperature

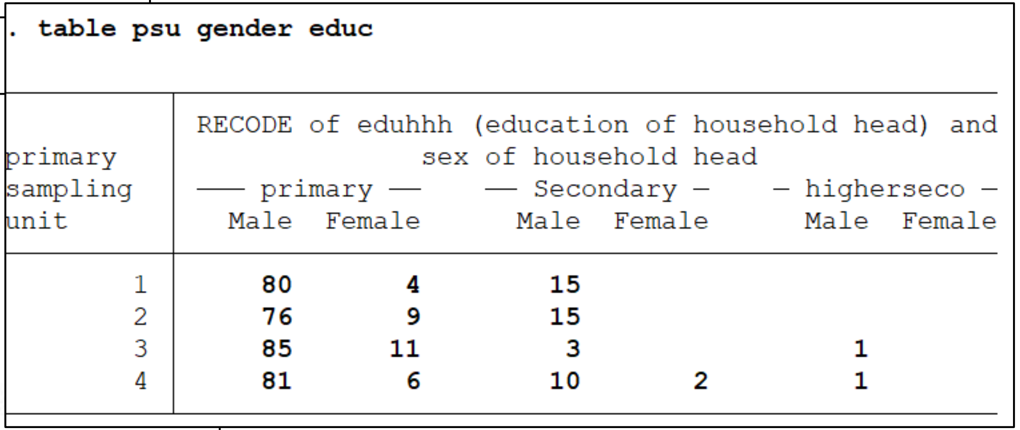
Descriptive statistics: categorical



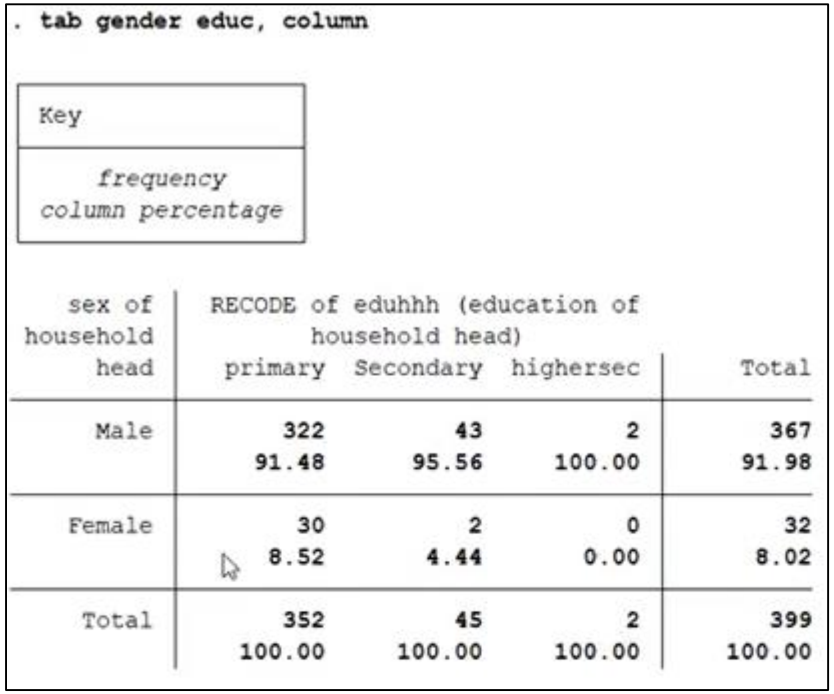
```
tab gender

     sex of |
  household |
       head |      Freq.     Percent        Cum.
------------+-----------------------------------
       Male |        367       91.98       91.98
     Female |         32        8.02      100.00
------------+-----------------------------------
      Total |        399      100.00
```



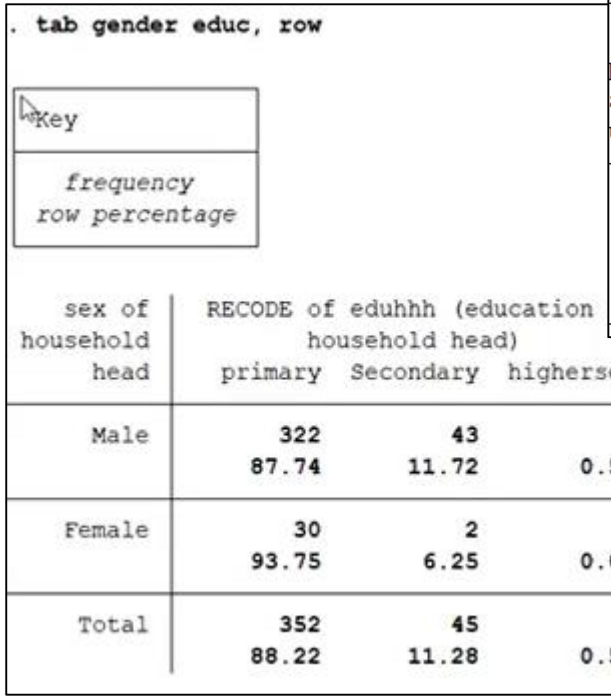
```
. tab gender educ

     sex of | RECODE of eduhhh (education of
  household |        household head)
       head |   primary  Secondary  highersec |     Total
------------+---------------------------------+----------
       Male |       322         43          2 |       367
     Female |        30          2          0 |        32
------------+---------------------------------+----------
      Total |       352         45          2 |       399
```



```
. table psu gender educ

-----------------------------------------------------------------------
primary   | RECODE of eduhhh (education of household head) and
sampling  |                sex of household head
unit      | -- primary --     -- Secondary -     - higherseco -
          |   Male  Female        Male  Female        Male  Female
----------+------------------------------------------------------------
        1 |     80       4          15
        2 |     76       9          15
        3 |     85      11           3                     1
        4 |     81       6          10       2             1
-----------------------------------------------------------------------
```

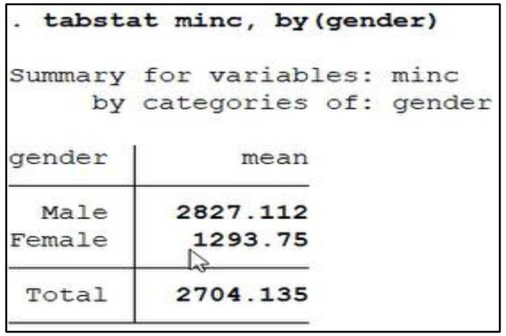


```
. tab gender educ, column

+-------------------+
| Key               |
|-------------------|
|     frequency     |
| column percentage |
+-------------------+

     sex of | RECODE of eduhhh (education of
  household |        household head)
       head |   primary  Secondary  highersec |     Total
------------+---------------------------------+----------
       Male |       322         43          2 |       367
            |     91.48      95.56     100.00 |     91.98
------------+---------------------------------+----------
     Female |        30          2          0 |        32
            |      8.52       4.44       0.00 |      8.02
------------+---------------------------------+----------
      Total |       352         45          2 |       399
            |    100.00     100.00     100.00 |    100.00
```

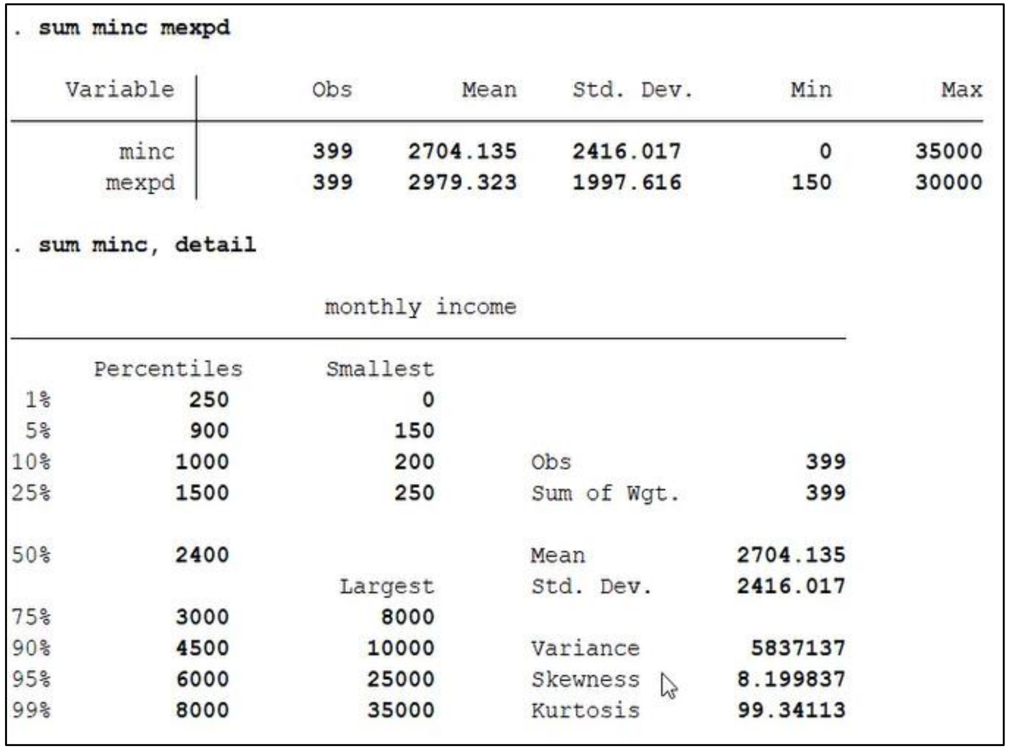


```
. tab gender educ, row

+----------------+
| Key            |
|----------------|
|   frequency    |
| row percentage |
+----------------+

     sex of | RECODE of eduhhh (education
  household |        household head)
       head |   primary  Secondary  highersec |     Total
------------+---------------------------------+----------
       Male |       322         43          2 |       367
            |     87.74      11.72       0.54 |    100.00
------------+---------------------------------+----------
     Female |        30          2          0 |        32
            |     93.75       6.25       0.00 |    100.00
------------+---------------------------------+----------
      Total |       352         45          2 |       399
            |     88.22      11.28       0.50 |    100.00
```

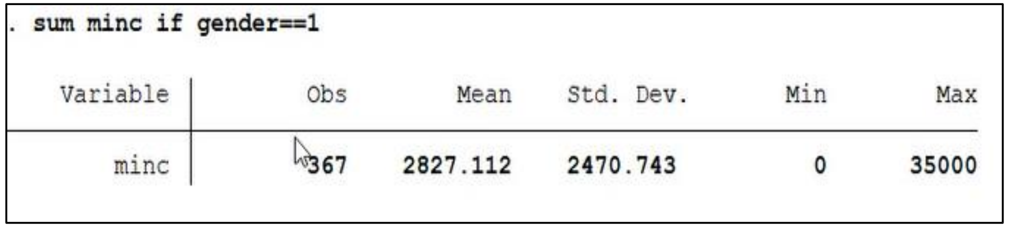
Descriptive statistics: numerical



```
. tabstat minc, by(gender)

Summary for variables: minc
     by categories of: gender

gender   |      mean
---------+----------
    Male |  2827.112
  Female |   1293.75
---------+----------
   Total |  2704.135
--------------------
```

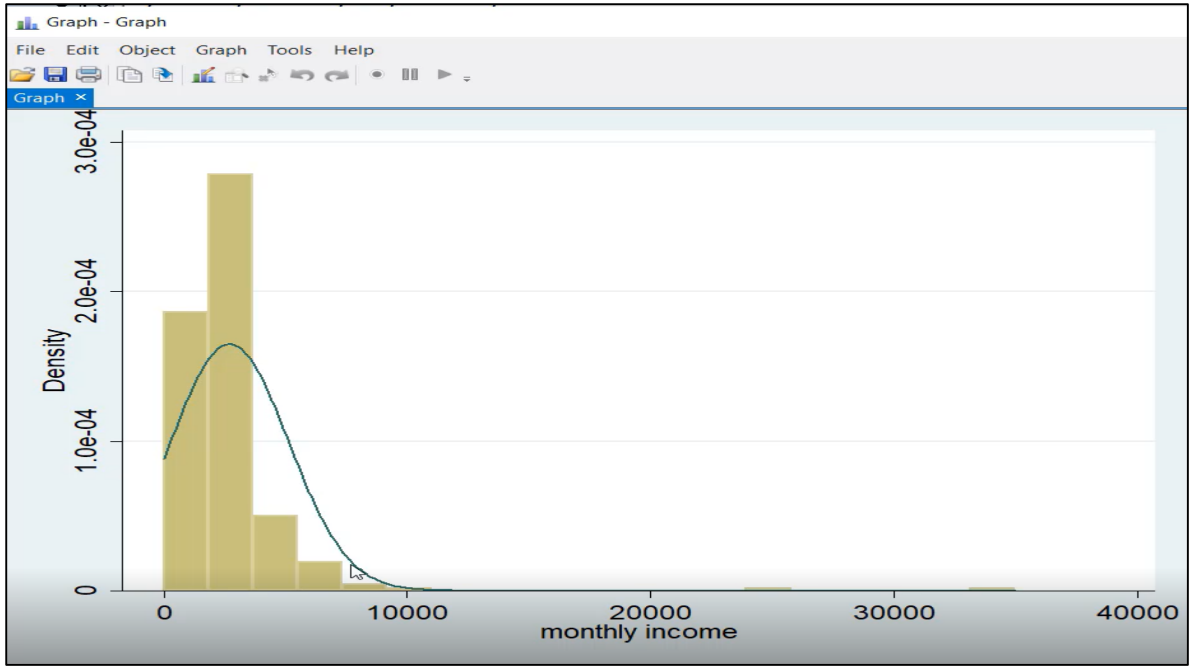


```
. sum minc mexpd

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        minc |        399    2704.135    2416.017          0      35000
       mexpd |        399    2979.323    1997.616        150      30000

. sum minc, detail

                       monthly income
-------------------------------------------------------------
      Percentiles      Smallest
 1%          250              0
 5%          900            150
10%         1000            200       Obs                 399
25%         1500            250       Sum of Wgt.         399

50%         2400                      Mean           2704.135
                        Largest       Std. Dev.      2416.017
75%         3000           8000
90%         4500          10000       Variance        5837137
95%         6000          25000       Skewness       8.199837
99%         8000          35000       Kurtosis       99.34113
```

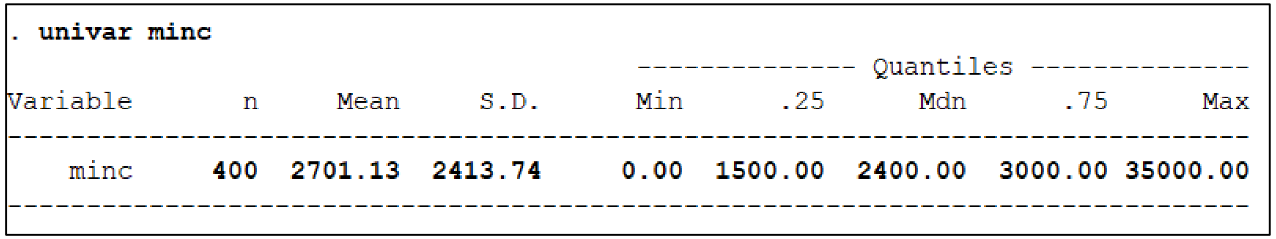


```
. sum minc if gender==1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        minc |        367    2827.112    2470.743          0      35000
```

hist minc, norm

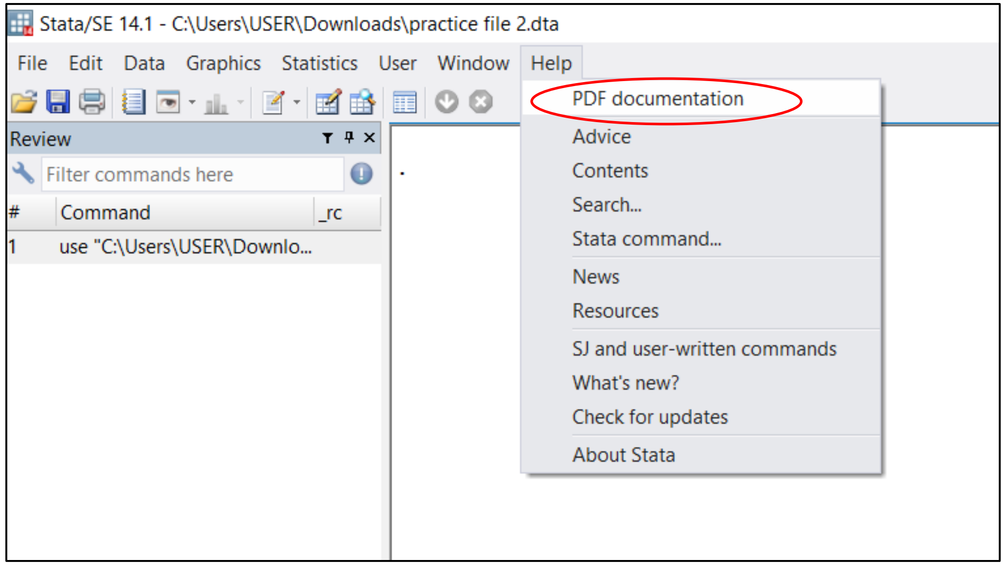


ssc install univar



```
. univar minc
                                        -------------- Quantiles --------------
Variable       n     Mean     S.D.      Min      .25      Mdn      .75      Max
-------------------------------------------------------------------------------
    minc     400  2701.13  2413.74     0.00  1500.00  2400.00  3000.00 35000.00
-------------------------------------------------------------------------------
```

Get help, find/install packages, and more..

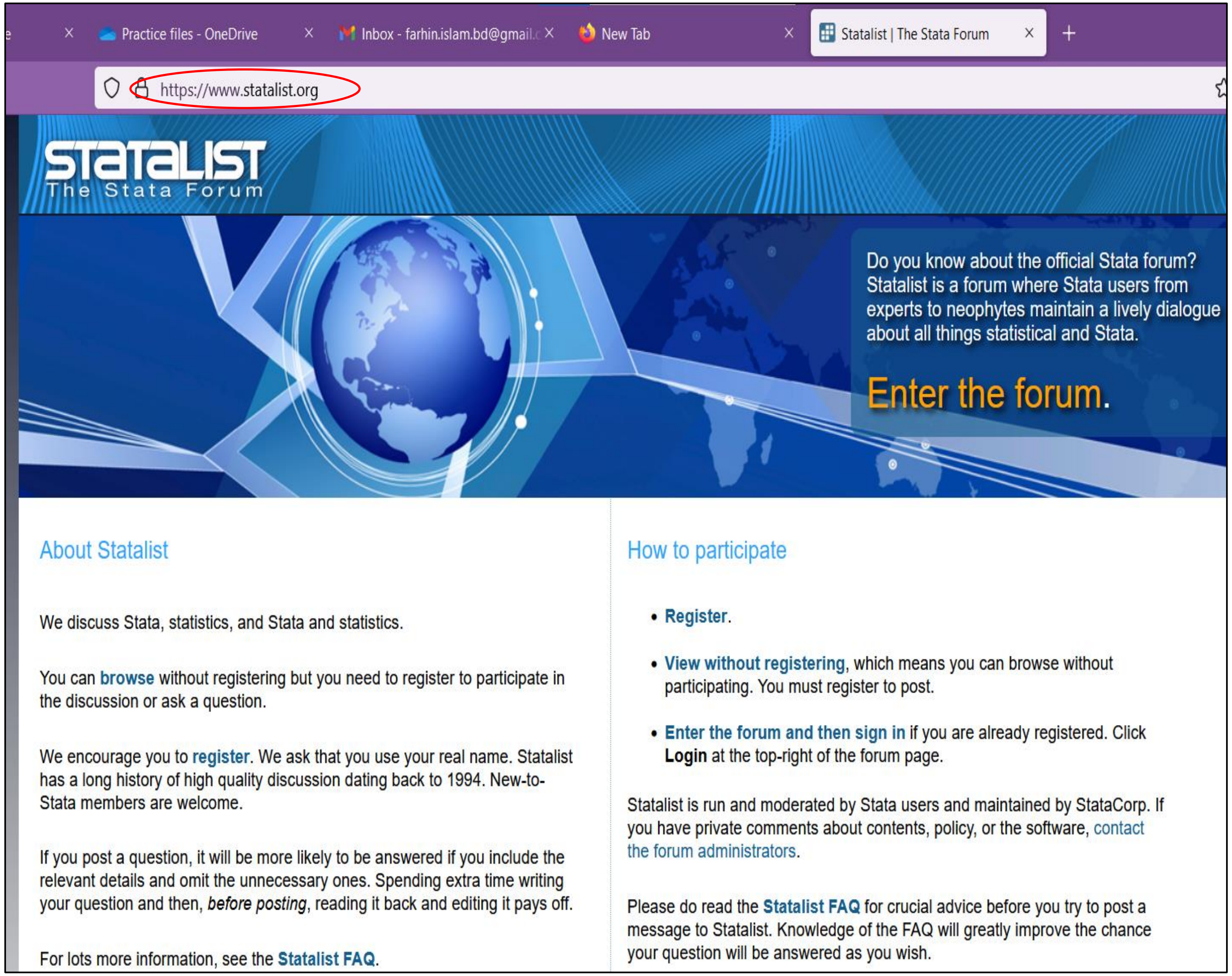


```
use "C:\Users\USER\Downlo...
```

- *ssc install extremes*
- *extremes minc*
- *ssc uninstall extremes*

- *ssc hot*

- *ssc install asdoc*
- *findit asdoc*
- *help asdoc*
- *help reg*


Practice files - OneDrive
Inbox - farhin.islam.bd@gmail.c
New Tab
Statalist | The Stata Forum
https://www.statalist.org
STATALIST
The Stata Forum

Do you know about the official Stata forum? Statalist is a forum where Stata users from experts to neophytes maintain a lively dialogue about all things statistical and Stata.

Enter the forum.

About Statalist

We discuss Stata, statistics, and Stata and statistics.

You can browse without registering but you need to register to participate in the discussion or ask a question.

We encourage you to register. We ask that you use your real name. Statalist has a long history of high quality discussion dating back to 1994. New-to-Stata members are welcome.

If you post a question, it will be more likely to be answered if you include the relevant details and omit the unnecessary ones. Spending extra time writing your question and then, *before posting*, reading it back and editing it pays off.

For lots more information, see the Statalist FAQ.

How to participate

- Register.
- View without registering, which means you can browse without participating. You must register to post.
- Enter the forum and then sign in if you are already registered. Click **Login** at the top-right of the forum page.

Statalist is run and moderated by Stata users and maintained by StataCorp. If you have private comments about contents, policy, or the software, contact the forum administrators.

Please do read the Statalist FAQ for crucial advice before you try to post a message to Statalist. Knowledge of the FAQ will greatly improve the chance your question will be answered as you wish.

Working directory

Before doing anything with an open Stata file, you should always define the **location of your working folder**. Your ‘working folder’ is a folder on (one of) your drives where you keep all of your data for a particular project or piece of analysis.

If you have never been so organized with your data sets, now is an excellent time to start! Doing a bunch of analyses using multiple datasets and having the files all jumbled in with other documents or spread all over the place in separate folders is not a correct way to use Stata!

Check current working directory with the command:

```
   pwd
```

Change working directory:

Go to File ⇒ Change Working Directory ⇒ Select specific folder.
Or, use following command:

```
cd "…folder path…\filename"
```

Do file

What is Do File?

Do files are simply text files whose names end with .do and which contain Stata commands exactly the way you'd type them into the command window. Sometimes people call them programs, though Stata uses this term for something else. You can write do files using any text editor, but the Do File Editor built into Stata has tools and features designed to help programmers so we recommend using it. Do not write Stata code using Word—it will automatically insert things like "smart quotes" and other formatting that Stata cannot understand. **You should save all your commands (total workflow) in a do file**.

Starting a Do file:

Start the Do File Editor by clicking on the button that looks like a pencil writing in a notebook or by typing doedit in the command line. Alternatively use Window ⇒ Do-file editor.

Use File ⇒ Open, from within the editor and look for the desired file. Alternatively, in the Review window you should see a copy of the command that was generated while executing commands. Copy this command and paste it into the do file. Save this file using File ⇒ Save, from within the do file editor. Only through this way, **your commands will not disappear after quitting Stata**.

Do file

Setting Up:

Almost every do file should start with the following commands (or something very much like them):

```
clear all
set more off
```

The first command clears the memory so you don't have to worry about what might have happened before your do file was run. The second tells Stata not to pause whenever the screen fills and wait for you to press a key.

Loading Data:

Next you will usually load a data set:

```
use dataset_name.dta
```

If the dataset is in the current working directory, you don't need to specify its location.

Do Your Work:

At this point you'll be ready to do your work. Generally, this means data preparation, exploratory analysis, or analysis you intend to report or publish. We recommend you to use do file for this. **Write the commands in the do file serially.**

Do file

Save your new dataset:

You may need to save your work (new data file after all the modifications) at the end:

```
save new_dataset_name.dta, replace
```

The replace option again allows Stata to overwrite the output from previous attempts to run the do file. Otherwise, the do file will not run 2nd time because there is already a dataset with that name.

Never, ever save your output data set over your input data set. (In other words, the starting use command and the ending save command should never act on the same file.) If you do, the data set your do file was written to work with will no longer exist. If it turns out you made a mistake, you can't easily recover.

Clearing everything from memory, loading the data set you want to use, and then saving any changes you make to a different file (with different name) makes your do file **reproducible**. You can run it again any time you want and get the exact same results. If the input data set changes, you'll be applying the exact same procedures to the new data. **If it turns out you made a mistake, all you need to do is correct the error in your code and run the do file again. If you need to make changes you can do so without starting over**. It may take a bit of effort at first to get into the habit of writing reproducible code, but the effort will pay off very quickly.

Do file

Comments in Do file:

Comments are text included in a do file for the **benefit of human readers, not for Stata**. Comments can explain what the do file does and why, and if anyone else ever needs to read and understand your do file they'll be very grateful for good comments. Or *you* are the most beneficiary of your comments, when you have to figure out how your do file works months or years after writing it. Be sure to comment any code that required particular cleverness to write.

Comments need to be marked as such so that Stata will not try to execute them. /* means Stata should ignore everything until it sees */, while // means Stata should ignore the rest of that line. Here's how one might comment:

```
//make a list of cars I might be interested in buying
list make price mpg rep78 if price<4000 | (price<5000 & rep78>3 & rep78<.)
/* Note:
Some cars will appear on the list even though they have a missing value
for rep78.
This is not an error.
If their price is less than $4,000 I don't care about their
repair record.
*/
```

Do file

Save Do file:
File ⇒ Save as (In do file window)

Running a Do File:

The easiest way to run a do file is to press **Ctrl+d** in the Do File Editor, or click the icon on the far right that looks like a "play" button over some code. **If you first select just part of the do file then only that part will be run.**

Running parts of your code rather than the entire do file can save a lot of time, but code taken out of context won't always work. For example, if you run a command that creates a variable x, realize you made a mistake, and then fix it, you can't simply select that command and run it again unless you first drop the existing version of x. If you find yourself getting confused by these kinds of issues, run the entire do file rather than a selection so everything is run in its proper context with appropriate serial.

Log file

What is Log File?

It is used to keep a record of the results obtained while using Stata. A research do file should have a corresponding log file which records all the commands the do file ran along with their results.

Start a Log:

To start logging, the command is:

```
log using logfilename.log, replace
```

where logfilename is the name of the file you want Stata to use as a log. **Give the log file the same name as the do file it records**, so it's obvious which log file goes with which do file. The **replace** option will overwrite the previous version. Alternatively, an append option will add to the previous contents of the log file.

Then do your work using do file.

Close your Log:

The command is:

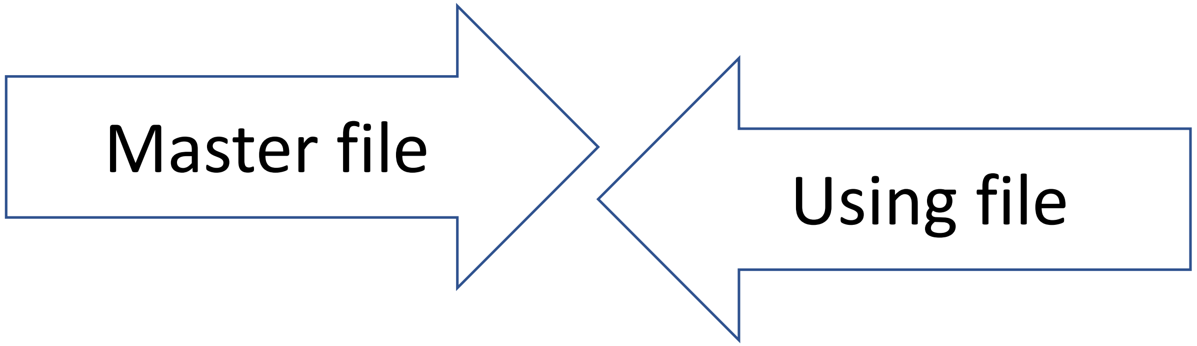
```
log close
```

Or, go to File ⇒ Log ⇒ Close

Merge & Append

……Joining two datasets together

Merge	Append
Adds variables/columns	Adds observations/rows



Master file

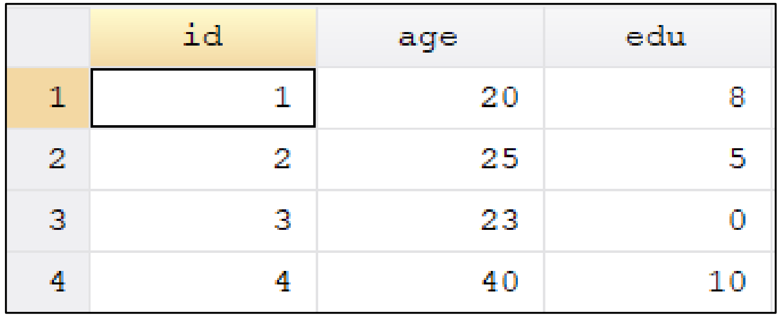
Using file

Append

- Append adds cases/observations to a dataset by joining two datasets together.
- In Append, we join two datasets:
  - **Master file:** the data file with which we will append the other file
  - **Using file:** the data file we will be appending with the master file.
- To append a using file with a master file, they must have:
  - Common variables with same name in two datasets for all observations.
  - Each variable must be in the same format in both the master file and the using file.
  - If you attempt to append a using dataset with variables that do not match with the master dataset, they will be added to the appended dataset as additional variables.

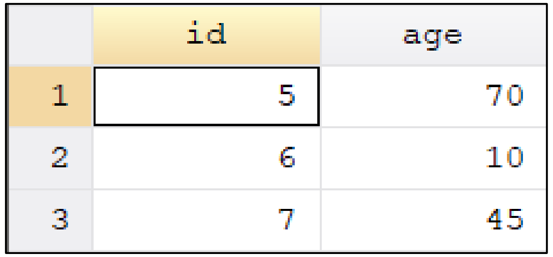
Append

Master file



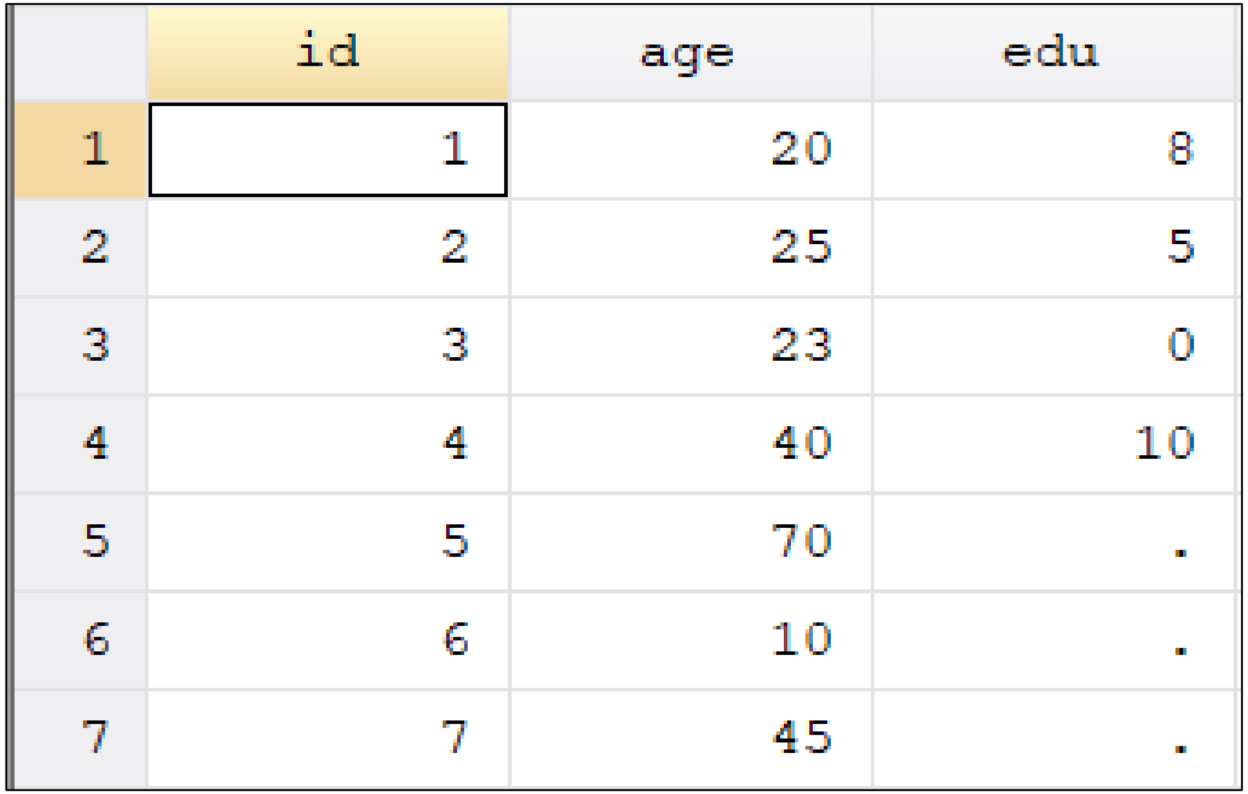
	id	age	edu
1	1	20	8
2	2	25	5
3	3	23	0
4	4	40	10

Using file



	id	age
1	5	70
2	6	10
3	7	45

Appended Dataset



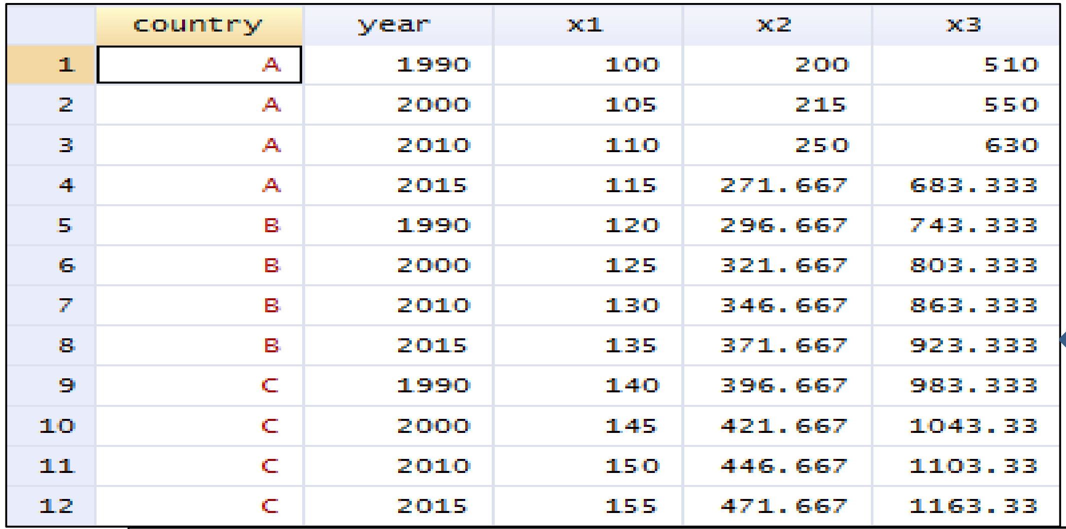
	id	age	edu
1	1	20	8
2	2	25	5
3	3	23	0
4	4	40	10
5	5	70	.
6	6	10	.
7	7	45	.

Merge

- Merge adds variables to a dataset by joining two datasets together.
- In merge, we join two datasets:
  - **Master file:** the data file with which we will merge the other file
  - **Using file:** the data file we will be merging with the master file.
- To merge a using file with a master file, they must have:
  - At least a common variable (key variable) based on which we will merge.
  - There can be more than one key variables
  - The key variables must have the same name.
  - The variable must be in the same format in both the master file and the using file.
  - If they are in *string (alphanumeric)* format in the both the files, their spelling must be same (i.e. country names, etc).
- Merge could be of four types:
  - 1:1 (one to one) merge, m:1 (many to one) merge, 1:m (one to many) merge and m:m (many to many) merge.

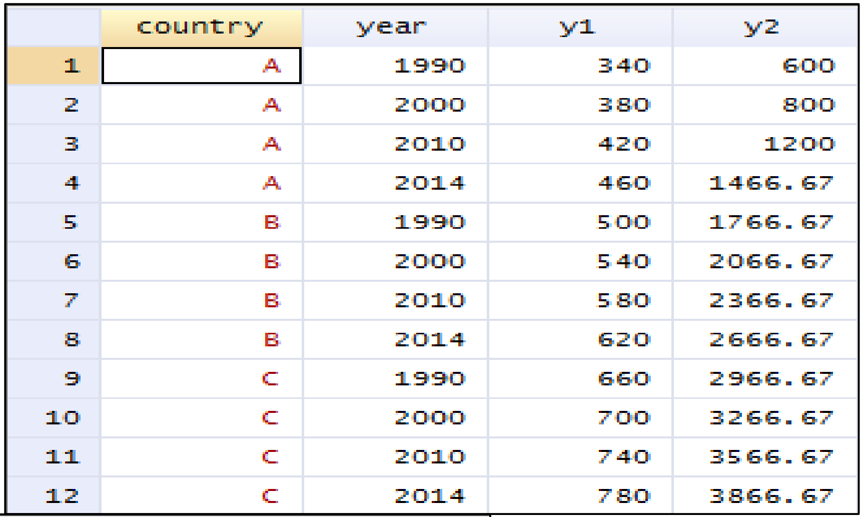
Merge (1:1)

Master file

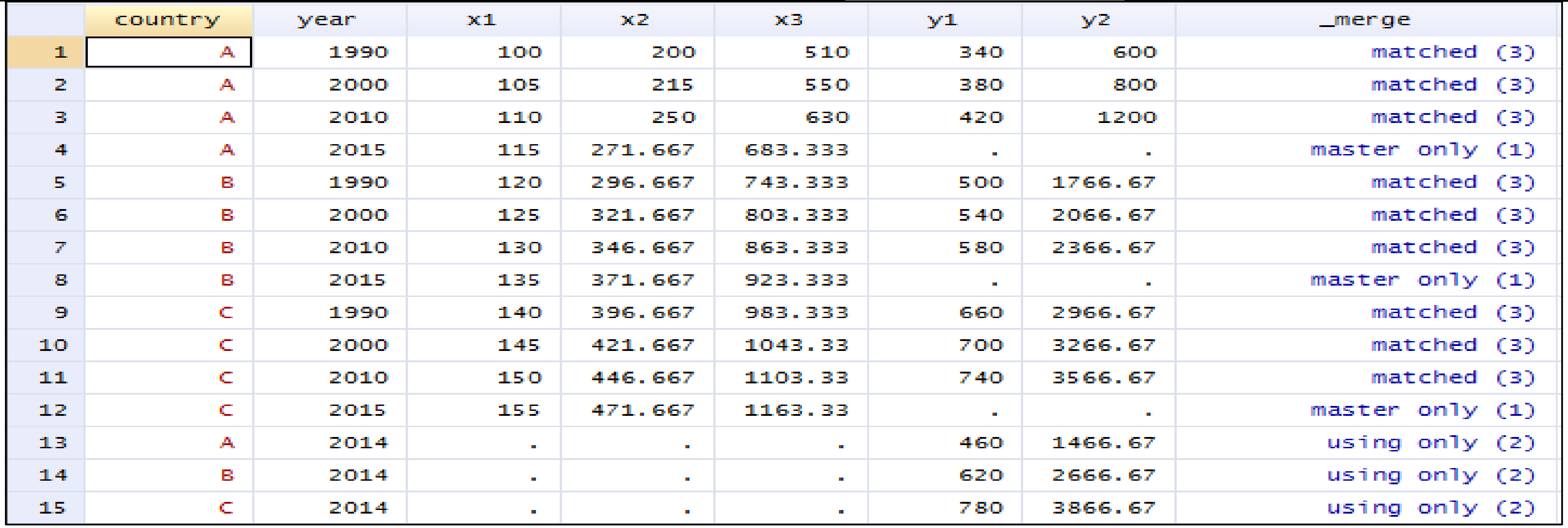


	country	year	x1	x2	x3
1	A	1990	100	200	510
2	A	2000	105	215	550
3	A	2010	110	250	630
4	A	2015	115	271.667	683.333
5	B	1990	120	296.667	743.333
6	B	2000	125	321.667	803.333
7	B	2010	130	346.667	863.333
8	B	2015	135	371.667	923.333
9	C	1990	140	396.667	983.333
10	C	2000	145	421.667	1043.33
11	C	2010	150	446.667	1103.33
12	C	2015	155	471.667	1163.33

Using file



	country	year	y1	y2
1	A	1990	340	600
2	A	2000	380	800
3	A	2010	420	1200
4	A	2014	460	1466.67
5	B	1990	500	1766.67
6	B	2000	540	2066.67
7	B	2010	580	2366.67
8	B	2014	620	2666.67
9	C	1990	660	2966.67
10	C	2000	700	3266.67
11	C	2010	740	3566.67
12	C	2014	780	3866.67

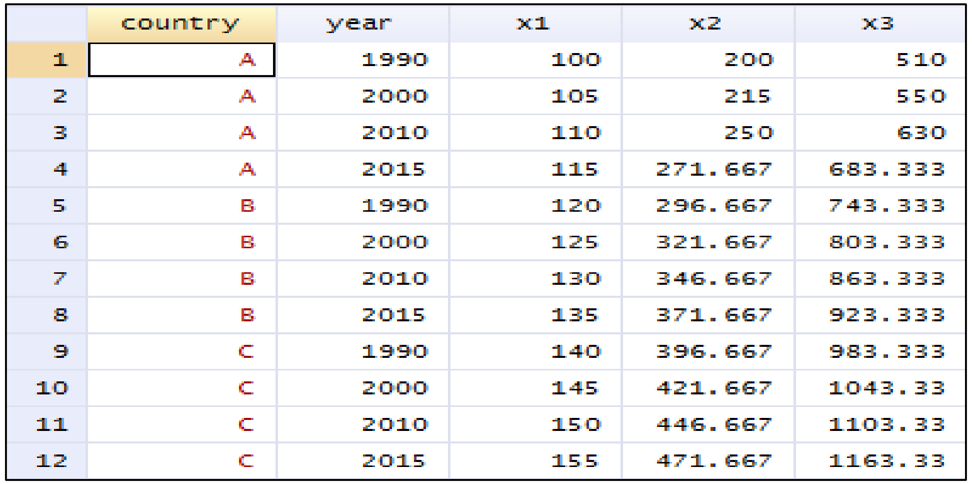


	country	year	x1	x2	x3	y1	y2	_merge
1	A	1990	100	200	510	340	600	matched (3)
2	A	2000	105	215	550	380	800	matched (3)
3	A	2010	110	250	630	420	1200	matched (3)
4	A	2015	115	271.667	683.333	.	.	master only (1)
5	B	1990	120	296.667	743.333	500	1766.67	matched (3)
6	B	2000	125	321.667	803.333	540	2066.67	matched (3)
7	B	2010	130	346.667	863.333	580	2366.67	matched (3)
8	B	2015	135	371.667	923.333	.	.	master only (1)
9	C	1990	140	396.667	983.333	660	2966.67	matched (3)
10	C	2000	145	421.667	1043.33	700	3266.67	matched (3)
11	C	2010	150	446.667	1103.33	740	3566.67	matched (3)
12	C	2015	155	471.667	1163.33	.	.	master only (1)
13	A	2014	.	.	.	460	1466.67	using only (2)
14	B	2014	.	.	.	620	2666.67	using only (2)
15	C	2014	.	.	.	780	3866.67	using only (2)

Merged dataset
(One to One)
1:1

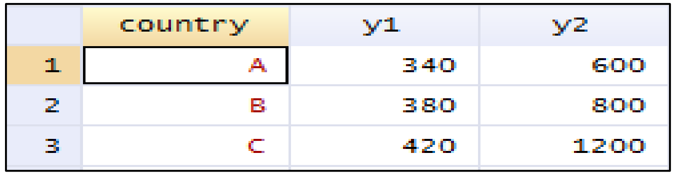
Merge (1:m)

Using file



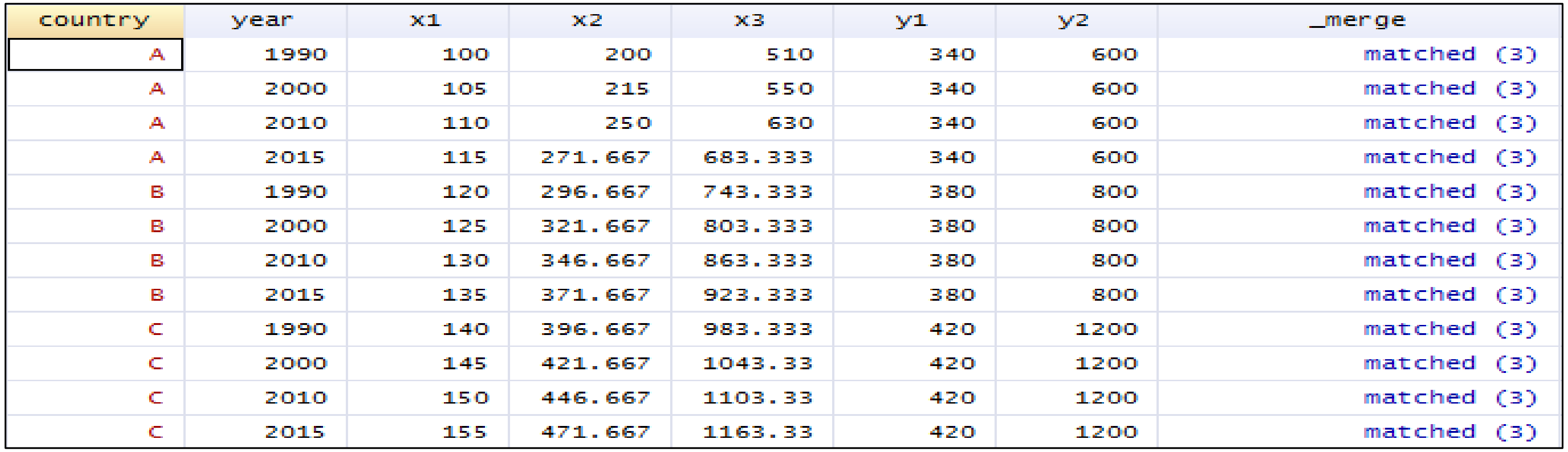
	country	year	x1	x2	x3
1	A	1990	100	200	510
2	A	2000	105	215	550
3	A	2010	110	250	630
4	A	2015	115	271.667	683.333
5	B	1990	120	296.667	743.333
6	B	2000	125	321.667	803.333
7	B	2010	130	346.667	863.333
8	B	2015	135	371.667	923.333
9	C	1990	140	396.667	983.333
10	C	2000	145	421.667	1043.33
11	C	2010	150	446.667	1103.33
12	C	2015	155	471.667	1163.33

Master file



	country	y1	y2
1	A	340	600
2	B	380	800
3	C	420	1200

Merged dataset
(One to Many)
1:m



country	year	x1	x2	x3	y1	y2	_merge
A	1990	100	200	510	340	600	matched (3)
A	2000	105	215	550	340	600	matched (3)
A	2010	110	250	630	340	600	matched (3)
A	2015	115	271.667	683.333	340	600	matched (3)
B	1990	120	296.667	743.333	380	800	matched (3)
B	2000	125	321.667	803.333	380	800	matched (3)
B	2010	130	346.667	863.333	380	800	matched (3)
B	2015	135	371.667	923.333	380	800	matched (3)
C	1990	140	396.667	983.333	420	1200	matched (3)
C	2000	145	421.667	1043.33	420	1200	matched (3)
C	2010	150	446.667	1103.33	420	1200	matched (3)
C	2015	155	471.667	1163.33	420	1200	matched (3)

Thank you!