

# Stata Tutorial

---

ECONOMETRIC ANALYSIS

SEPT 2023



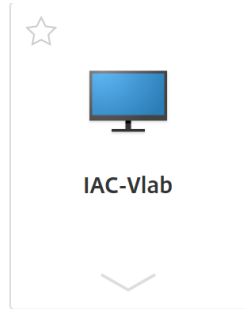
# Use Stata via Vlab

---

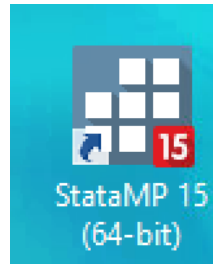
1, log on through

<https://mycloud.gatech.edu/vpn/index.html>

2, click desktop -> click



3, click the Stata shortcut on the desktop



# Get Started on Stata

The screenshot shows the Stata 12.1 interface. The main window is divided into three panes:

- Command view** (left pane): Shows a list of commands entered in the Command window.

#	Command	_rc
1	sysuse auto.dta	
2	replace price = 4098 ...	
3	help	
- Result window** (middle pane): Displays the output of the commands. It shows the Stata logo, version 12.1, copyright information (1985-2011 StataCorp LP), and the license details for the Department of Economics at Texas A&M University. It also shows the results of the `sysuse auto.dta` command, indicating that the data is from 1978 and that the price variable was changed for the first observation.

MP - Parallel Edition

10-user 4-core Stata network perpetual license:  
Serial number: 50120555476  
Licensed to: Department of Economics  
Texas A&M University

Notes:  
1. (/v# option or -set maxvar-) 5000 maximum variables  
2. New update available; type -update all-

. sysuse auto.dta  
(1978 Automobile Data)

. replace price = 4098 in 1  
(1 real change made)

. help

.
- Variables** (right pane): A list of variables in the current dataset, including make, price, mpg, rep78, headroom, trunk, weight, length, turn, displacement, gear\_ratio, and foreign. Below this is a Properties window showing details for the 'make' variable, such as its name, label, type, format, and value labels.

Command window

Stata is a command-driven package. Its source code is written in C/C++.

You can enter commands in following ways:

1. Type the first command window and execute it, then the next,...
2. Do-file: type up a list commands in a 'do-file', and execute the do-file.

# Directory and folder

---

- `cd (or pwd)`: display the current directory
- `dir`: list contents of directory or folder
- `mkdir "C:\Stata"`: creates a new directory in
- `cd " C:\Stata"` : change directory to `"C:\stata"`

\*Almost the same as windows command line.

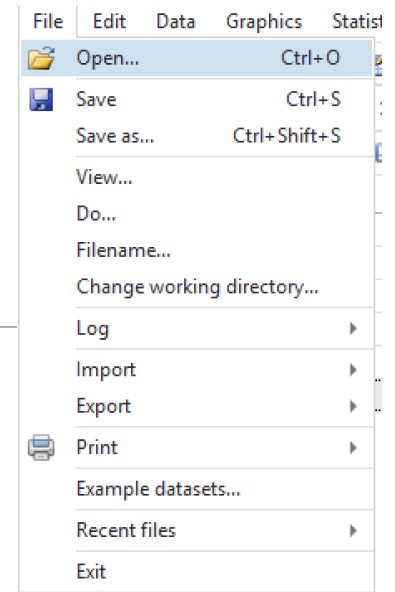
# Reading data into stata

- Stata data format: dta
- If your data in stata format, then simply read it as:  
use "C:\Users\xyan306\Desktop\stataDataset\wage2.dta", clear  
or if you have changed the directory, you can just type "use wage2.dta, clear"
- If your data is originally from Excel or some other format, you need to save the data as either csv or txt-file for reading into stata.

The first line in the spreadsheet should have the variable names, and the second line onwards are the data

Once your data is the csv or txt format, you can read it as:

```
import excel "C:\Users\xyan306\Desktop\stataDataset\wage2.xls", sheet("Sheet1") firstrow clear
```



# Examine the data

---

- list in 2/5 (list row 2 to row 5, including 2 and 5)
- list if wage==1000 (list the row(s) if and only if price equals to 1000)
- count if wage==1000
- describe(describe the type and format)
- summarize (generate summary statistics)
- summarize, detail (with percentile and moments)
- scatter price weight (plot the data with scatter)

help scatter to see more properties on scatter command

# Do-files

---

`clear` (clear all data in the RAM allocated to Stata)

`cd "c:\stata"` (change directory)

# More commands

---

- gen: creates a new variable

gen logwage= log(wage)

- replace: changes the value of a variable or observation

replace wage2=2\*wage

- drop: deletes a variable or observation

drop wage2

- keep: specifies which variables to leave in the dataset

- regress

- predict



# More commands

---

- predict r, residual
- predict fitted, xb
- regress .....if.....
- test

test educ=0: test the coefficient on educ is zero

test educ=exper: test the coefficients on the price and length are the same

test educ exper: test coefficients on price and length are jointly zero

# More commands

---

- `ivreg` (instrumental variable regression/2SLS)
- `xtreg` (Panel data regression: Random effects/fixed effects estimation)
- `esttab`: produces a pretty regression table in a format that you specify (.csv, .tex, .txt and others)
- `esttab using example.tex, label nostar title (Regression table \label{tab1})`
- `esttab using example.tex, title (Regression table \label{tab1})`

# Documentation

---

- Like any other programming languages, it has official documentation.
- Google the key words + “stata”

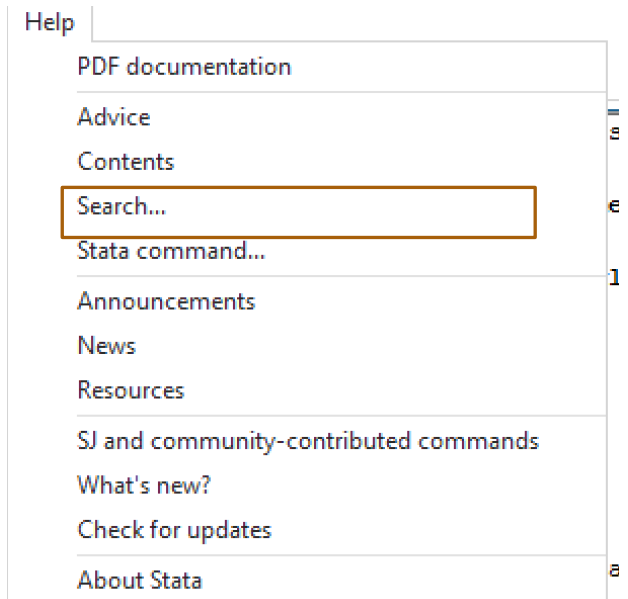


t-test stata

## ttest - Stata

**ttest** performs **t tests** on the equality of means. In the first form, **ttest** tests that varname has a mean of #. In the second form, **ttest** tests that varname has the same ...

- The search function under the help tab is helpful too.



# Output

regress wage hours iq educ exper tenure age married black urban south

Source	SS	df	MS	Number of obs	=	935
				F(10, 924)	=	29.70
Model	37148718.4	10	3714871.84	Prob > F	=	0.0000
Residual	115567450	924	125072.998	R-squared	=	0.2433
				Adj R-squared	=	0.2351
Total	152716168	934	163507.675	Root MSE	=	353.66

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
hours	-3.21257	1.622245	-1.98	0.048	-6.396282	-.0288583
iq	3.549025	.9693936	3.66	0.000	1.646556	5.451493
educ	52.46424	7.011271	7.48	0.000	38.70437	66.2241
exper	9.619544	3.5974	2.67	0.008	2.559522	16.67957
tenure	5.479802	2.41179	2.27	0.023	.7465807	10.21302
age	10.21389	4.550158	2.24	0.025	1.284049	19.14374
married	177.7986	37.88222	4.69	0.000	103.4534	252.1437
black	-125.9797	38.63195	-3.26	0.001	-201.7963	-50.1632
urban	168.8338	26.09792	6.47	0.000	117.6158	220.0519
south	-50.41342	25.57057	-1.97	0.049	-100.5965	-.230294
_cons	-702.3279	175.3895	-4.00	0.000	-1046.536	-358.12

# End

---

**Have fun with Stata!**