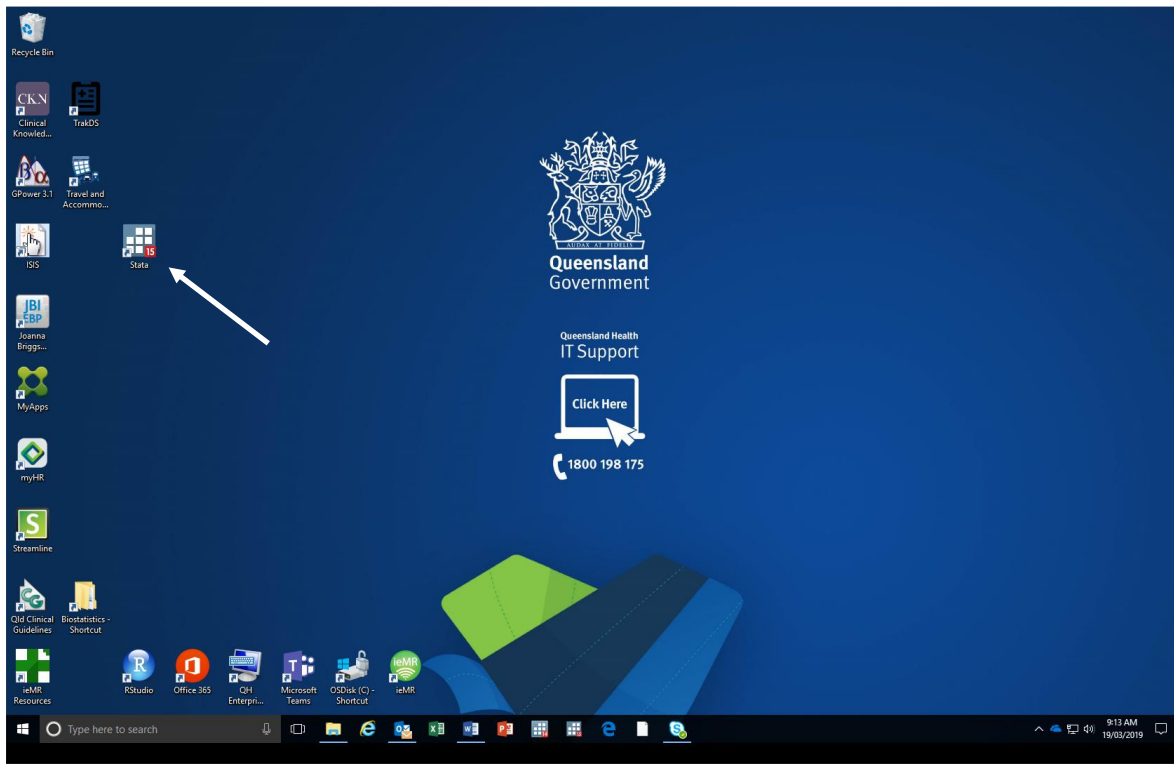
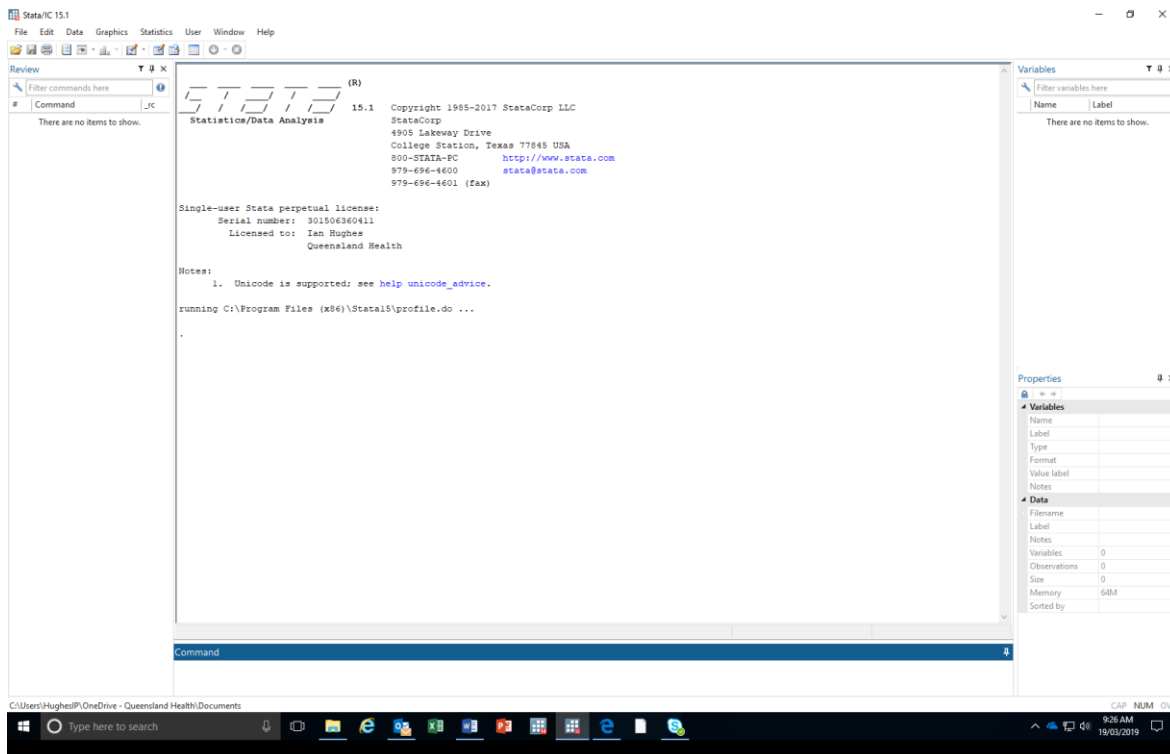


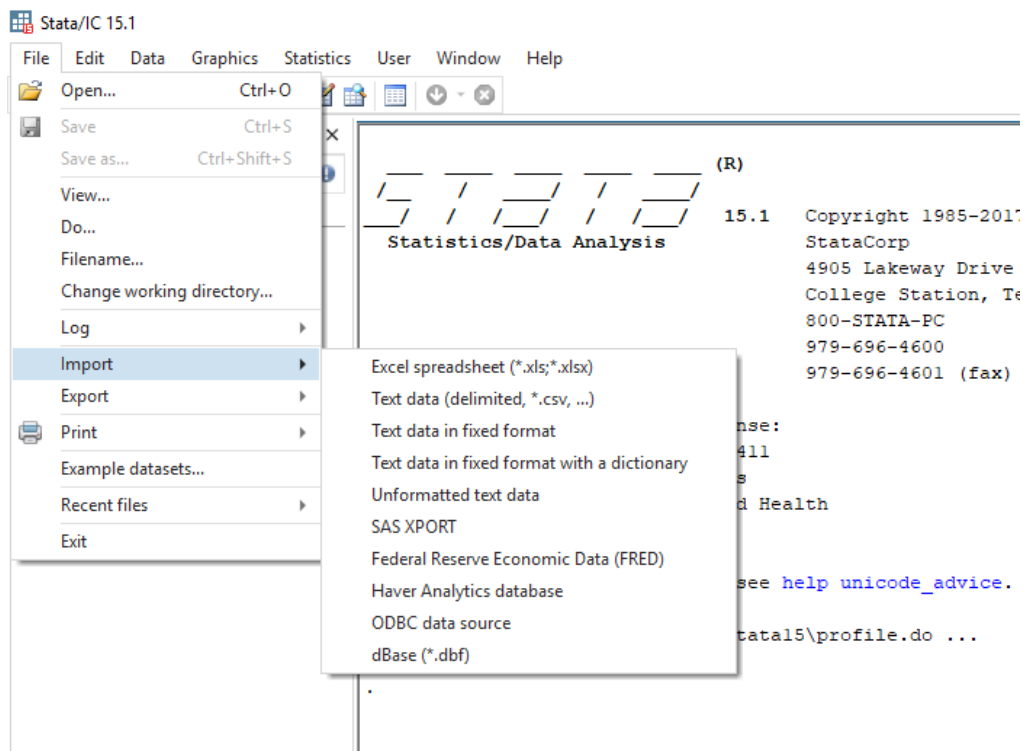
Introduction to Using Stata15 at Gold Coast Health

The Stata 15 icon should be located in the start menu and can be moved to the desk top. Double click on the icon to open Stata.

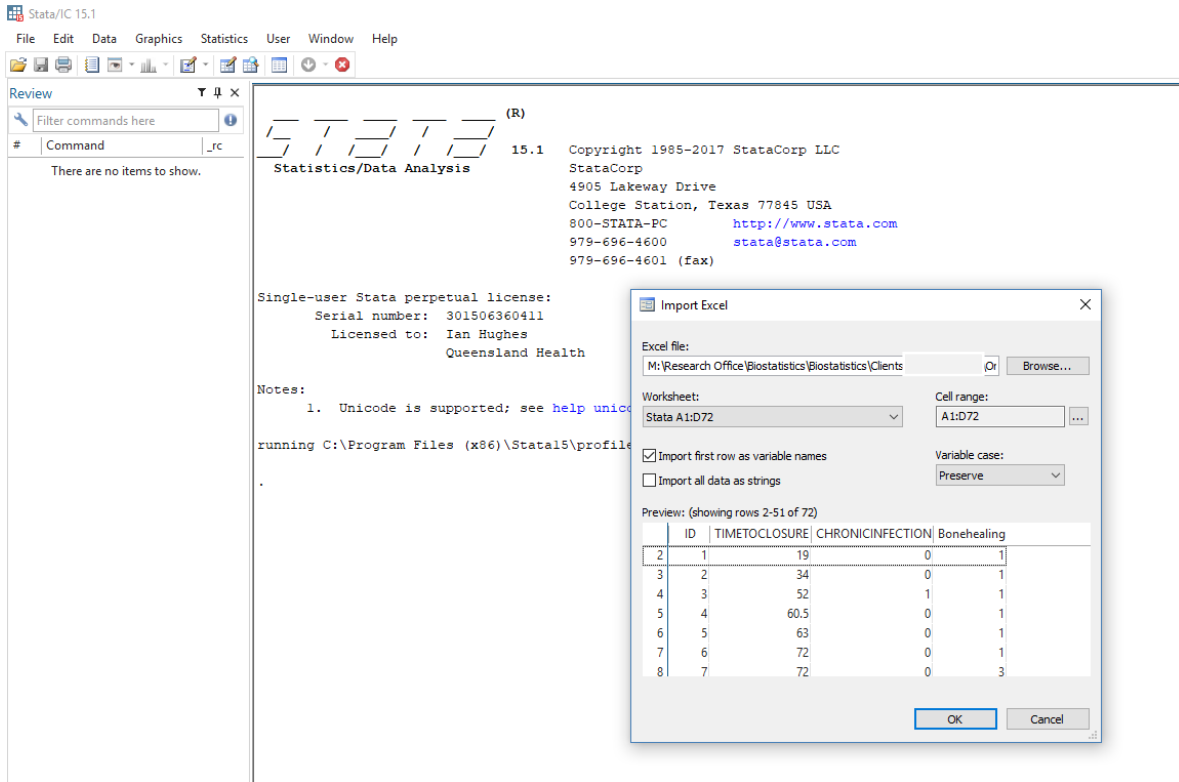




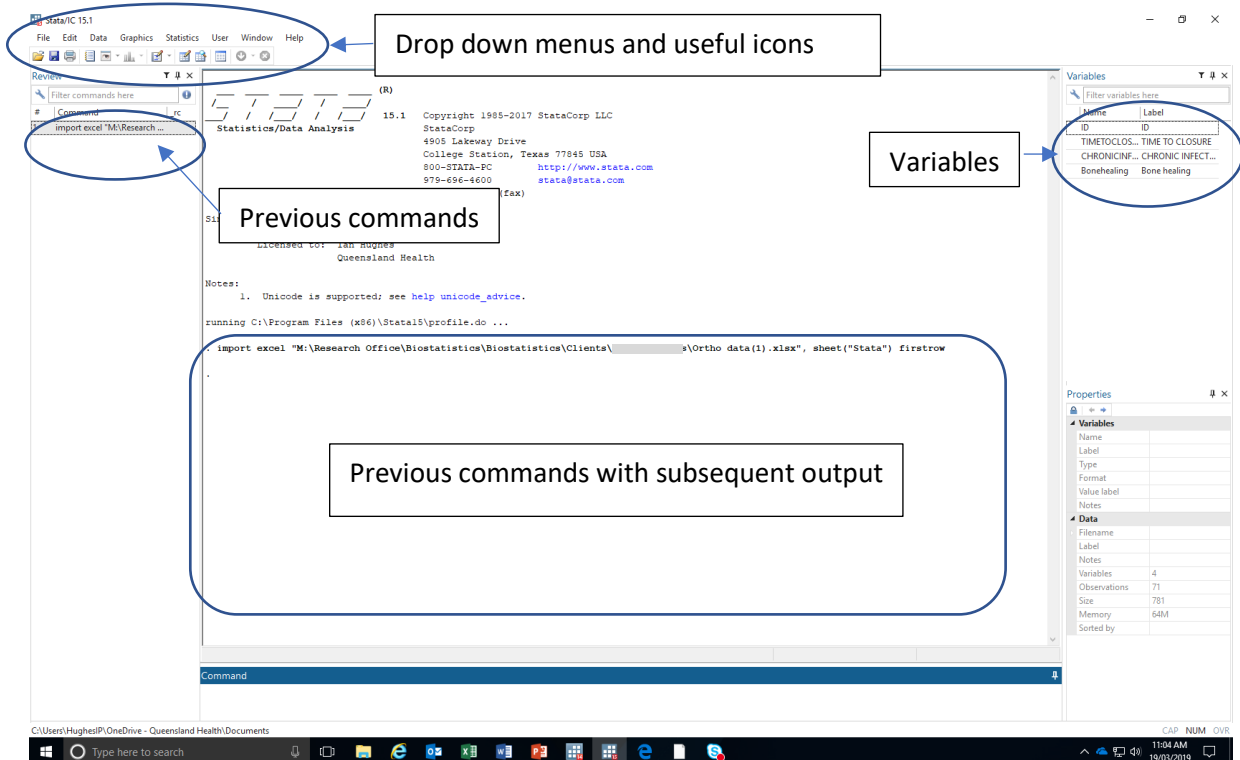
The Stata screen should appear as above. To import data from an Excel file click the File tab, Import, Excel spreadsheet as shown below.



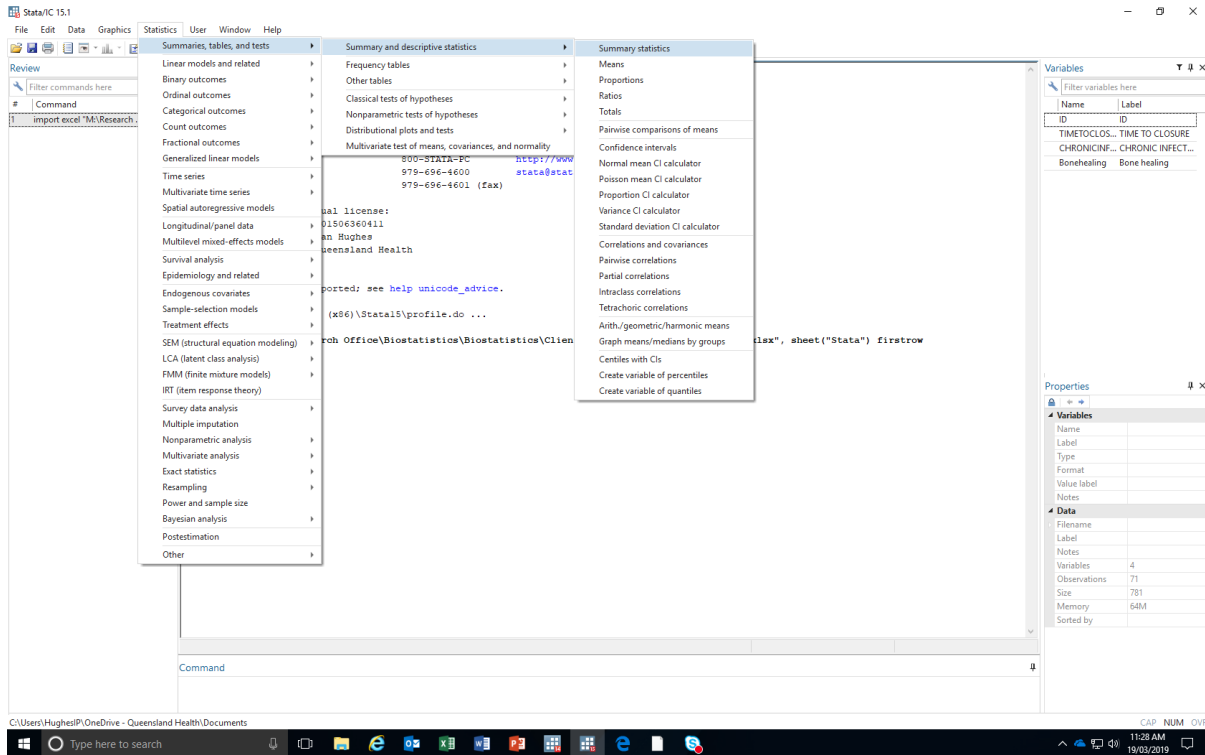
Browse to the file you want to import. Pick the worksheet you want, tick "Import first row as variable names", click ok.



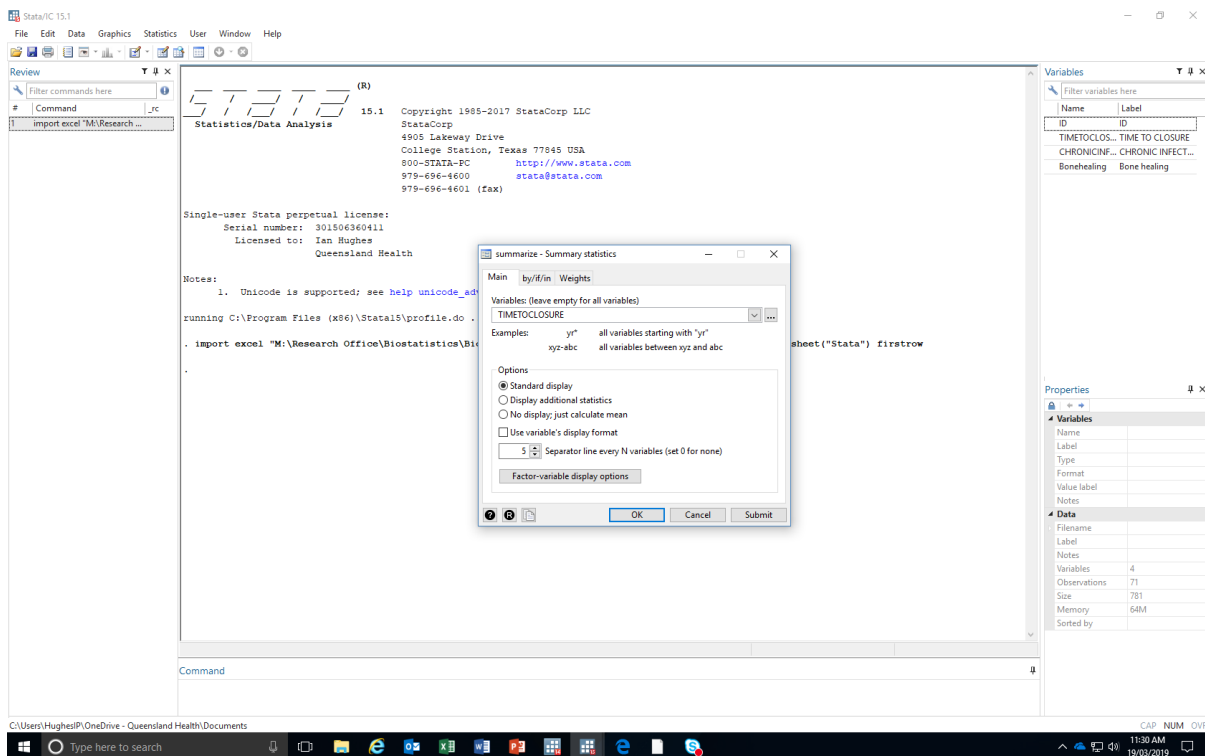
The figure below shows the layout of the Stata home screen once data has been imported.



The drop-down menus can be used to select statistics, graphics etc you may be interested in. To investigate the data we can do some summary statistics. Click on Statistics, Summaries..., Summary and descriptive, Summary statistics as shown below.



A dialog box appears. Click on the arrow button of the Variables box and select the variable you would like to find the summary statistics of. Click ok.



The results are displayed in the main window. Immediately above the results is the syntax that produced the results. You could type this into the Command window to get the same

results as from using the dropdown menus. The syntax is also shown in the left-hand panel which shows a history of commands. Clicking on a command here will copy it to the command window.

The screenshot shows the Stata 15.1 interface. The left-hand panel (Review) contains a list of commands:


```
1 import excel "M:\Research Office\Biostatistics\Biostatistics\Clients\... Ortho data(1).xlsx", sheet("Stata") firstrow
2 summarize TIMETOCLOSURE
```

 The main window displays the output of the `summarize` command:


```

    Copyright 1985-2017 StataCorp LLC
    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)

    Single-user Stata perpetual license:
    Serial number: 301506360411
    Licensed to: Ian Hughes
    Queensland Health

    Notes:
    1. Unicode is supported; see help unicode_advice.

    running C:\Program Files (x86)\Stata15\profile.do ...

    . import excel "M:\Research Office\Biostatistics\Biostatistics\Clients\... Ortho data(1).xlsx", sheet("Stata") firstrow
    . summarize TIMETOCLOSURE

    Variable | Obs   Mean   Std. Dev.   Min   Max
    -----+-----
    TIMETOCLOS-E | 71   414.9155   411.4733    19   2424
    
```

 The right-hand panel (Variables) shows the variable `TIMETOCLOSURE` with its label `TIME TO CLOSURE`. The Properties panel shows the variable's details, including its type, format, and size.

Click on the last command in the left-hand panel. Go to the Command window and add a comma and “detail”, as shown below. Click enter.

The screenshot shows the Stata Command window. The command `summarize TIMETOCLOSURE, detail` is entered into the command window. The background shows a Windows taskbar with the time 11:32 AM on 19/03/2019.

We now have a more complete set of summary statistics.

```

. summarize TIMETOCLOSURE

Variable | Obs      Mean      Std. Dev.   Min      Max
-----+-----+-----+-----+-----+-----
TIMETOCLOS~E |      71  414.9155  411.4733    19     2424

. summarize TIMETOCLOSURE, detail

                TIME TO CLOSURE
-----+-----+-----+-----+-----+-----
Percentiles  Smallest
1%           19           19
5%          60.5          34
10%         72           52      Obs           71
25%        180          60.5     Sum of Wgt.    71

50%         288
75%         560          Largest
90%        843.7         1224     Mean           414.9155
95%        1224         1446     Std. Dev.      411.4733
99%        2424         2424     Variance      169310.2
                                           Skewness      2.437968
                                           Kurtosis      10.5143

```



Click on the Graphics tab. Select histogram. Choose a variable and select whether the data is continuous or discrete. Select what you want on the Y axis

The screenshot shows the Stata 15.1 interface. The command window contains the following text:

```

Single-user Stata perpetual license:
Serial number: 301506360411
Licensed to: Ian Hughes
Queensland Health

Notes:
1. Unicode is supported; see help unicode_advice.

running C:\Program Files (x86)\Stata15\profile.do ...

. import excel "M:\Research Office\Biostatistics\Bio...

. summarize TIMETOCLOSURE

Variable | Obs      Mean      Std. Dev.
-----+-----+-----+-----
TIMETOCLOS~E |      71  414.9155  411.4733

. summarize TIMETOCLOSURE, detail

                TIME TO CLOSURE
-----+-----+-----+-----+-----+-----
Percentiles  Smallest
1%           19           19
5%          60.5          34
10%         72           52      Obs           71
25%        180          60.5     Sum of Wgt.    71

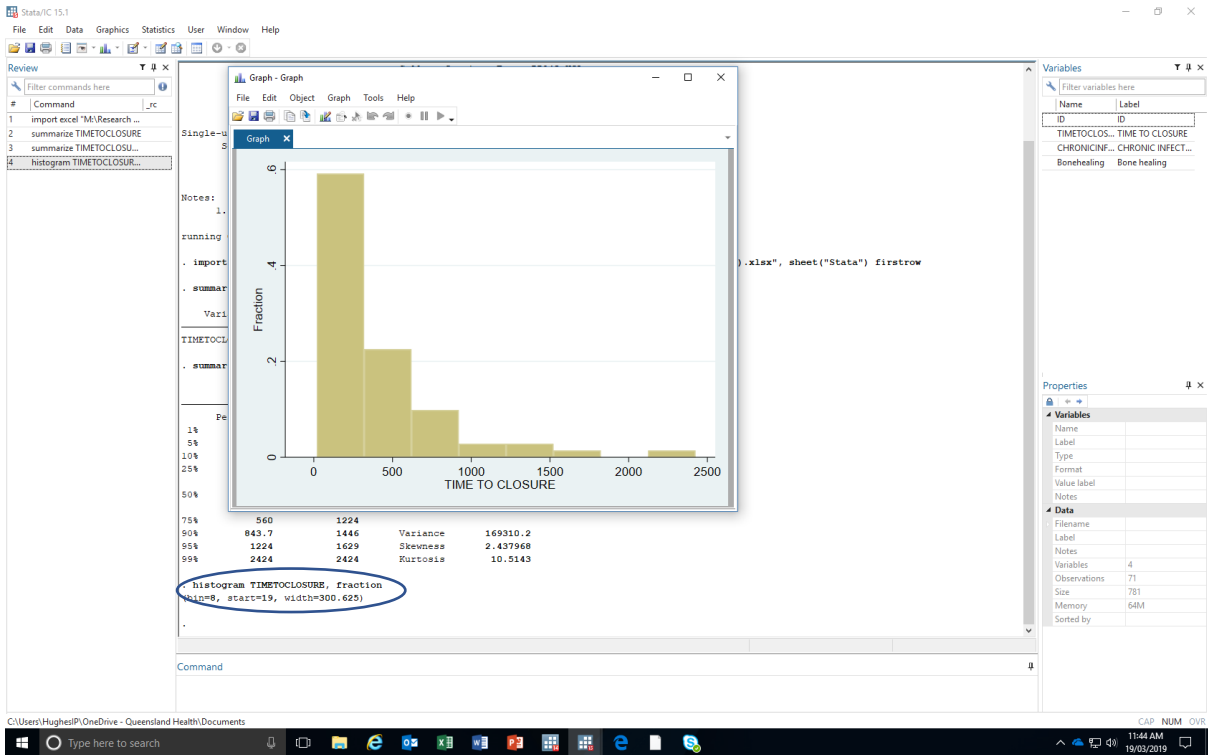
50%         288
75%         560          Largest
90%        843.7         1224     Mean           414.9155
95%        1224         1446     Std. Dev.      411.4733
99%        2424         2424     Variance      169310.2
                                           Skewness      2.437968
                                           Kurtosis      10.5143

```

The histogram dialog box is open, showing the following options:

- Variable: TIMETOCLOSURE
- Data are continuous
- Data are discrete
- Y axis: Density, Fraction, Frequency, Percent
- Number of bins: 10
- Width of bins: []
- Lower limit of first bin: []
- Add height labels to bars
- Recalculate bin sizes when (y) is specified

The graphic is shown in a separate window. It can be copied to a Word file. There are many options available to make graphics look really good.

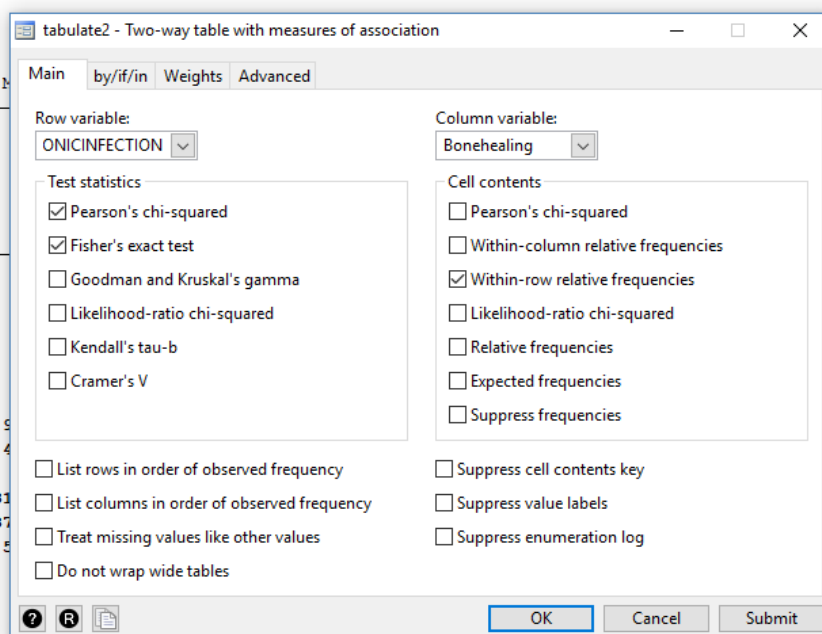


To look for an association between two categorical variables follow the tabs below.

The screenshot shows the Stata/IC 15.1 interface with the 'Statistics' menu open. The path to 'Two-way table with measures of association' is highlighted:

- Statistics
 - Summaries, tables, and tests
 - Frequency tables
 - Two-way table with measures of association

ev.
33
t.
414.9
411.4
16931
2.437
10.5



The results are shown below.

Review

#	Command	25%	50%	75%	90%	95%	99%	Sum of Wgt.	Mean	Largest	Std. Dev.
1	import excel "M&M Research ..."							71	414.9155		411.4733
2	summarize TIMETOCLOSURE		288	560	843.7	1224	1446				
3	summarize TIMETOCLOSURE				1224	1629	2424		169310.2		2.437968
4	histogram TIMETOCLOSURE									2424	10.5143
5	tabulate CHRONICINFECTION Bonehealing, chi2 exact row										

```

. histogram TIMETOCLOSURE, fraction
  (bin=9, start=19, width=300.625)

. tabulate CHRONICINFECTION Bonehealing, chi2 exact row

Key
-----
frequency
row percentage

Enumerating sample-space combinations:
stage 5: enumerations = 1
stage 4: enumerations = 2
stage 3: enumerations = 5
stage 2: enumerations = 16
stage 1: enumerations = 0

CHRONIC INFECTION | Bone healing | Total
-----|-----|-----
              | 1       2       3       4       5       |
0          | 38      2       5       3       1       | 49
          | 77.55   4.08   10.20   6.12   2.04   | 100.00
1          | 14      3       2       3       0       | 22
          | 63.64   13.64   9.09   13.64   0.00   | 100.00
Total     | 52      5       7       6       1       | 71
          | 73.24   7.04   9.86   8.45   1.41   | 100.00

Pearson chi2(4) = 3.8521  Pr = 0.426
Fisher's exact = 0.402

```

Command

```

. tabulate CHRONICINFECTION Bonehealing, chi2 exact row

```

Say we actually wanted the table the other way around. It is easiest to click on the last command (left panel) and then change the order of the variables in the command window as shown below.


```

|       73.24       7.04       9.86       8.45       1.41 |       100.00

Pearson chi2(4) =   3.8521   Pr = 0.426
Fisher's exact =                0.402

Command
tabulate Bonehealing CHRONICINFECTION, chi2 exact row

```

Stata/IC 15.1

File Edit Data Graphics Statistics User Window Help

Review

```

Filter commands here
# Command _jc
1 import excel "M:\Research..."
2 summarize TIMETOCLOSURE
3 summarize TIMETOCLOSUR...
4 histogram TIMETOCLOSUR...
5 tabulate CHRONICINFECTI...
6 tabulate Bonehealing CHR...

```

```

Pearson chi2(4) =   3.8521   Pr = 0.426
Fisher's exact =                0.402

. tabulate Bonehealing CHRONICINFECTION, chi2 exact row

```

Bone healing	CHRONIC INFECTION		Total
	0	1	
1	38 73.08	14 26.92	52 100.00
2	2 40.00	3 60.00	5 100.00
3	5 71.43	2 28.57	7 100.00
4	3 50.00	3 50.00	6 100.00
5	1 100.00	0 0.00	1 100.00
Total	49 69.01	22 30.99	71 100.00

```

Enumerating sample-space combinations:
stage 5: enumerations = 1
stage 4: enumerations = 2
stage 3: enumerations = 5
stage 2: enumerations = 16
stage 1: enumerations = 0

Command

```

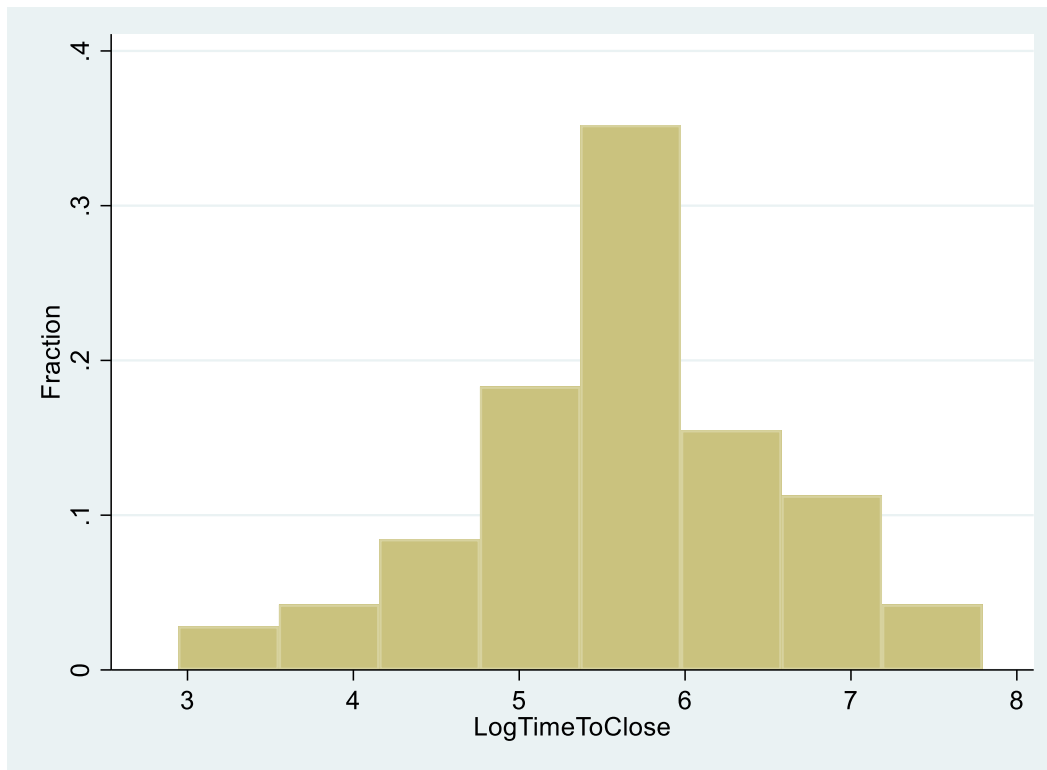
To generate a new variable you can follow the Data tab then select “create or change data”. Or, you can go to the Command window and write generate (or gen) the name you choose for the new variable, =, and then some function that creates the new variable. In the example below, as TIME TO CLOSURE was a skewed distribution I decided to log transform it.

```

Command
gen LogTimeToClose = ln( TIMETOCLOSURE)

```

d Health\Documents



The LogTimeToClose variable looks quite symmetrical. We can use a formal test of normality. Eg the Shapiro-Wilk test.

Stata/IC 15.1

File Edit Data Graphics Statistics User Window Help

Review

Filter commands here

Command

- 1 import excel "M:\Research .
- 2 summarize TIMETOCLOSUR
- 3 summarize TIMETOCLOSUR
- 4 histogram TIMETOCLOSUR
- 5 tabulate CHRONICINFECTI
- 6 tabulate Bonehealing CHR.
- 7 gen LogTimeToClose = ln(
- 8 ranksum TIMETOCLOSURE,
- 9 tabstat TIMETOCLOSURE, ..
- 10 histogram LogTimeToClos.

Statistics

- Summaries, tables, and tests
 - Summary and descriptive statistics
 - Frequency tables
 - Other tables
 - Classical tests of hypotheses
 - Nonparametric tests of hypotheses
 - Distributional plots and tests
 - Symmetry plot
 - Quantiles plot
 - Normal quantile plot
 - Normal probability plot, standardized
 - Chi-squared quantile plot
 - Chi-squared probability plot
 - Quantile-quantile plot
 - Stem-and-leaf display
 - Letter-value display
 - Generate cumulative distribution
 - Skewness and kurtosis normality test
 - Shapiro-Wilk normality test**
 - Shapiro-Francia normality test
 - Ladder of powers
 - Ladder-of-powers histograms
 - Ladder-of-powers quantile-normal plot
 - Multivariate test of means, covariances, and normality
- sum (Mann-Whitney) test

rank	sum	expected
1708.5		1764
847.5		792
	2556	2556
- Time series
- Multivariate time series
- Spatial autoregressive models
- Longitudinal/panel data
- Multilevel mixed-effects models
- Survival analysis
- Epidemiology and related
 - 6468.00
 - 3.69
- Endogenous covariates
- Sample-selection models
 - 6464.31
- Treatment effects
- SEM (structural equation modeling)
- LCA (latent class analysis)
- FMM (finite mixture models)
- IRT (item response theory)
- Survey data analysis
- Multiple imputation
- Nonparametric analysis
- Multivariate analysis

mean	sd	p25	p50	p75
1098	299.188	184.4	241.6	479.8
3273	582.7058	168	300	768
9155	411.4733	180	288	560
- Exact statistics
- Resampling
- Power and sample size
- Bayesian analysis
- Postestimation
- Other

```

. histogram LogTimeToClose, fraction
(bin=8, start=2.9444389, width=.60609192)

```

```

. histogram LogTimeToClose, fraction
(bin=8, start=2.9444389, width=.60609192)

. swilk LogTimeToClose

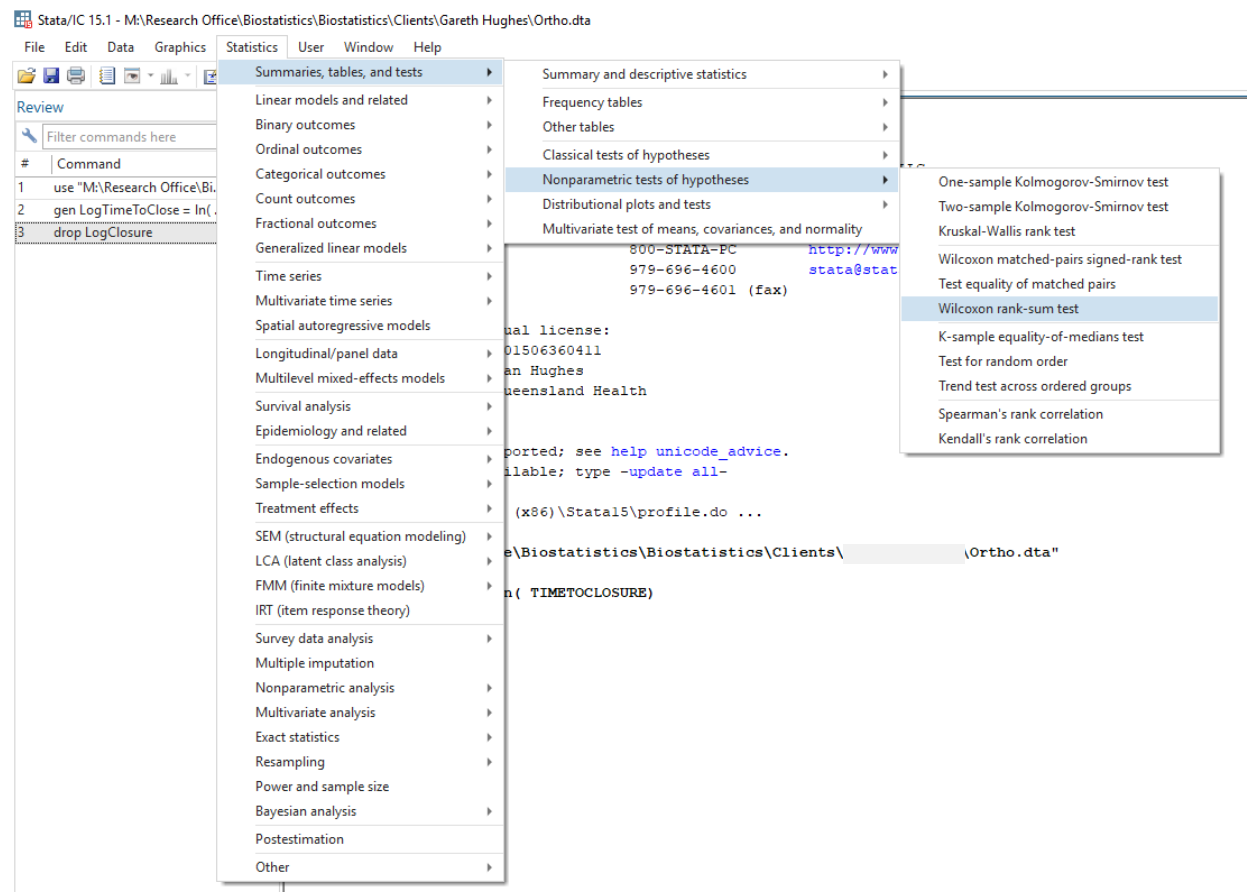
```

Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Prob>z
LogTimeToClose	71	0.98790	0.753	-0.616	0.73112

Command

As suspected, the transformed variable is not significantly different from normal.

We can thus test to see if time to closure is different between the chronic infection group and the non-chronic infection group by either doing a non-parametric test (Wilcoxon rank-sum test, also called the Mann-Whitney U test) on the original data or we can do a t-test to compare the means of the log-transformed time to closure.



Stata/IC 15.1

File Edit Data Graphics Statistics User Window Help

Review

Filter commands here

#	Command	_rc
1	import excel "M:\Research ...	
2	summarize TIMETOCLOSURE	
3	summarize TIMETOCLOSUR...	
4	histogram TIMETOCLOSUR...	
5	tabulate CHRONICINFECTI...	
6	tabulate Bonehealing CHR...	
7	gen LogTimeToClose = ln(...	
8	ranksum TIMETOCLOSURE, ...	
9	tabstat TIMETOCLOSURE, ...	

Total

Pea

Fi

. gen LogTime

. ranksum TIM

Two-sample Wi

CHRONICINF~N

0

1

49

22

1708.5

847.5

1764

792

Help

- PDF documentation
- Advice
- Contents
- Search...
- Stata command...
- Announcements
- News
- Resources
- SJ and community-contributed commands
- What's new?
- Check for updates
- About Stata

Stata command

Command:

tabstat

OK Cancel

1 05 50 251

Viewer - help tabstat

File Edit History Help

help tabstat

help tabstat help tabstat x

Dialog Also see Jump to

[R] **tabstat** — Compact table of summary statistics
 (View complete PDF manual entry)

Syntax

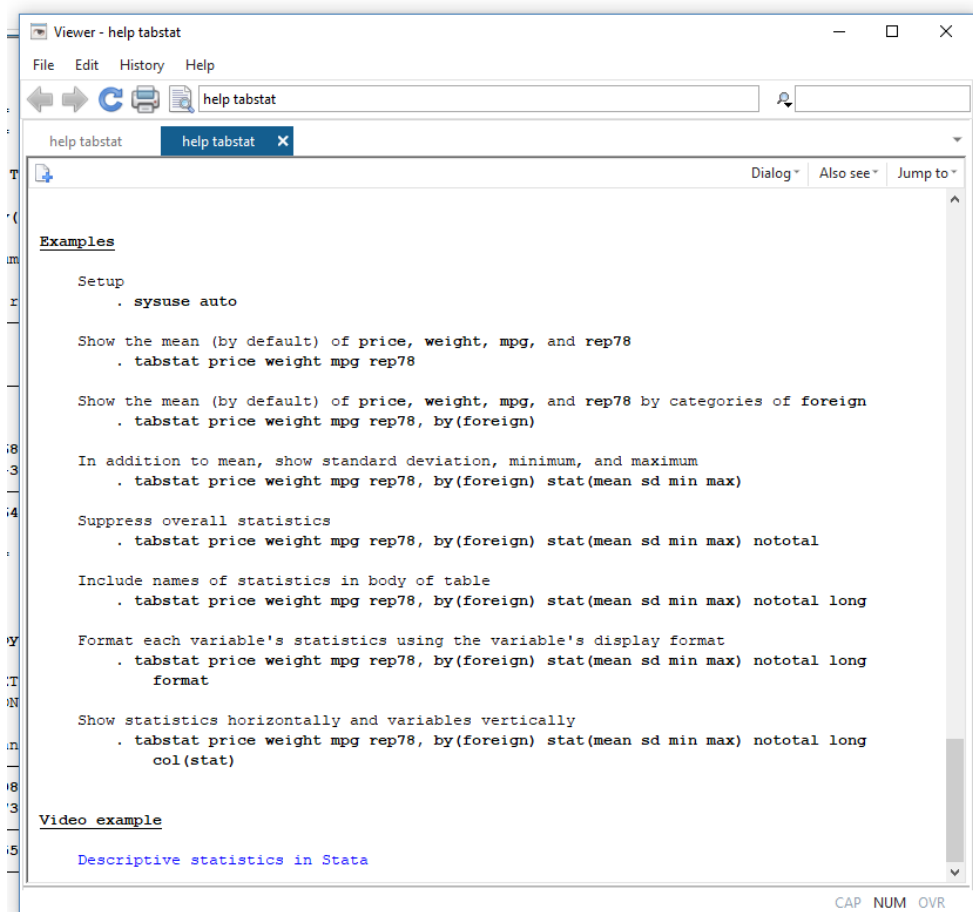
```
tabstat varlist [if] [in] [weight] [, options]
```

options	Description
Main	
<code>by(varname)</code>	group statistics by variable
<code>statistics(statname [...])</code>	report specified statistics
Options	
<code>labelwidth(#)</code>	width for by() variable labels; default is labelwidth(16)
<code>varwidth(#)</code>	variable width; default is varwidth(12)
<code>columns(variables)</code>	display variables in table columns; the default
<code>columns(statistics)</code>	display statistics in table columns
<code>format(%fmt)</code>	display format for statistics; default format is %9.0g
<code>casewise</code>	perform casewise deletion of observations
<code>nototal</code>	do not report overall statistics; use with by()
<code>missing</code>	report statistics for missing values of by() variable
<code>noseparator</code>	do not use separator line between by() categories
<code>longstub</code>	make left table stub wider
<code>save</code>	store summary statistics in r()

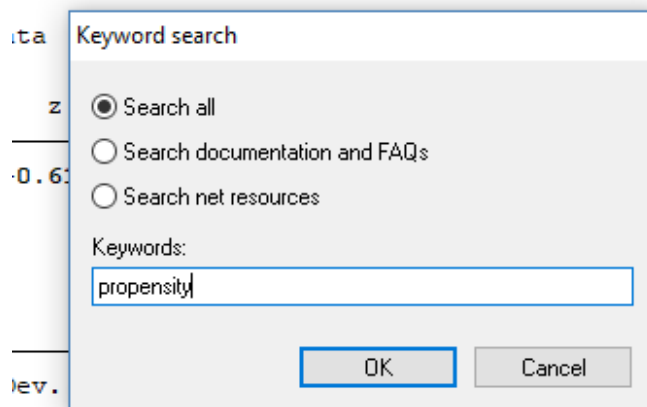
by is allowed; see [D] by.
 aweights and fweights are allowed; see weight.

Menu

CAP NUM OVR

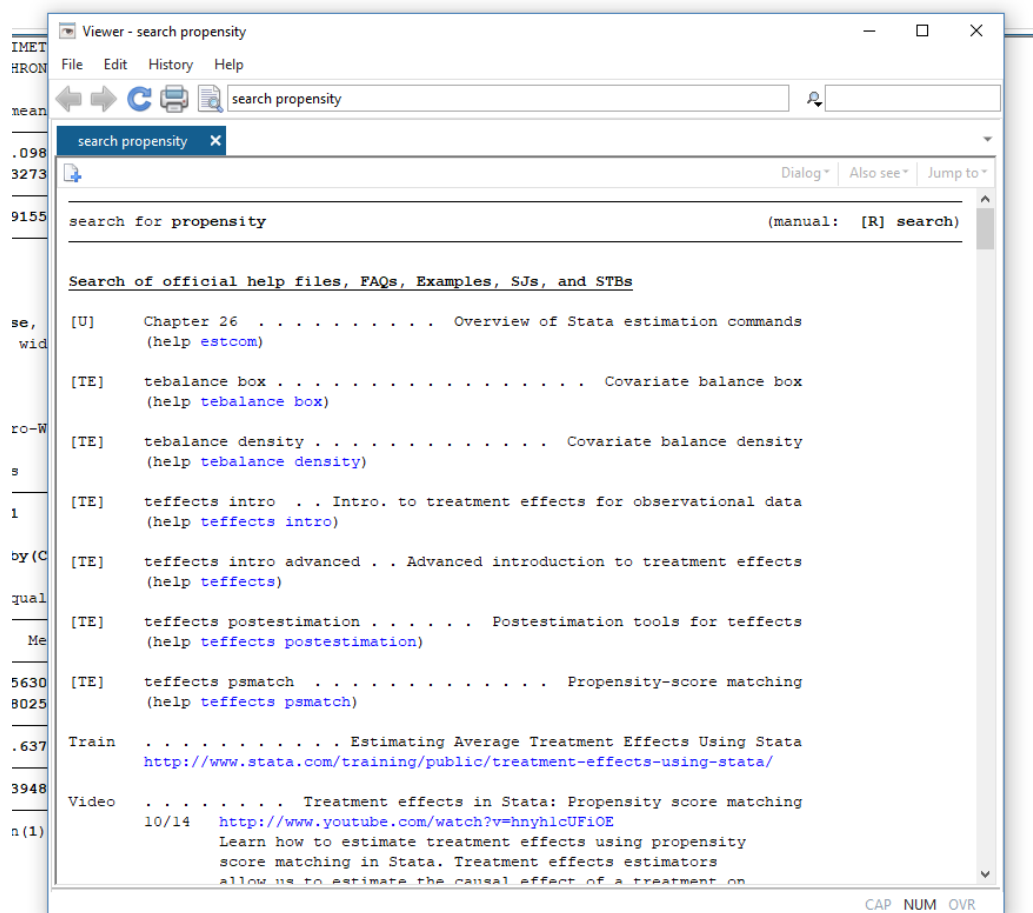


If you don't know the name of the command but want to find out if Stata has any commands that might be useful for a particular topic you can click on Help then search. For example, say we are interested in propensity scores.

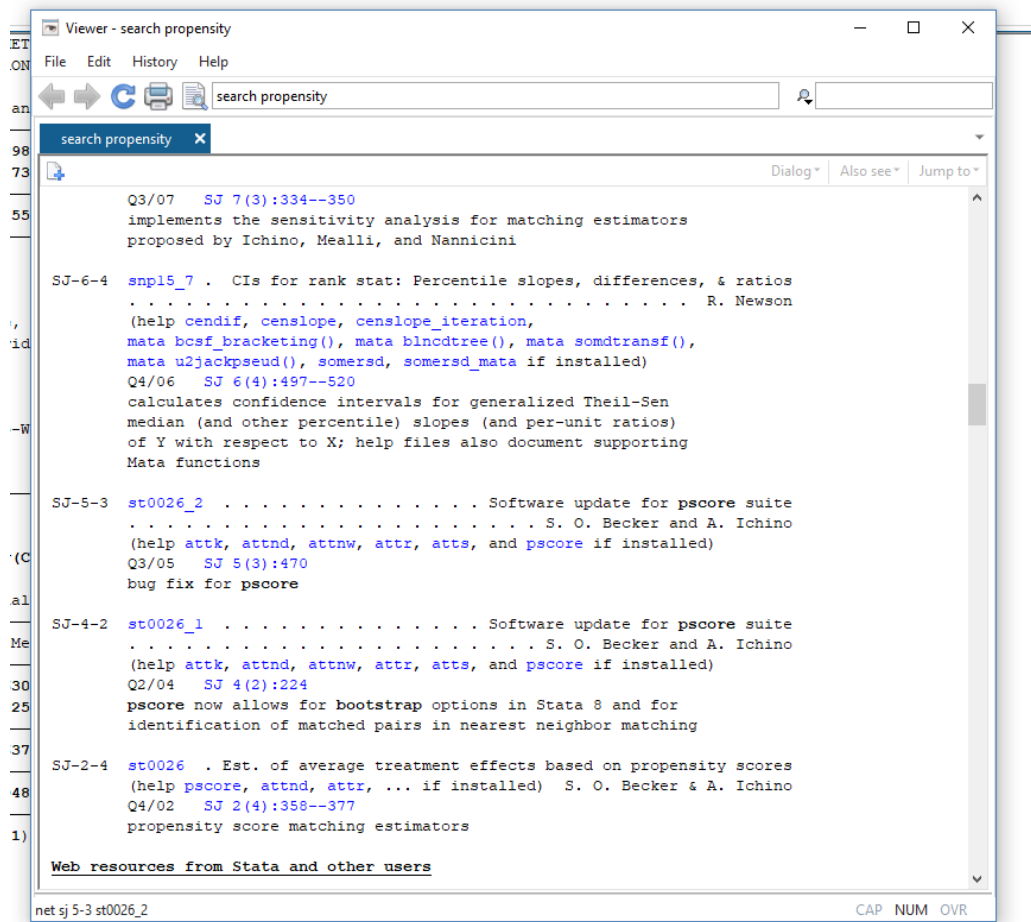


166 5.307818 5.81823

Various sources of information are found.



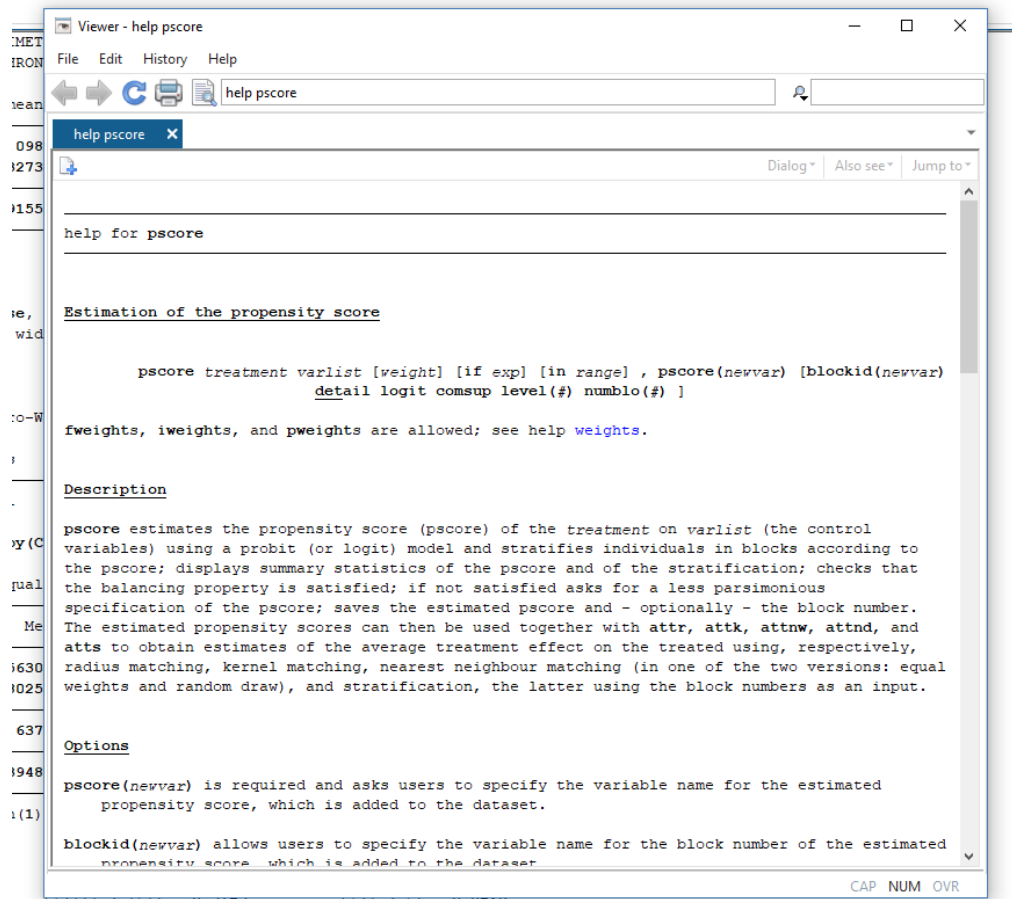
Further down you will find links to Stata packages that may be useful for calculation of propensity scores.



The latest update for the [pscore](#) suite looks like it is the best option. Click on [st0026_2](#).

```
Viewer - net sj 5-3 st0026_2
File Edit History Help
net sj 5-3 st0026_2
net sj 5-3 st0026_2 x
Dialog Also see Jump to
package st0026_2 from http://www.stata-journal.com/software/sj5-3
TITLE
  SJ5-3 st0026_2. Estimation of average treatment effects
DESCRIPTION/AUTHOR(S)
  Estimation of average treatment effects
  by Sascha O. Becker, University of Munich
  Andrea Ichino, EUI
  Support: so.b@gmx.net, andrea.ichino@iue.it
  After installation, type help attk, atnd, atnw,
  attr, atts, and pscore
INSTALLATION FILES (click here to install)
  st0026_2/attk.ado
  st0026_2/attk.hlp
  st0026_2/atnd.ado
  st0026_2/atnd.hlp
  st0026_2/atnw.ado
  st0026_2/atnw.hlp
  st0026_2/attr.ado
  st0026_2/attr.hlp
  st0026_2/atts.ado
  st0026_2/atts.hlp
  st0026_2/pscore.hlp
  st0026_2/pscore.ado
(click here to return to the previous screen)
CAP NUM OVR
```

Click on “click here to install”. To check if it has installed, click on Help, Stata command., and type pscore. Now the help file should come up telling you how to use the command you just downloaded.



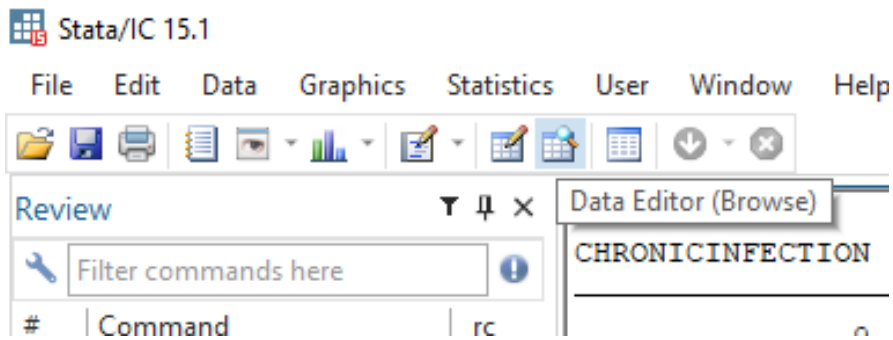
The following illustrates how to create, manipulate, and view variables. You can create a variable that is a random selection of values from a normal distribution.

```
. gen random = rnormal()
.
```

Command

The r is for “random”. Other distributions could be used too. For details search random number generation.

We can view the data, including the new variable we called “random” by clicking on the Data Editor (Browse) icon.



Data Editor (Browse) - [Untitled]

File Edit View Data Tools

ID	1					
ID	TIME TO CLOS-E	CHRONIC INF-N	Bonehealing	LogTimeToC-e	random	
1	1	19	0	1	2.944439	2.025588
2	2	34	0	1	3.526361	1.042631
3	3	52	1	1	3.951244	.2977124
4	4	60.5	0	1	4.102643	-1.722132
5	5	63	0	1	4.143135	-.7291995
6	6	72	0	1	4.276666	.8618261
7	7	72	0	3	4.276666	-.239354
8	8	72	1	4	4.276666	.516549
9	9	96	0	1	4.564348	-1.812016
10	10	109	1	1	4.691348	-1.015124
11	11	115	0	1	4.744932	1.095368

We can change a variable by using the replace command. In this case we are going to change all values in random to 0 if the original value was less than or equal to 0.

```
. replace random = 0 if random <=0
(41 real changes made)
.
```

Command

Check to see what happened.

Data Editor (Browse) - [Untitled]

File Edit View Data Tools

ID[1] 1

	ID	TIMETOCLOS~E	CHRONICINF~N	Bonehealing	LogTimeToC~e	random		
1	1	19	0	1	2.944439	2.025588		
2	2	34	0	1	3.526361	1.042631		
3	3	52	1	1	3.951244	.2977124		
4	4	60.5	0	1	4.102643	0		
5	5	63	0	1	4.143135	0		
6	6	72	0	1	4.276666	.8618261		
7	7	72	0	3	4.276666	0		
8	8	72	1	4	4.276666	.516549		
9	9	96	0	1	4.564348	0		
10	10	109	1	1	4.691348	0		
11	11	115	0	1	4.744932	1.095368		
12	12	141	1	1	4.94876	1.332649		

Now replace all of the positive values with 1.

```
. replace random = 0 if random <=0
(41 real changes made)

. replace random = 1 if random >0
(30 real changes made)

.
```

Command

Data Editor (Browse) - [Untitled]

File Edit View Data Tools

ID[1] 1

	ID	TIMETOCLOS~E	CHRONICINF~N	Bonehealing	LogTimeToC~e	random		
1	1	19	0	1	2.944439	1		
2	2	34	0	1	3.526361	1		
3	3	52	1	1	3.951244	1		
4	4	60.5	0	1	4.102643	0		
5	5	63	0	1	4.143135	0		
6	6	72	0	1	4.276666	1		
7	7	72	0	3	4.276666	0		
8	8	72	1	4	4.276666	1		
9	9	96	0	1	4.564348	0		
10	10	109	1	1	4.691348	0		
11	11	115	0	1	4.744932	1		

And rename the variable "sex". Obviously we would usually not be randomly assigning sex to participants but it gives me the excuse to demonstrate how to assign labels to categorical variables.

```

. replace random = 1 if random >0
(30 real changes made)

. rename random sex

.

```

Command

Here we create a label file called Gender which assigns the label "Female" to the 0s and "male" to the 1s. In the next line I tell Stata to label the values of the variable sex according to the definitions in the label file Gender.

```

. label define Gender 0 "Female" 1 "Male"

. label values sex Gender

.

```

Data Editor (Browse) - [Untitled]

File Edit View Data Tools



var11[17]							
	ID	TIMETOCLOS-E	CHRONICINF-N	Bonehealing	LogTimeToC~e	sex	
1	1	19	0	1	2.944439	Male	
2	2	34	0	1	3.526361	Male	
3	3	52	1	1	3.951244	Male	
4	4	60.5	0	1	4.102643	Female	
5	5	63	0	1	4.143135	Female	
6	6	72	0	1	4.276666	Male	
7	7	72	0	3	4.276666	Female	
8	8	72	1	4	4.276666	Male	

Now, when you produce tables for example they have the labels.

```

. tabulate CHRONICINFECTION sex

```

CHRONIC INFECTION	sex		Total
	Female	Male	
0	29	20	49
1	12	10	22
Total	41	30	71

This should be enough to get you started with Stata. There are many resources available beginning with the Help icon in Stata. Below are some really useful resources. The last one,

the UCLA site is very good, it gives the same examples in Stata, SPSS, and R. And of course Googling “How to in Stata” works!

Free resources – webinars (<https://www.stata.com/training/webinar/>); manuals (<https://surveydesign.com.au/manuals.html>); or YouTube videos (<https://www.youtube.com/channel/UCV4k4G4nEtBS4tLOyHqustDA>)

Paid resources – Stata Press books (<https://surveydesign.com.au/statabooks.html>); Stata online NetCourses (<https://surveydesign.com.au/events.html>) or other public training events (<https://www.stata-au.com/events>)

<https://stats.idre.ucla.edu/stata/>

Good luck with Stata!

Ian Hughes

Biostatistician GCHHS