CHULALONGKORN BUSINESS SCHOOL
FLAGSHIP FOR LIFE

AACSB ACCREDITED

EFMD EQUIS ACCREDITED

EFMD EPAS ACCREDITED

MSF Chula

Financial Econometrics

Lecture IV:
Inference and Rescaling

Narapong Srivisal, Ph.D.

Population

$y = \beta_0 + \beta_1 x + \ldots + u$

$\beta_1 \overset{?}{=} 0$

Sample

OLS + data

$\Downarrow$

$\hat{\beta}_1 = \overset{3}{a}$ may not equal to 0

because OLS is random

want to answer

Inference :

make conclusion about parameter based on $\hat{\beta}$

Estimator :

- consistent – close to parameter (β must be closed to $\hat{\beta}$)

  ↳ think / ask $\beta = b$

  estimate $\hat{\beta} = a$

  b and a close

  possible that $\beta = b$

  a, b far from each other

  → unlikely that $\beta = b$

Research question : x has impact on y or not ?

$y = \beta_0 + \beta_1 x + \text{controls} + u$

$\hat{\beta}_1 = 1$
0.1
0.00001

check $\beta_1 \overset{?}{=} 0$

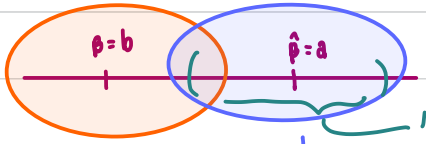
cannot make conclusion ⇒ need to look at context of x and y

e.g. y = wage x = year

y = ln(wage)

need statistical standard to check whether $\hat{\beta}_1$ close to β_1 or not

---



$\hat{\beta} \pm se(\hat{\beta}) \cdot c$

confidence interval

$H_0: \beta = b$

inside or outside interval

reject H_0

interval wide enough to believe that it should cover β

hypothesis testing

confidence interval

* we have to know distribution

Hypothesis test:

Assume $\beta = b$ is true

↓

prob to observe $\hat{\beta}$ further than a

↓

low

↳ reject H_0

Outline

- Review of Important Distributions

Inference Outline

- Hypothesis Testing
  - Test for one parameter
  - Test for a linear combination of parameters
  - Test multiple linear restrictions
- Confidence Interval

Rescaling

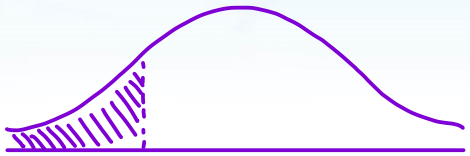
Normal Distribution

- A random variable X is distributed as Normal with expected value equal to μ and variance equal to σ^2 if it has the PDF:

don't need to remember

pdf

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$



- We use the notation $X \sim N(\mu, \sigma^2)$.
- Standard Normal is the Normal Distribution with zero mean and a unit variance, i.e. $Z \sim N(0,1)$
- The Normal PDF function does not have a closed-form integration.
- To find $\Pr\{X \le x\}$, we can transform X to the Standard Normal and use the table for CDF of Standard Normal, which is often called the Z-table.

$$\Pr\{X \le x\} = \Pr\left\{\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right\} = \Pr\left\{Z \le \frac{x-\mu}{\sigma}\right\} = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

value can be only positive

Chi-squared Distribution

- If $Z_1, Z_2, \dots, Z_m \sim iid\ N(0,1)$, then

sum of z_i^2

$$\sum_{i=1}^{m} Z_i^2 \sim \chi_m^2$$

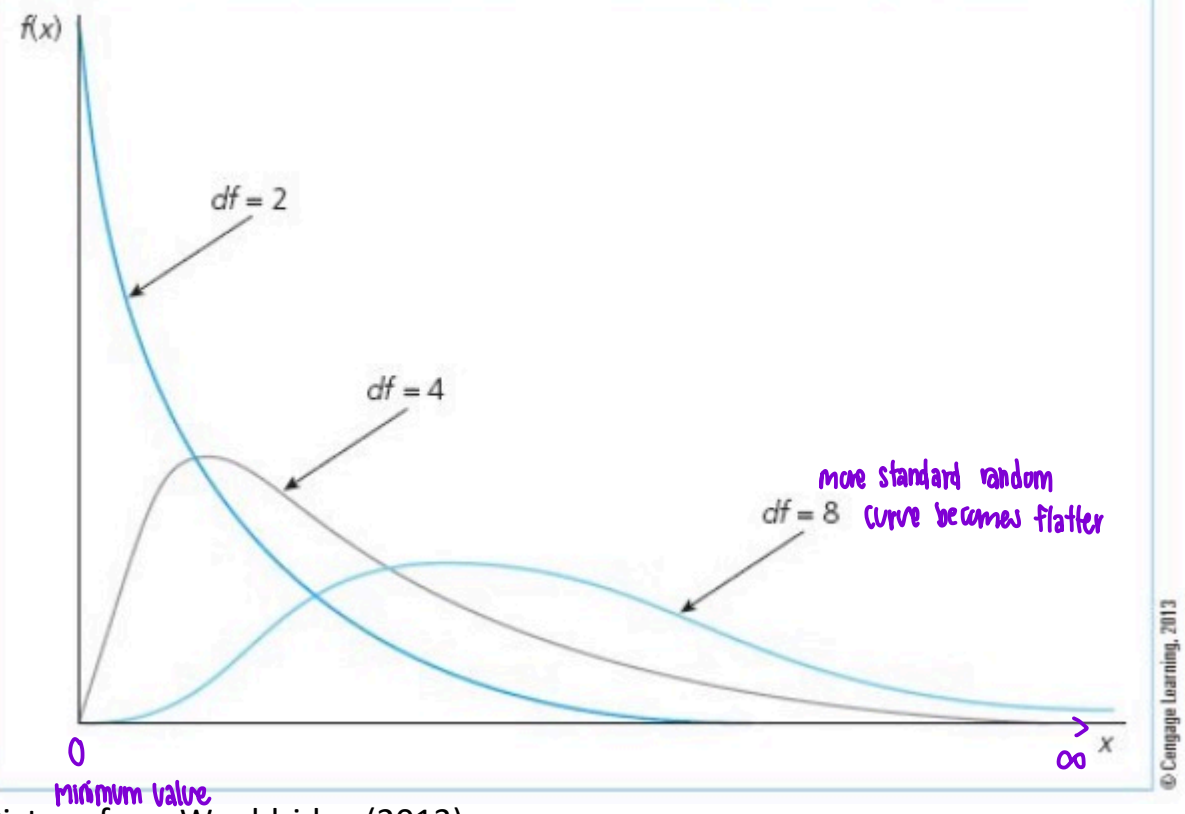
degree of freedom

- We call it **Chi-squared distribution with m degrees of freedom.**
- If $S_1 \sim \chi_{r_1}^2, \dots, S_p \sim \chi_{r_p}^2$ are p independent Chi-squared variables, then

$$\sum_{i=1}^{p} S_i = \chi_{\sum_{j=1}^{p} r_j}^2$$

- If $Z = (Z_1, \dots, Z_m) \sim N(0, \Omega)$, and Ω^{-1} exists, then $Z'\Omega^{-1}Z \sim \chi_m^2$
- If $Z = (Z_1, \dots, Z_m) \sim N(0, I_m)$ and Q is an idempotent matrix ($QQ = Q$) with $\text{rank}(Q) = k$, then $Z'QZ \sim \chi_k^2$
- If $X \sim \chi_m^2$, then $E[X] = m$ and $Var[X] = 2m$.

FIGURE B.9 The chi-square distribution with various degrees of freedom.



Picture from Wooldridge (2013)

Student's *t* Distribution

- If $Z \sim N(0,1)$, $W \sim \chi^2_m$, and the two random variables are independent, then

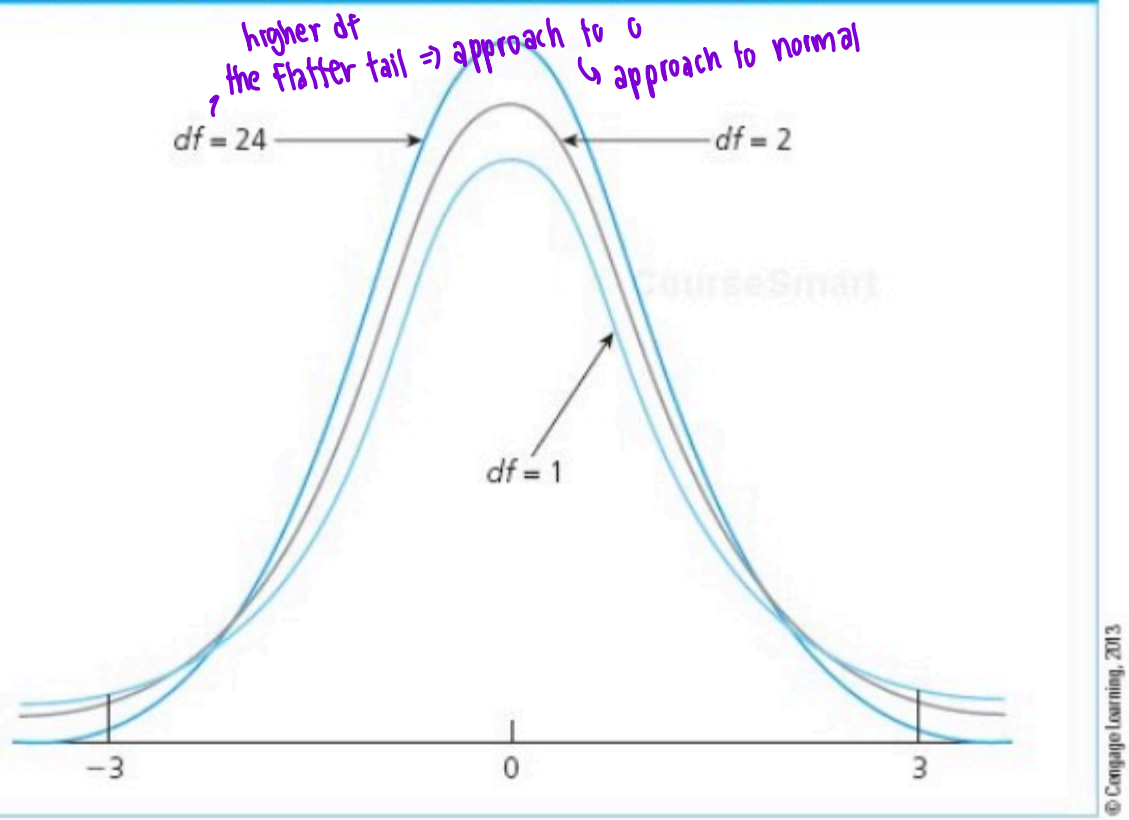
$$\frac{Z}{\sqrt{W/m}} \sim t_m$$

Chi-square

- We call it **Student's *t* distribution with *m* degrees of freedom**.
- If $X \sim t_m$, then $E[X] = 0$ and $Var[X] = \frac{m}{m-2}$, for $m > 2$.
- The larger degree of freedom *m*, the thinner tail is.
- For large value of *m*, the distribution is approximately the same as the standard normal distribution $N(0,1)$.

df higher than the df available given in table ⇒ use z-table instead!!!

FIGURE B.10 The t distribution with various degrees of freedom.



Picture from Wooldridge (2013)

F Distribution

- If $W \sim \chi^2_m$ and $V \sim \chi^2_n$, then

chi-square

$$\frac{\left(\frac{W}{m}\right)}{\left(\frac{V}{n}\right)} \sim F_{m,n}$$

two df from two chi-square

chi-square

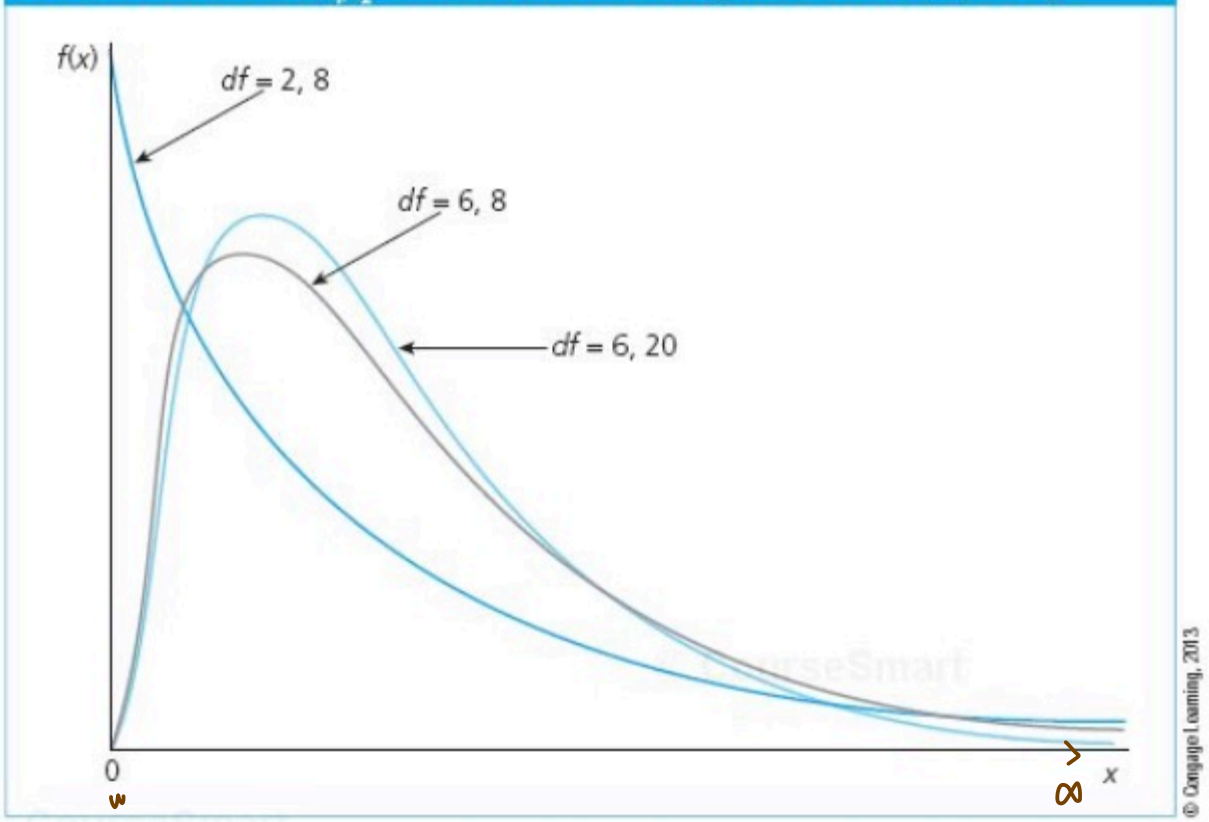
- We call it the ***F* distribution with *m* numerator degrees of freedom and *n* denominator degrees of freedom**.
- If *n* is large, then $mF_{m,n} \approx \chi^2_m$
- **If $t \sim t_n$, then $t^2 \sim F_{1,n}$**

square t dist ⇒ F dist

$$\left(\frac{z}{\sqrt{w/m}} \sim t_m\right)^2 \Rightarrow \frac{z^2}{w/m} \sim F_{1,m}$$

If this is not equare to 1 ⇒ you cannot √ it to become t dist

FIGURE B.11 The F_{k_1,k_2} distribution for various degrees of freedom, k_1 and k_2.



Picture from Wooldridge (2013)

Inference

- Recall that a parameter θ is what we don't know; so, we need to estimate it.
- Statistical inference helps us **make a conclusion about the parameter conditional on an estimated number we have**.
- Suppose somebody suggests that $\theta = b$, but our estimator gives an estimated value $\hat{\theta}_n = a$. Should we conclude that the suggestion is wrong?
- In order to judge this, we may consider whether claimed value b is "*close to*" a.
- Two mainstream processes of making inference: **hypothesis test** and **confidence interval** analyses.

Hypothesis Testing

- In order to judge whether the claimed parameter value $\theta = b$ is "*close to*" an observed estimate $\hat{\theta}_n = a$, the hypothesis testing concept **assumes that $\theta = b$ is true** and examines the probability that one can observe $\hat{\theta}_n$ further from b than a is under this assumption.
- The condition assumed to be true is called **Null Hypothesis, H_0**
- E.g. Suppose we have a linear model $Y = \beta_0 + \beta_1 X + U$, and we want to know if X has any impact on Y. So, we may state the null hypothesis that $\beta_1 = 0$ and see if we can reject this hypothesis.
- If the probability is "*low*," then the statement $\theta = b$ should be rejected.
- Should we conclude that X affects Y if $\hat{\beta}_1 = 1$? Consider answering in the following two contexts:

$$wage = \hat{\beta}_0 + 1(edu) + \hat{U}$$

$$\ln(wage) = \hat{\beta}_0 + 1(edu) + \hat{U}$$

- The same value of estimates may be "far" or "near" in different contexts. So, we must find a standardized way to make a conclusion.
- In order to make a decision on this, we have to find a **Test Statistics**, which

  1st step

  - **is a function of parameter β_1 and random estimator $\hat{\beta}_1$**, say $T(\beta_1, \hat{\beta}_1)$

    to know whether it is likely/unlikely to be the claimed value

  - **has a distribution in the form that we know** E.g. Standard Normal, Student-*t*, Chi-squared, or *F* distribution.

- Then, we can see where $T(\beta = 0, \hat{\beta} = 1)$ is on its distribution and determine if the probability of observing further estimates than $\hat{\beta}_1 = 1$ is high or low.
- Probability of observing further estimates is called the **p-value.**

How we get test statistic for linear regression model

- two kind of test statistic in midterm!
- >10 " ——— n in final

Hypothesis Test for One Parameter

- From CLT, the OLS estimator $\hat{\beta}$ is asymptotically Normal. Thus, each component $\hat{\beta}_j, 0 \leq j \leq k$, is also asymptotically Normal

$$\sqrt{n}(\hat{\beta}_j - \beta_j)/\sqrt{\sigma_{jj}} \xrightarrow{d} N(0,1)$$

  where σ_{jj} is the the variance of the OLS estimator $\hat{\beta}_j$

- Since we don't observe U_i, we don't know the parameter σ_{jj}, which is a function of standard deviation of U.

- Instead, we typically use a **standard error, $se(\hat{\beta}_j)$**, which is an estimator for σ_{jj} based on an appropriated estimator of standard deviation of U, but the distribution will not be Normal but

estimate — parameter according to assumed hypothesis

$$\frac{(\hat{\beta}_j - \beta_j)}{se(\hat{\beta}_j)} \sim t_{n-k-1}$$

n - (k+1) — # betas in the model

n: # of observations

get from software

Standard error of $\hat{\beta}$ ⇒ estimate of S.D.

random ⇒ we also need to estimate it because it also includes parameter that we cannot see

- So, we can define the **test statistics for $H_0: \beta_j = b$** based on

$$\frac{(\hat{\beta}_j - b)}{se(\hat{\beta}_j)} \sim t_{n-k-1}$$

- Then, calculate the p-value, which is the probability that the defined test statistic, $T_n(b, \hat{\beta}_j)$ takes higher value than what we observe, in order to see if it is less than a pre-specified **significance level** (α).
- or we can use the t-table statistics to see if $T_n > c_\alpha$, the **critical value** that corresponding to the significance level α.
- All of the popular statistical software usually reports $se(\hat{\beta}_j)$ for all $0 \leq j \leq k$ along with the OLS estimates.

Two-Sided Test

two-tailed

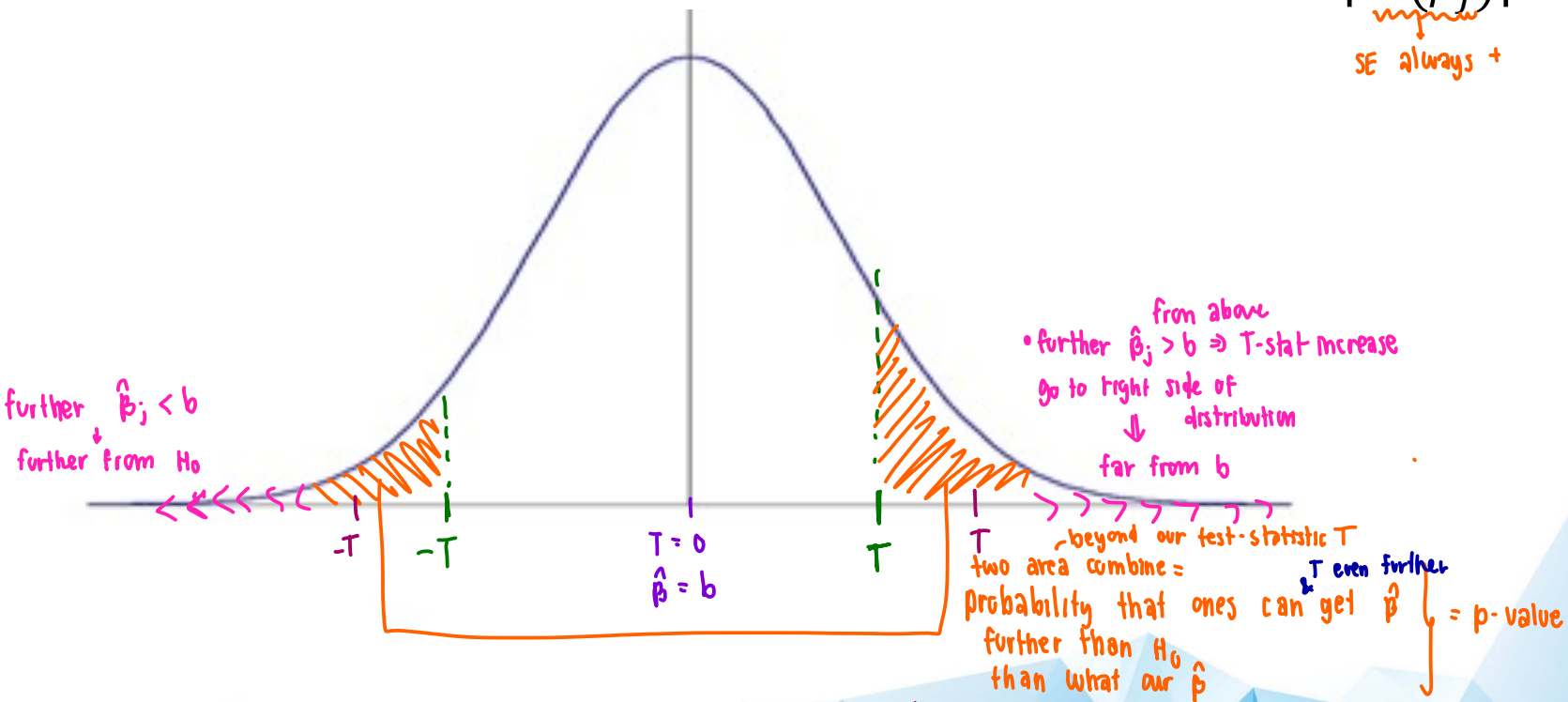
Null Hypothesis $H_0: \beta_j = b$

Alternative Hypothesis $H_1: \beta_j \neq b$

we will reject H_0 if $\hat{\beta}_j$ is far from b

$$T_n(b, \hat{\beta}_j) = \left| \frac{\hat{\beta}_j - b}{se(\hat{\beta}_j)} \right|$$

SE always +



Hypothesis testing :

alternative hypothesis H_a , H_1

1) $H_0 : \beta = b$, $H_1 : \beta \overset{<}{\underset{\neq}{>}} b$

2) Test statistics

3.2 $\alpha \Rightarrow$ find critical value (C_α)

3) 3.1 Test - stat $\Rightarrow$ p - value (area under the distribution curve)

4) 3.2 reject H_0 if p-value is lower than α (significant value) $\longrightarrow$ standard level is 1% , 5% 10%

- 1% → strong evidence
- 5% → standard level of acceptable
- 10% → not strong but acceptable

4.2 reject H_0 if test stat is further than C_α

Compare area

t - stat $\Rightarrow$ $\boxed{\text{p - value} \Rightarrow \alpha \text{ significant level}}$

α $\Rightarrow$ $\boxed{\text{critical value} \Rightarrow \text{t - stat}}$

Compare value on axis

$\rightarrow$ must obtain the same conclusion

grade #hr

$y = \beta_0 + \beta_1 x + \text{control} + u$

+

care only effect is positive

One-Sided Test

Null Hypothesis $H_0: \beta_j = b$ (≤ 0)

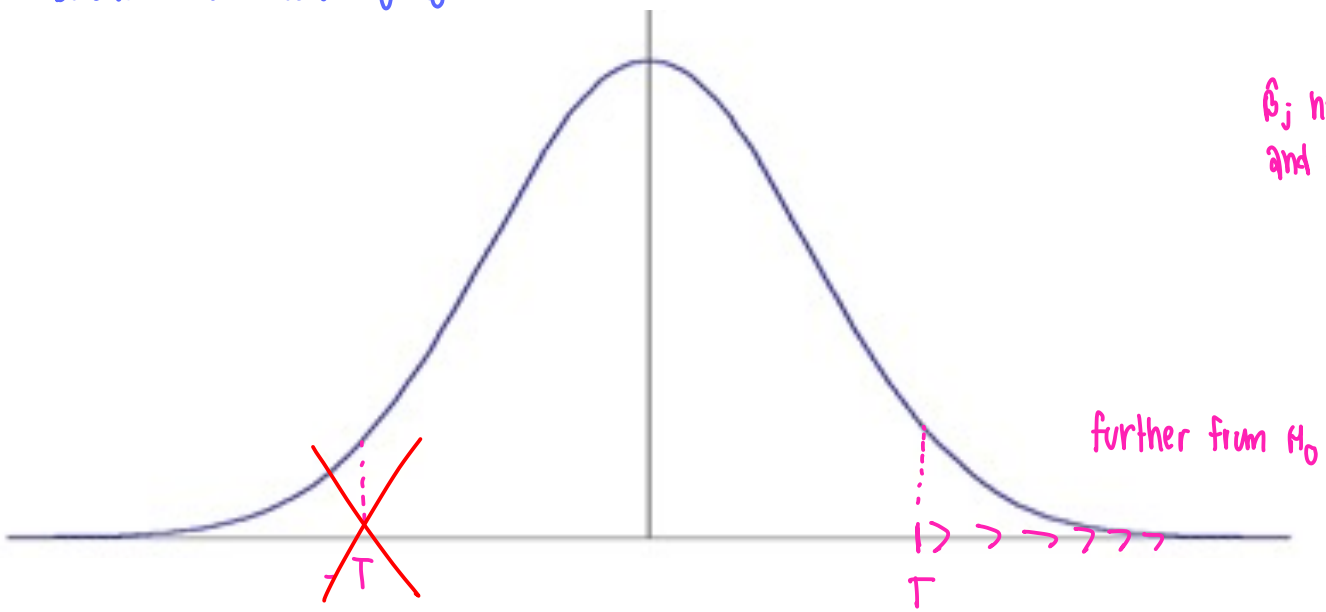
Alternative Hypothesis $H_1: \beta_j > b$ (0)

$H_0: \beta \leq b$; we count only positive

$H_0: \beta = 0$; can also use this (=) because we use b as 0 anyway

$$T_n(b, \hat{\beta}_j) = \frac{\hat{\beta}_j - b}{se(\hat{\beta}_j)}$$

$\hat{\beta}_j$ has to be positive and further from b



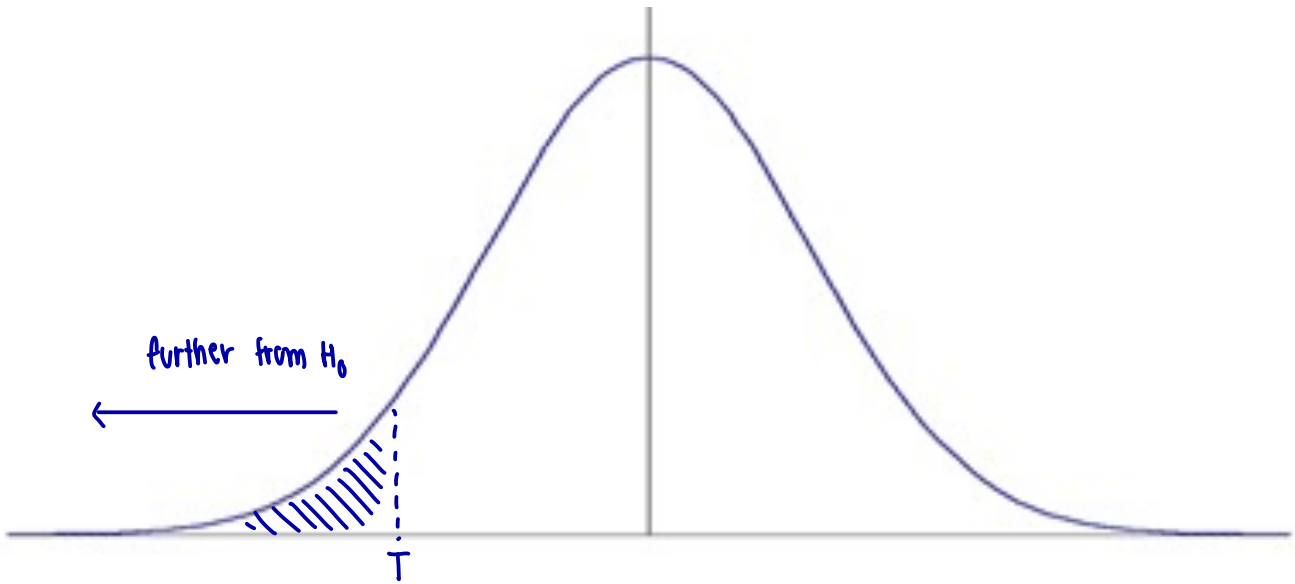
$\beta < 0$?

One-Sided Test

Null Hypothesis $H_0: \beta_j = b$

Alternative Hypothesis $H_1: \beta_j < b$

$$T_n(b, \hat{\beta}_j) = \frac{\hat{\beta}_j - b}{se(\hat{\beta}_j)}$$

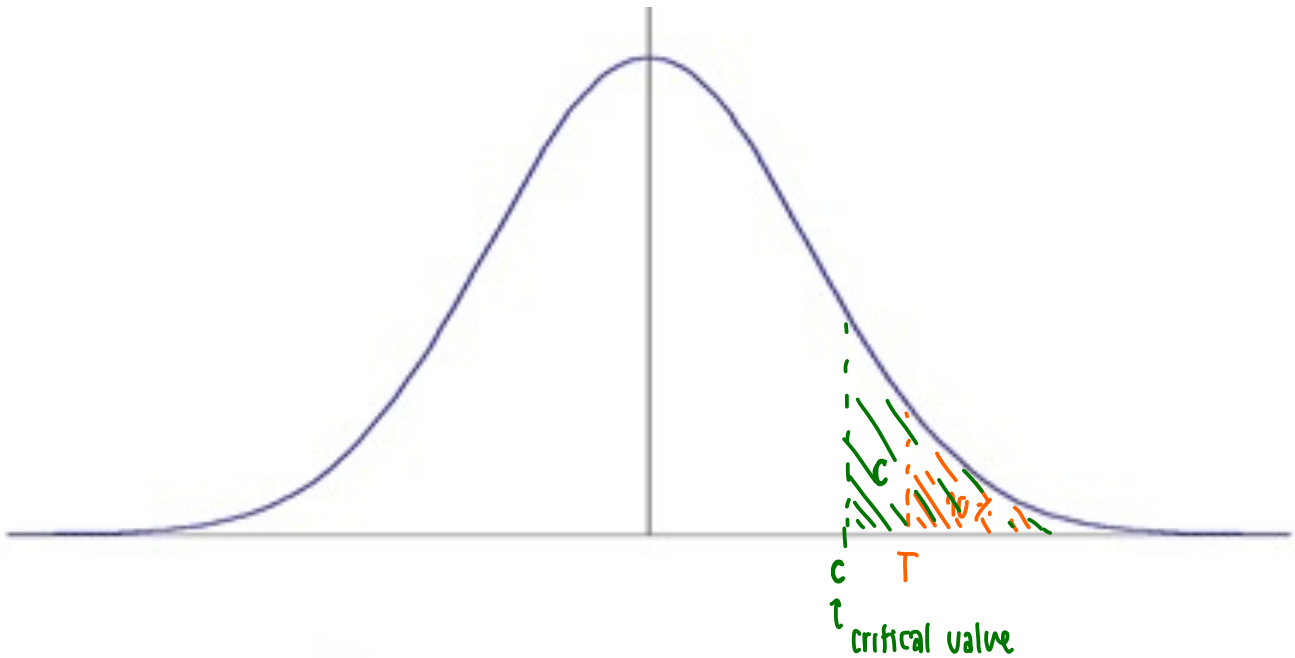


One-Sided Test

Null Hypothesis $H_0: \beta_j = b$

Alternative Hypothesis $H_1: \beta_j > b$

$$T = T_n(b, \hat{\beta}_j) = \frac{\hat{\beta}_j - b}{se(\hat{\beta}_j)}$$



give model
translate from statement to β ⋮ harder!!! give research question ⇒ write down model

difficult!

Hypothesis Statement
One- or Two-sided Test

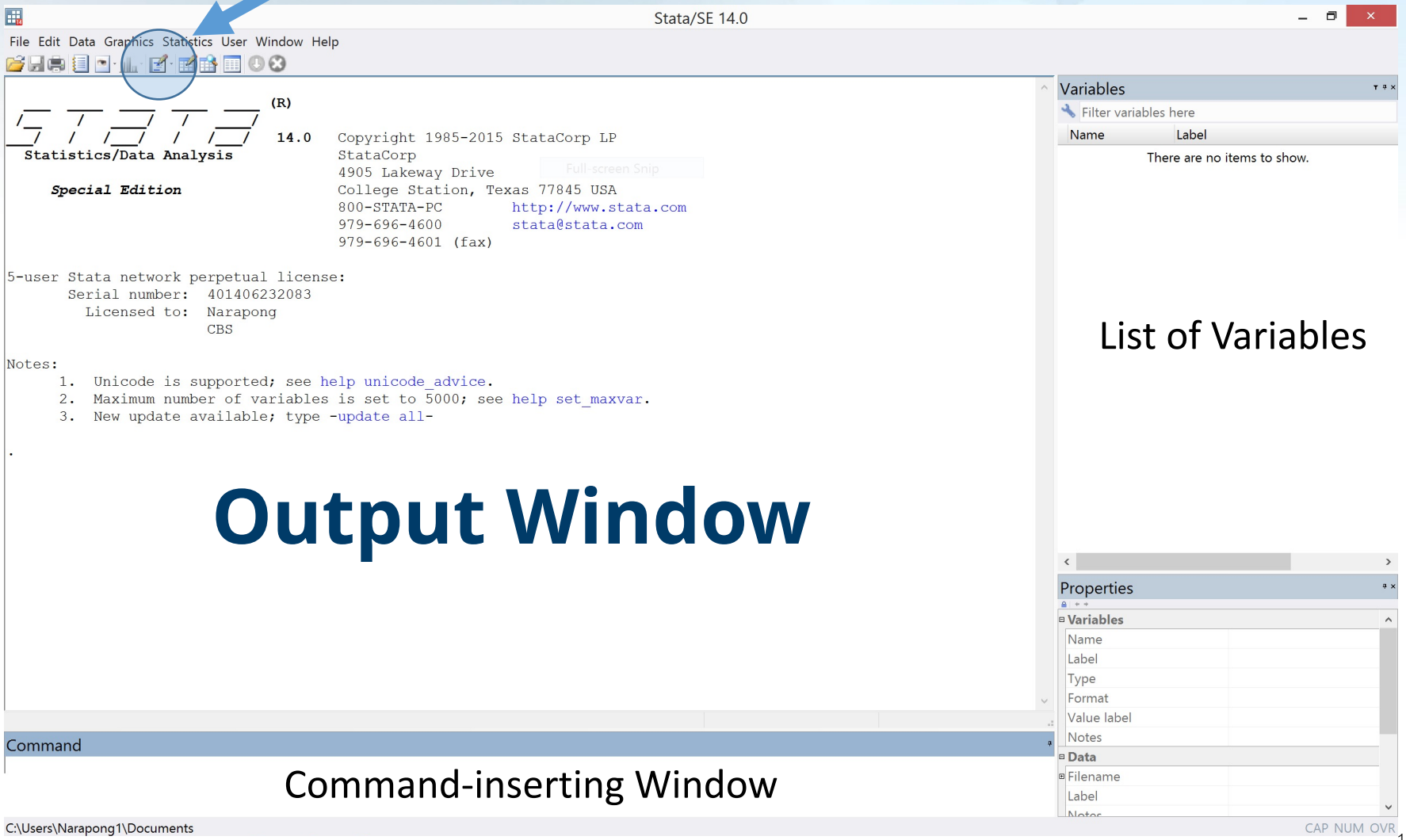
Test Statistics

P-Value or Critical Value

Conclusion of the Test

say also what it means by reject e.g. x has impact on y
conclusion in plain English

Open do-file editor window



List of Variables

Output Window

Command-inserting Window

Stata/SE 14.0

File Edit Data Graphics Statistics User Window Help

```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   14.0   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
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

5-user Stata network perpetual license:
       Serial number:  401406232083
         Licensed to:  Narapong
                       CBS

Notes:
      1.  Unicode is supported; see help unicode_advice.
      2.  Maximum number of variables is set to 5000; see help set_maxvar.
      3.  New update available; type -update all-

.
```

Command

C:\Users\Narapong1\Documents

Do-file Editor - finometricsBBA2019.do

File Edit View Project Tools

finometricsBBA2... × Untitled.do ×

```
clear all
set more off

* change directory
cd "C:\Users\Narapong1\Dropbox\Teaching\Econometrics\BBA 2019\lab

*******************************
******* Data Management *******
*******************************

* open file
use "wage.dta", clear

* summary statistics
summarize

* sort variables
sort educ female
gsort -educ female wage

* order variables in the data file
order educ female wage

* rename and label
rename female fem
label var fem "dummy for female"

* depict data
gr twoway scatter wage exper

* save graph
graph save "wage-exper.gph", replace
gr export "wage-exper.png", replace
erase "wage-exper.gph"
erase "wage-exper.png"

* save file
save "wage_new.dta", replace
erase "wage_new.dta"
```

Line: 1, Col: 1 CAP NUM OVR

CAP NUM OVR

do-file editor window

Estimating by OLS

```
. regress wage educ exper

      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(2, 523)       =     75.99
       Model |  1612.25452         2  806.127261   Prob > F        =    0.0000
    Residual |  5548.15979       523  10.6083361   R-squared       =    0.2252
-------------+----------------------------------   Adj R-squared   =    0.2222
       Total |  7160.41431       525  13.6388844   Root MSE        =     3.257

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .6442721   .0538061    11.97   0.000     .5385695    .7499747
       exper |   .0700954   .0109776     6.39   0.000     .0485297    .0916611
       _cons |   -3.39054   .7665661    -4.42   0.000    -4.896466   -1.884613
------------------------------------------------------------------------------
```

$\hat{\beta}$ $\quad$ $se(\hat{\beta})$ $\quad$ **t-stat** $\quad$ **p-value**

Estimating by OLS

```
. reg wage educ exper, nohead
------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .6442721   .0538061    11.97   0.000     .5385695    .7499747
       exper |   .0700954   .0109776     6.39   0.000     .0485297    .0916611
       _cons |   -3.39054   .7665661    -4.42   0.000    -4.896466   -1.884613
------------------------------------------------------------------------------


. reg ln(wage) educ exper
variable ln not found
r(111);
```

Generating a New Variable

0.216 0.097 0.0103

```
. gen lwage = ln(wage)
```

$\ln[wage] = \hat{\beta}_0 + \hat{\beta}_1 edu + \hat{\beta}_2 exper + u$

```
. reg lwage educ exper
```

degree of freedom = # of obs - # of beta
526 - 3

```
      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(2, 523)       =     86.86
       Model |  36.9850401         2  18.4925201   Prob > F        =    0.0000
    Residual |   111.34471       523  .212896195   R-squared       =    0.2493
-------------+----------------------------------   Adj R-squared   =    0.2465
       Total |   148.32975       525  .282532857   Root MSE        =    .46141

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  .0979356   .0076224    12.85   0.000     .0829613    .1129099
       exper |  .0103469   .0015551     6.65   0.000     .0072919     .013402
       _cons |  .2168544    .108595     2.00   0.046     .0035183    .4301904
------------------------------------------------------------------------------
```

coef. / SE

$H_0: \beta_i = 0$ for hypothesis

$\alpha = 5\%$

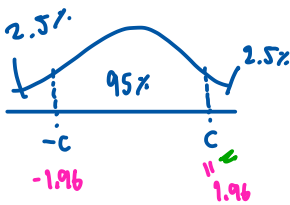
z-table

upper bound

β_1 β_2 β_0

0.216 - 0.1085 · c

0.216 + 0.1085 · c



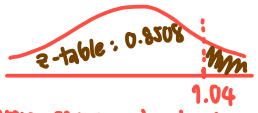
$\hat{\beta} \pm se(\hat{\beta}) \cdot$ critical value

only this change if α change

If educ ↑ 1yr → wage increase by 9%

Ex: $H_0: \beta_1 \le 0.09$ vs. $H_1: \beta_1 > 0.09$

$t\text{-stat} = \dfrac{\hat{\beta} - 0.09}{se(\hat{\beta})} \sim t_{523}$



df is large, can us standard normal

$= \dfrac{0.0979356 - 0.09}{0.0076224} = 1.04$

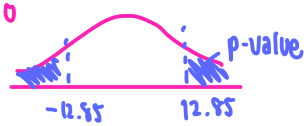
p-value = 1 - 0.8508 = 0.1492 = 14.92% > 10%, 5%, 1% ⇒ fail to reject H_0 because p-value > 10%

from z-table

∴ we have no enough statistical evidence to show that the impact of educ on wage is more than 9%

it is possible to be 9% or less than 9%

$H_0: \beta = 0$ vs. $\beta \neq 0$



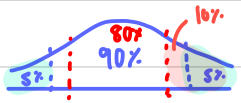
0.09 is in CI or not

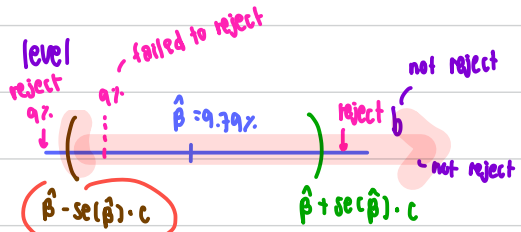
EX $H_0: \beta_1 \overset{<}{=} 0.09$ vs. $H_1: \beta_1 > 0.09$

use confidence interval to see if we can reject at 10% significance level

find CI : $\hat{\beta} \pm se(\hat{\beta}) \cdot c$

$0.0979356 \pm 0.0076224 \cdot c$





$H_1: \beta \neq$

$H_1: \beta > b$

when estimate is far above b

$0.0979356 - 0.0076224(1.282) = 0.088$

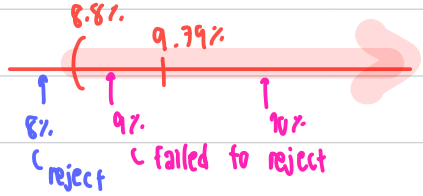
critical value 10% one-sided

want to test $\beta = 9\%$ vs. $\beta > 9\%$

8%

(8.8%, ∞)

90% Confidence interval for 1 sided test



* Construct CI whether impact on educ on wage > 10%

$H_0: \beta = b$ vs

$H_1: \beta > b$

90% confidence interval :

$Pr\{\beta \leq \text{interval}\} = 90\%$

$Pr\{t\text{-stat} > -c\} = 90\%$

$Pr\{t\text{-stat} < c\} = 90\%$

$\frac{(\hat{\beta} - \beta)}{se(\hat{\beta})} < c \rightarrow \hat{\beta} - \beta < c \cdot se(\hat{\beta})$

$\hat{\beta} - c \cdot se(\hat{\beta}) < \beta$

$\frac{(\hat{\beta} - \beta)}{se(\hat{\beta})} > -c$

$\hat{\beta} - \beta > se(\hat{\beta}) \cdot (-c)$

$\beta < \hat{\beta} + se(\hat{\beta}) \cdot c$

Multicollinearity

```
. gen male = 1-female

. reg wage educ exper male female, nohead

note: female omitted because of collinearity
------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .6025802   .0511174    11.79   0.000     .5021591    .7030012
       exper |   .0642417   .0104003     6.18   0.000     .0438101    .0846734
        male |   2.155517   .2703055     7.97   0.000     1.624497    2.686537
      female |          0  (omitted)
       _cons |  -3.889998   .7271441    -5.35   0.000    -5.318487    -2.46151
------------------------------------------------------------------------------
```

Outreg2: Professional Format

```
. ssc install outreg2
checking outreg2 consistency and verifying not already installed...
all files already exist and are up to date.

. reg wage educ exper

      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(2, 523)       =     75.99
       Model |  1612.25452         2  806.127261   Prob > F        =    0.0000
    Residual |  5548.15979       523  10.6083361   R-squared       =    0.2252
-------------+----------------------------------   Adj R-squared   =    0.2222
       Total |  7160.41431       525  13.6388844   Root MSE        =     3.257

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .6442721   .0538061    11.97   0.000     .5385695    .7499747
       exper |   .0700954   .0109776     6.39   0.000     .0485297    .0916611
       _cons |   -3.39054   .7665661    -4.42   0.000    -4.896466   -1.884613
------------------------------------------------------------------------------

. outreg2 using "wagereg.doc", replace ctitle(benchmark)
wagereg.doc
dir : seeout
```

Export to word document named “wagereg.doc”

Outreg2: Professional Format

VARIABLES	(1) benchmark
educ	0.644***
	(0.0538)
exper	0.0701***
	(0.0110)
Constant	-3.391***
	(0.767)
Observations	526
R-squared	0.225

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Outreg2: Professional Format

```
. reg wage educ exper, nohead
------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .6442721   .0538061    11.97   0.000     .5385695    .7499747
       exper |   .0700954   .0109776     6.39   0.000     .0485297    .0916611
       _cons |   -3.39054   .7665661    -4.42   0.000    -4.896466   -1.884613
------------------------------------------------------------------------------

. outreg2 using "wagereg.xls", replace adjr2
wagereg.xls
dir : seeout
.
. reg wage educ exper exper2, nohead
------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .5953429   .0530251    11.23   0.000     .4911741    .6995118
       exper |    .268287   .0368969     7.27   0.000     .1958023    .3407717
      exper2 |  -.0046123    .000822    -5.61   0.000     -.006227   -.0029975
       _cons |   -3.96489   .7521526    -5.27   0.000    -5.442508   -2.487272
------------------------------------------------------------------------------

. outreg2 using "wagereg.xls", append adjr2
wagereg.xls
dir : seeout
```

Reported adjusted R^2 instead

Export to excel file named "wagereg.xls"

Outreg2: Professional Format

	(1)	(2)
VARIABLES	wage	wage
educ	0.644***	0.595***
	(0.0538)	(0.0530)
exper	0.0701***	0.268***
	(0.0110)	(0.0369)
exper2		-0.00461***
		(0.000822)
Constant	-3.391***	-3.965***
	(0.767)	(0.752)
Observations	526	526
Adjusted R-squared	0.222	0.265

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Hypothesis Test for a Linear Combination of β's

- For example, consider the following regression model:

$$returns_{it} = \beta_0 + \beta_1 Q1_{it} + \beta_2 Q2_{it} + \beta_3 Q3_{it} + U_{it}$$

- Suppose we want to test whether, on average, stocks have positive returns in the first quarter, then

$$H_0: \beta_0 + \beta_1 = 0 \quad \textbf{vs} \quad H_1: \beta_0 + \beta_1 > 0$$

- Suppose we want to test whether, on average, returns of stock in quarter 1 and quarter 2 are equal, then

$$H_0: \beta_1 = \beta_2 \quad \textbf{vs} \quad H_1: \beta_1 \neq \beta_2$$

$$H_0: \beta_1 - \beta_2 = 0 \quad \textbf{vs} \quad H_1: \beta_1 - \beta_2 \neq 0$$

- Suppose now that we want to test $H_0: \beta_1 + \beta_2 = 0$ against the alternative hypothesis $H_1: \beta_1 + \beta_2 \neq 0$.
- Then, we can still use a **similar form of test statistics**:

$$\boldsymbol{T_n = \left| \frac{(\hat{\beta}_1 + \hat{\beta}_2) - 0}{se(\hat{\beta}_1 + \hat{\beta}_2)} \right| \sim t_{n-k-1}}$$

- Notice however that we need to find the standard error of $\hat{\beta}_1 + \hat{\beta}_2$
- Recall: $Var(\hat{\beta}_1 + \hat{\beta}_2) = Var(\hat{\beta}_1) + Var(\hat{\beta}_2) + 2cov(\hat{\beta}_1, \hat{\beta}_2)$. So, our standard error is

$$se(\hat{\beta}_1 + \hat{\beta}_2) = \left\{ se(\hat{\beta}_1)^2 + se(\hat{\beta}_2)^2 + 2s_{12} \right\}^{\frac{1}{2}}$$

where s_{12} is the estimated covariance derived from the estimated covariance matrix $\hat{\Sigma}$ of $\hat{\beta}$

- In general, we can test

$$H_0: r'\beta = b \text{ vs } H_1: r'\beta \neq b$$

where r is a vector of $k+1$ dimension, β is the vector of all the $k+1$ parameters, and b is a constant.

- Then, our test statistics is

$$T_n = \left|\frac{r'\hat{\beta} - b}{se\left(r'\hat{\beta}\right)}\right| = \left|\frac{r'\hat{\beta} - b}{\sqrt{r'\left(\frac{\hat{\Sigma}}{n}\right)r}}\right| \sim t_{n-k-1}$$

- Most of the statistical software has a command to compute the estimated covariance matrix $\hat{\Sigma}$ or $\hat{\Sigma}/n$, or automatically provide the covariance matrix as a result of the regression. But we have to compute $se\left(r'\hat{\beta}\right)$ from the covariance matrix.

Test for a Linear Combination of β_j

Find covariance matrix of the estimator

```
. reg lwage educ female feduc, nohead

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0772279   .0089875     8.59   0.000     .0595718    .0948841
      female |  -.3600645   .1854296    -1.94   0.053    -.7243444    .0042154
       feduc |  -.0000641   .0145035    -0.00   0.996    -.0285565    .0284283
       _cons |   .8259546   .1180502     7.00   0.000     .5940427    1.057867
------------------------------------------------------------------------------

. vce

Covariance matrix of coefficients of regress model

        e(V) |       educ      female       feduc       _cons
-------------+-------------------------------------------------
        educ |  .00008078
      female |  .00103298   .03438413
       feduc | -.00008078  -.00262903   .00021035
       _cons | -.00103298  -.01393586   .00103298   .01393586


.
```

Transformation Technique

2

- We can transform the regression such the interested linear combination of β's becomes a coefficient of a regressor in the transformed model.
- For example, starting from the linear regression model:

$$Y = \beta_0 + \beta_1 \overset{edu}{X_1} + \beta_2 \overset{f}{X_2} + \beta_3 \overset{f \cdot edu}{X_3} + U$$

- If we are interested in testing $\beta_1 + \beta_3 = 0.1$, we can transform the original model to

run regression on this

$$Y = \beta_0 + (\boldsymbol{\beta_1 + \beta_3})\overset{edu}{X_1} + \beta_2 \overset{f}{X_2} + \beta_3 (\overset{f \cdot edu - edu}{X_3 - X_1}) + U$$

- Then, estimating the regression of Y on X_1, X_2, and $X_3 - X_1$.
- The estimated coefficient of X_1 and its standard error will then be readily available for use.

$$\ln[wage] = \beta_0 + \beta_1 edu + \beta_2 f + \beta_3 f \cdot edu$$

effect of edu on wage

$$[\beta_1 + \beta_3 f]\, 100\%$$

$$\Downarrow$$

male : $\beta_1\, 100\%$

female : $(\beta_1 + \beta_3)\, 100\% \cong 10\%$

$H_0 : \beta_1 + \beta_3 = 0.1$

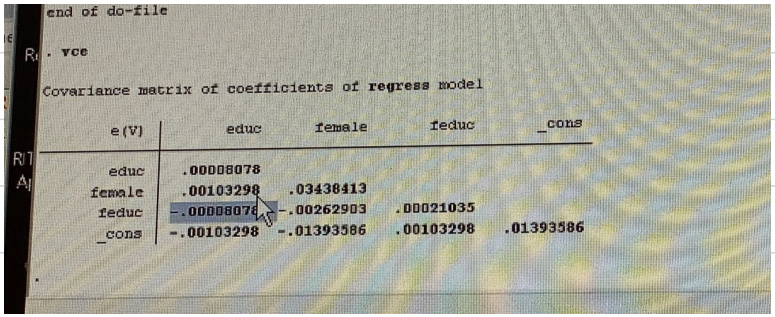
we have to find $se(\hat{\beta}_1 + \hat{\beta}_3)$

3 ways

① $cov(\hat{\beta}_1, \hat{\beta}_3)$

$$\Rightarrow se(\hat{\beta}_1 + \hat{\beta}_3) = \sqrt{se(\hat{\beta}_1)^2 + se(\hat{\beta}_2)^2 + 2cov(\hat{\beta}_1, \hat{\beta}_2)}$$

↳ write <u>vce</u> in the STATA



```
end of do-file

. vce

Covariance matrix of coefficients of regress model
```

e(V)	educ	female	feduc	_cons
educ	.00008078			
female	.00103298	.03438413		
feduc	-.00008078	-.00262903	.00021035	
_cons	-.00103298	-.01393586	.00103298	.01393586

② Transformation

Test for a Linear Combination of β_j

Employ transformation technique

```
. reg lwage educ female feduc, nohead
```

$\hat{\beta}_1$

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0772279	.0089875	8.59	0.000	.0595718	.0948841
female	-.3600645	.1854296	-1.94	0.053	-.7243444	.0042154
feduc	-.0000641	.0145035	-0.00	0.996	-.0285565	.0284283
_cons	.8259546	.1180502	7.00	0.000	.5940427	1.057867

$\hat{\beta}_3$

generate new

```
. gen xxx = feduc-educ
. reg lwage educ female xxx, nohead
```

get this

$se(\hat{\beta}_1 + \hat{\beta}_3)$!!

$\hat{\beta}_1 + \hat{\beta}_3$

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0771639	.0113831	6.78	0.000	.0548015	.0995262
female	-.3600645	.1854296	-1.94	0.053	-.7243444	.0042154
xxx	-.0000641	.0145035	-0.00	0.996	-.0285565	.0284283
_cons	.8259546	.1180502	7.00	0.000	.5940427	1.057867

Test for a Linear Combination of β_j

③ **Use STATA command "test"**

give F-stat

$H_0: \beta_1 + \beta_3 = 0.1$

$t\text{-stat} = \dfrac{\{0.0372239 - 0.0000641\} - 0.1}{se(\hat{\beta}_1 + \hat{\beta}_3)} = \ominus$

positive

Is the impact of education on wage for female equal to 10%?

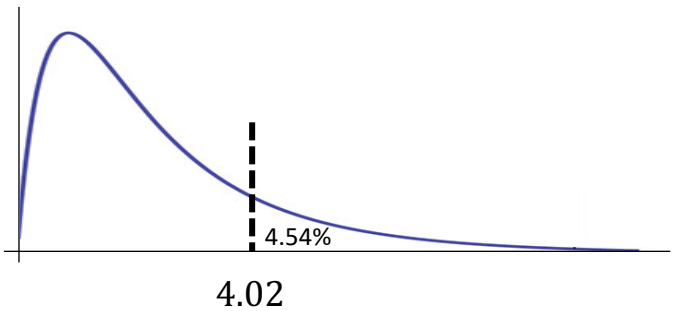
```
. test educ+feduc = 0.1

 ( 1)  educ + feduc = .1

       F(  1,   522) =    4.02
            Prob > F =    0.0454
```

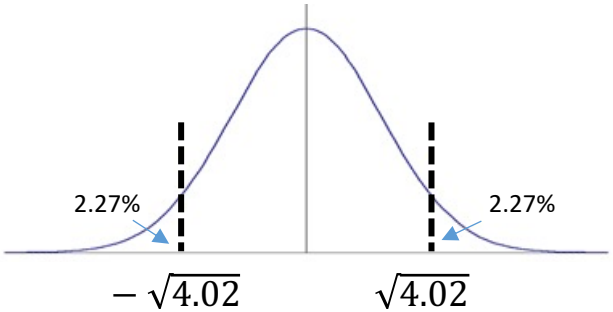
t-stat should be $-\sqrt{4.02}$

F-stat = $(t\text{-stat})^2$

$$F_{1,522} = 4.02 = (t_{522})^2$$

$$\Pr\{F_{1,522} > 4.02\} = 4.54\% = \Pr\{(t_{522})^2 > 4.02\}$$





Test for a Linear Combination of β_j

Use STATA command "`test`"

Is the impact of education on wage for female less than 10%?

```
. test educ+feduc = 0.1

 ( 1)  educ + feduc = .1

       F(  1,   522) =    4.02
            Prob > F =    0.0454
```

$t^2 \rightarrow$ t-stat $\begin{cases} \cancel{+\sqrt{4.02}} \\ -\sqrt{4.02} \end{cases}$

$$F_{1,522} = 4.02 = (t_{522})^2$$

(t-stat)^2

$$\Pr\{F_{1,522} > 4.02\} = 4.54\% = \Pr\{(t_{522})^2 > 4.02\}$$

p-value → don't have to do anything else

$H_1: \beta_1 + \beta_3 \neq 0.1 \rightarrow 4.54\%$

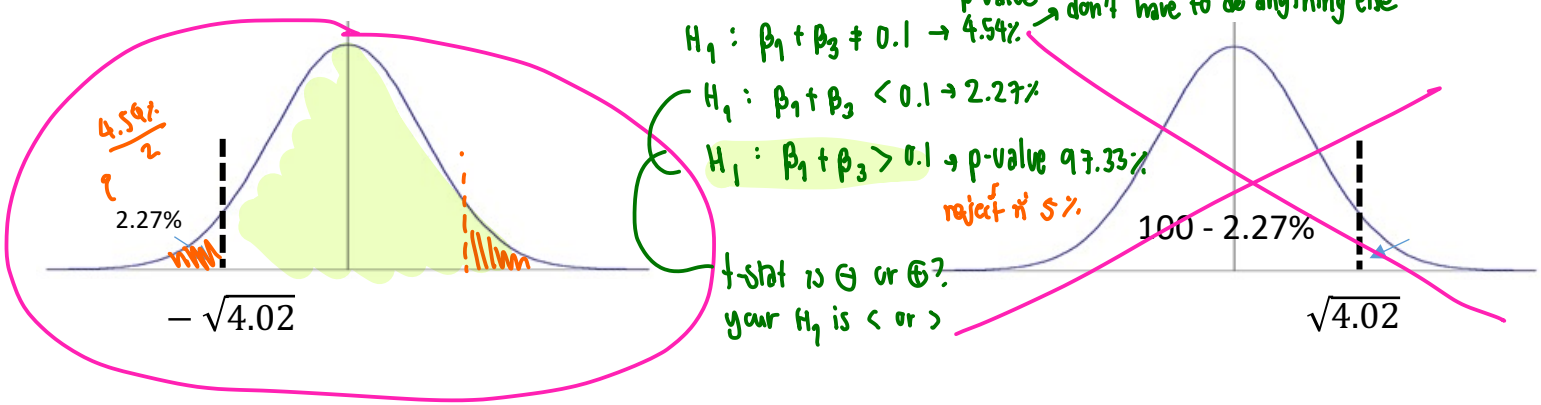
$H_1: \beta_1 + \beta_3 < 0.1 \rightarrow 2.27\%$

$H_1: \beta_1 + \beta_3 > 0.1 \rightarrow$ p-value 97.33%

reject if 5%

t-stat is ⊖ or ⊕?

your H_1 is < or >



$$wage = \beta_0 + \beta_1 edu + \beta_2 edu^2 + controls + u$$

edu has impact on wage ?

diff with respect to edu

$$\beta_1 + 2\beta_2 \cdot edu = 0$$

$H_0 : \beta_1 = 0$ and $\beta_2 = 0$ || cannot use t-stat

$H_1 : \beta_1 \neq 0$ or $\beta_2 \neq 0$ ↳ only F-stat

Hypothesis Test for Multiple Linear Restrictions

- Consider the following regression model

$$Y = \beta_0 + \beta_1 \boldsymbol{X_1} + \beta_2 \boldsymbol{X_1^2} + \beta_3 X_3 + U$$

- Suppose we want to test whether X_1 has an impact on Y.
- Since, the impact is $\beta_1 + 2\beta_2 X_1$, we have to test

$$\mathbf{H_0}\text{: } \boldsymbol{\beta_1 = 0 \text{ and } \beta_2 = 0}$$

$$\mathbf{H_1}\text{: } \boldsymbol{\beta_1 \neq 0 \text{ or } \beta_2 \neq 0}.$$

- In matrix notation, we can write the null hypothesis as

$$H_0\text{: } \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

which is simply a constant matrix multiplies with β

- In general, you can write the test as

$$H_0: \underbrace{A}_{d\times(k+1)}\underbrace{\beta}_{(k+1)\times 1} = \underbrace{\bar{b}}_{d\times 1} \text{ vs } H_1: A\beta \neq \bar{b}$$

- We know from CLT, that our OLS estimator $\hat{\beta}$ is asymptotically normal.

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N(0, \Sigma)$$

- Multiplying by a constant matrix A, and note that $A\beta = \bar{b}$ under the null hypothesis; so, we get

$$\sqrt{n}\left(A\hat{\beta} - A\beta\right) = \sqrt{n}\left(A\hat{\beta} - \bar{b}\right) \xrightarrow{d} N(0, A\Sigma A')$$

$$\left[\sqrt{n}\underbrace{\left(A\hat{\beta} - \bar{b}\right)}_{d\times 1}\right]' (A\Sigma A')^{-1} \left[\sqrt{n}\underbrace{\left(A\hat{\beta} - \bar{b}\right)}_{d\times 1}\right] \xrightarrow{d} \chi^2_d$$

$$\left(A\hat{\beta} - \bar{b}\right)' \left(A\left(\frac{\Sigma}{n}\right)A'\right)^{-1} \left(A\hat{\beta} - \bar{b}\right) \xrightarrow{d} \chi^2_d$$

- Again, since we don't know the parameter Σ, we will have to use the sample estimate $\hat{\Sigma}$ in place of Σ.
- The distribution of the test statistic will become

restriction/equations

don't have to remember

$$F = \left(A\hat{\beta} - \bar{b}\right)' \left(A\left(\frac{\hat{\Sigma}}{n}\right)A'\right)^{-1} \left(A\hat{\beta} - \bar{b}\right) \xrightarrow{d} F_{d,n-k-1}$$

obs - # beta in model not hypothesis

where d **is the number of restrictions** or equations we want to test.

- The test statistic F can also be computed by running regressions twice
  - **First, the unrestricted model: e.g.** $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + U$
  - **Second, the restricted model under the null: e.g.** $Y = \beta_0 + \beta_3 X_3 + U$

$$F = \frac{\left(\frac{RSS_r - RSS_{ur}}{d}\right)}{\left(\frac{RSS_{ur}}{n-k-1}\right)} = \frac{\left(\frac{R_{ur}^2 - R_r^2}{d}\right)}{\left(\frac{1 - R_{ur}^2}{n-k-1}\right)}$$

where the subscript r denotes restricted model, and subscript ur denotes unrestricted model

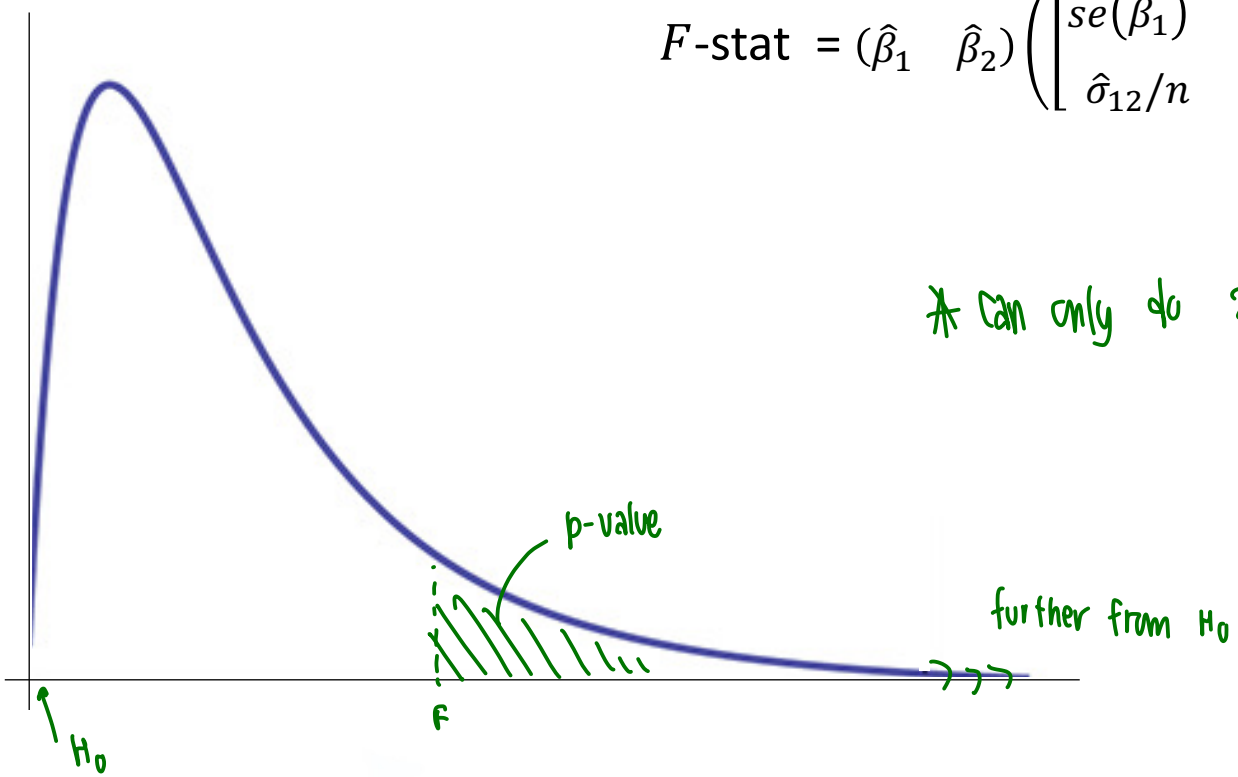
F-Test

Null Hypothesis $H_0: \beta_1 = 0$ and $\beta_2 = 0$

Alternative Hypothesis $H_1: \beta_1 \neq 0$ or $\beta_2 \neq 0$

$$F\text{-stat} = (\hat{\beta}_1 \quad \hat{\beta}_2)\left(\begin{bmatrix} se(\hat{\beta}_1)^2 & \hat{\sigma}_{12}/n \\ \hat{\sigma}_{12}/n & se(\hat{\beta}_2)^2 \end{bmatrix}\right)^{-1}\begin{pmatrix}\hat{\beta}_1 \\ \hat{\beta}_2\end{pmatrix}$$

* can only do 2-tail test
not 1-tail test



- Because we compute the *F*-statistics by squaring the estimates, the F-test is a two-sided test.
- We can also apply this test for the case of one restriction, i.e. the number of restriction $d = 1$, in this case the F-statistics is equal to the square of t-statistics.
- For $d = 1$, the F-test and two-sided t-test give the same p-value and conclusion, as

$$\Pr\left\{F_{1,n-k-1} > \left(\frac{\hat{\beta} - \beta}{se(\hat{\beta})}\right)^2\right\} = \Pr\left\{t_{n-k-1} > \left|\frac{\hat{\beta} - \beta}{se(\hat{\beta})}\right|\right\}$$

Test for Multiple Linear Restrictions

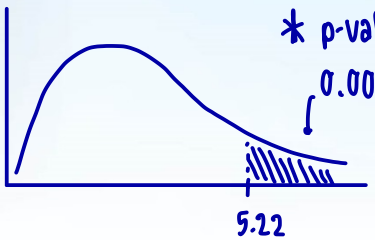
```
. gen feduc = female*educ
. reg wage educ female feduc

      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(3, 522)       =     61.07
       Model |  1860.2444         3  620.081466   Prob > F        =    0.0000
    Residual |  5300.16991       522  10.1535822   R-squared       =    0.2598
-------------+----------------------------------   Adj R-squared   =    0.2555
       Total |  7160.41431       525  13.6388844   Root MSE        =    3.1865

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |    .539476   .0642229     8.40   0.000     .4133089    .6656432
      female |  -1.198523    1.32504    -0.90   0.366    -3.801589    1.404543
       feduc |   -.085999   .1036388    -0.83   0.407    -.2895994    .1176014
       _cons |   .2004962   .8435616     0.24   0.812    -1.456696    1.857689
------------------------------------------------------------------------------

. test educ female feduc

 ( 1)  educ = 0
 ( 2)  female = 0
 ( 3)  feduc = 0

       F(  3,   522) =   61.07
            Prob > F =    0.0000
```

Test for Multiple Linear Restrictions

$H_0: \beta^{educ} = 0.1 \text{ and } \beta^{female*educ} = 0$

```
. test (educ=0.1) (feduc=0)

 ( 1)  educ = .1
 ( 2)  feduc = 0

       F(  2,   522) =   32.86
            Prob > F =    0.0000
```

df



* p-value
0.0057 < 1%
∴ reject H_0

$\ln wage = \beta_0 + \beta_1 edu + \beta_2 female + \beta_3 fem \cdot edu$

translate to simple statement!!!

$H_0: \beta^{educ} = 0.1 \text{ and } \beta^{educ} + \beta^{female*educ} = 0.1$

```
. test (educ=0.1) (educ+feduc=0.1)

 ( 1)  educ = .1
 ( 2)  educ + feduc = .1

       F(  2,   522) =   32.86
            Prob > F =    0.0000
```

Confidence Interval

- Confidence Interval analysis is a counterpart of the hypothesis testing.
- It is a **random interval that has the probability of covering the true parameter equal to $1-\alpha$**
- Thus, for β_j, we can construct the interval based on the same test statistics T_n:

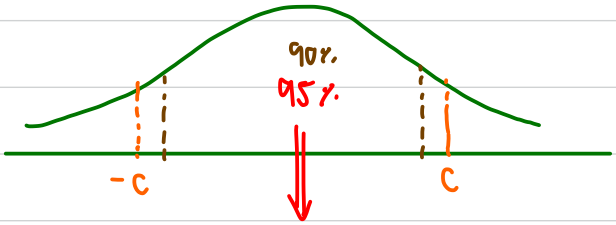
$$
\begin{aligned}
1-\alpha &= \Pr\left\{T_n = \left|\frac{\hat{\beta}-\beta}{se(\hat{\beta})}\right| < C_\alpha\right\} \\
&= \Pr\{|\hat{\beta}-\beta| < C_\alpha \cdot se(\hat{\beta})\} \\
&= \Pr\{-C_\alpha \cdot se(\hat{\beta}) < \hat{\beta}-\beta < C_\alpha \cdot se(\hat{\beta})\} \\
&= \Pr\{\hat{\beta}-C_\alpha \cdot se(\hat{\beta}) < \beta < \hat{\beta}+C_\alpha \cdot se(\hat{\beta})\}
\end{aligned}
$$

- **A confidence interval is $CI_{1-\alpha} = \left(\hat{\beta}_j - C_\alpha \cdot se(\hat{\beta}_j), \hat{\beta}_j + C_\alpha \cdot se(\hat{\beta}_j)\right)$**

95% Confidence Interval

$\Pr\{\beta \text{ in CI}\} = 95\%$

90% → interval become narrower



t-stat ~ $t_{\#\text{obs} - \#\text{beta}}$

$\Pr\{-c < \text{t-stat} < c\} = 95\%$

99%
90%
95% confidence = 1%, 10%, 5% significance level tail

Standard level

$(1-\alpha)$ Confidence interval

confidence level

$$\hat{\beta} \pm se(\hat{\beta}) \cdot c_{\alpha}$$

significance level

$$\Pr\left\{-c < \frac{\hat{\beta} - \beta}{se(\hat{\beta})} < c\right\} = 95\%$$

$-c \cdot se(\hat{\beta}) < \hat{\beta} - \beta$

$\beta < \hat{\beta} + c \cdot se(\hat{\beta})$

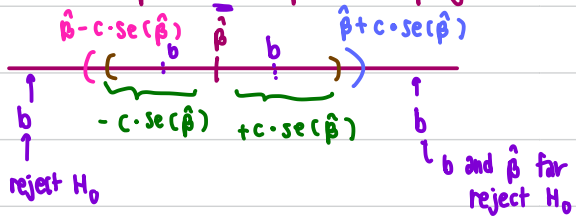
$\hat{\beta} - \beta < c \cdot se(\hat{\beta})$

$\hat{\beta} - c \cdot se(\hat{\beta}) < \beta$

$\Downarrow$

$$\Pr\{\hat{\beta} - c \cdot se(\hat{\beta}) < \beta < \hat{\beta} + c \cdot se(\hat{\beta})\} = 95\%$$

this should cover β at 95%



$H_0 : \beta = b$

$H_1 : \beta \neq b$

$H_0 : \beta_1 = b$

$H_1 : \beta_1 \neq b$

- Hypothesis test $\Leftrightarrow$ p-value $\overset{>\alpha}{<} \alpha$ $\Rightarrow$ reject H_0
- confidence interval $b \notin CI$ $\Rightarrow$ reject H_0

"Fail to reject H_0" / "cannot reject H_0" / "don't have enough evidence to reject H_0"

* b is not far to $\hat{\beta}_1$ $\Rightarrow$ it is not that $\hat{\beta}_1 = b$ $\Rightarrow$ cannot say "accept" !!!

random

CI :

$$\Pr\{\underbrace{\hat{\beta}}_{\text{random}} - c \cdot \underbrace{se(\hat{\beta})}_{\text{random}} < \underbrace{\beta}_{\text{fix}} < \underbrace{\hat{\beta}}_{\text{random}} + c \cdot \underbrace{se(\hat{\beta})}_{\text{random}}\} = 1 - \alpha$$

CI random depends on observation

e.g. $\hat{\beta} \pm c \cdot se(\hat{\beta}) \rightarrow (7, 11)$ $\boxed{\text{95\% of the times CI over } \beta}$ *

✗ the prob that β is in $(7, 11)$ is 95%

random

just 1 time

$$\ln wage = \beta_0 + \beta_1 edu + \beta_2 edu \cdot male + \dots$$

Q: The effect of edu on wage for male is 10%

diff by edu

$\beta_1 + \beta_2$ male

female → β_1

male → $\beta_1 + \beta_2$

$H_0: \beta_1 + \beta_2 = 0$ → can still use t-stat

$H_0: \beta_1 + \beta_2 = 0.1$

$$t\text{-stat} = \frac{\overset{\hat{\beta}_1 + \hat{\beta}_2}{\text{estimate}} - \overset{0.1}{\text{parameter}}}{se(\text{estimator})}$$

$se(\hat{\beta}_1 + \hat{\beta}_2)$

⇓

3-ways

from software

↳ $se(\hat{\beta}_1)$

$se(\hat{\beta}_2)$

Test for linear combination of β's

1) stata command
2) compute se ( ) manually
3) transformation

s.e. = estimate of S.D. $\sqrt{var}$

#2 Compute se( ) manually

$se(\hat{\beta}_1 + \hat{\beta}_2)$

$\Downarrow$

$\sqrt{\hat{var}(\hat{\beta}_1 + \hat{\beta}_2)}$

compute it later

get from STATA

$= \hat{var}(\hat{\beta}_1) + \hat{var}(\hat{\beta}_2) + 2\,\hat{cov}(\hat{\beta}_1, \hat{\beta}_2)$

$[se(\hat{\beta}_1)]^2$ $\quad$ $[se(\hat{\beta}_2)]^2$

#3 Transformation

pg. 33

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \boxed{\beta_3 x_3} + u$ → transform to have regressor as $(\beta_1 + \beta_2)$

select either x_1 or x_3 as your regressor with $(\beta_1 + \beta_2)$

to keep equal

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + [(\beta_1 + \beta_3) x_3 - \beta_3 x_3] + u$

1st regressor, 2nd, 3rd → use it in STATA

$y = \beta_0 + \beta_1 [x_1 - x_3] + \beta_2 x_2 + (\beta_1 + \beta_3) x_3 + u$

do math first then STATA

Transformation technique :

can be written as $2\beta_1\left(\frac{x_1}{2}\right)$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

coefficient of one regressor

1) $H_0 : 2\beta_1 + \beta_2 = 0$

$$y = \beta_0 + \beta_1 x_1 + (2\beta_1 + \beta_2) x_2 - 2\beta_1 x_2 + \beta_3 x_3 + u$$

$$y = \beta_0 + \beta_1 (x_1 - 2x_2) + (2\beta_1 + \beta_2) x_2 + \beta_3 x_3 + u$$

2) $H_0 : \boxed{2\beta_1 + 3\beta_2} = 0 \iff \beta_1 + \frac{3}{2}\beta_2 = 0$ or $\frac{2}{3}\beta_1 + \beta_2 = 0$

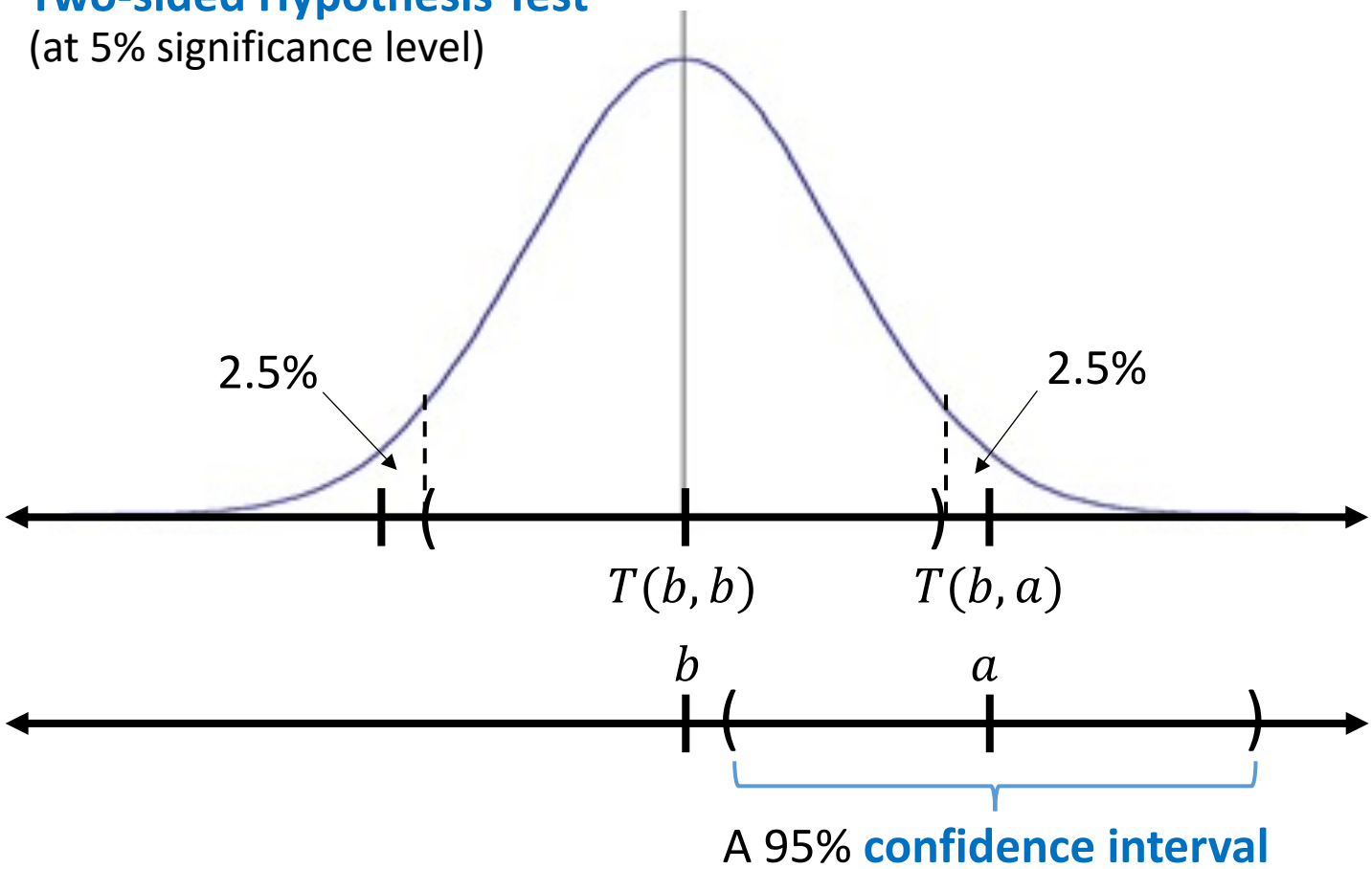
$$y = \beta_0 + \left(\beta_1 + \frac{3}{2}\beta_2\right) x_1 - \frac{3}{2}\beta_2 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

$$y = \beta_0 + \underbrace{\left(\beta_1 + \frac{3}{2}\beta_2\right) x_1}_{(2\beta_1 + 3\beta_2)\frac{x_1}{2}} + \beta_2 \left(x_2 - \frac{3}{2}x_1\right) + \beta_3 x_3 + u$$

3) $H_0 : \beta_1 + \beta_2 + \beta_3 = 0$

$$y = \beta_0 + (\beta_1 + \beta_2 + \beta_3) x_1$$

Two-sided Hypothesis Test
(at 5% significance level)

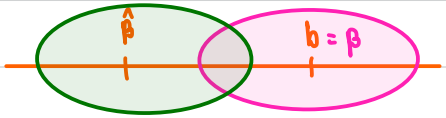


```
. reg lwage educ exper exper2

      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(  3,   522) =   74.67
       Model |  44.5393718     3  14.8464573           Prob > F      =  0.0000
    Residual |  103.790378   522  .198832142           R-squared     =  0.3003
-------------+------------------------------           Adj R-squared =  0.2963
       Total |   148.32975   525  .282532857           Root MSE      =  .44591

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0903658    .007468    12.10   0.000     .0756948    .1050368
       exper |   .0410089   .0051965     7.89   0.000     .0308002    .0512175
      exper2 |  -.0007136   .0001158    -6.16   0.000     -.000941   -.0004861
       _cons |   .1279975   .1059323     1.21   0.227    -.0801085    .3361035
------------------------------------------------------------------------------
```

Review:

Sample $\hat{\beta}$ $\xrightarrow{\text{Inference}}$ Population $\beta = b$ ↗ hypothesis

is $\hat{\beta}$ closed to b?



$(1-\alpha)$ Confidence interval

$\hat{\beta} \pm se(\hat{\beta}) \cdot c_\alpha$

└ significant level

Hypothesis

b is the actual data of population

Hypothesis test

1) Define H_0, H_1

2) Find test statistics
- function of both $\hat{\beta}$ and β
- we have to know its distribution

→ 1 equation to test → use t-test

→ # in model not hypothesis

t-statistics

$t_{\#obs - \#beta}$

↳ More than 1 equation → use F-test

$\sim F_{\alpha, \#obs - \#betas}$

3) Use test statistics to find p-value

↗ low → far from hypothesis → reject

use significance level to find critical value (c_α)

4) Reject H_0 if p-value < 10%, 5%, 1% (significant level)

if test stat further than c_α

* use STATA

$H_0: \beta = b$

t - stat

$$\frac{\overset{\hat{\beta}}{\text{estimate}} - \overset{b}{\text{parameter}}}{\underset{se(\hat{\beta})}{se(\text{estimator})}}$$

More complicate:

$H_0: \beta_1 + \beta_2 = b$

$$\frac{\overset{\hat{\beta}_1 + \hat{\beta}_2}{\text{estimate}} - \overset{b}{\text{parameter}}}{se(\text{estimator})}$$

$se(\hat{\beta}_1 + \hat{\beta}_2)$

STATA cannot automatically find this

need more work to do

1) find $cov(\hat{\beta}_1, \hat{\beta}_2) \rightarrow se(\hat{\beta}_1 + \hat{\beta}_2)$

2) transform model to have $\beta_1 + \beta_2$ as coefficient

3) use STATA test command

Goodness of fit

① SER → lower = better fit

↳ it has unit the same as Y, if you have

diff in unit

e.g. $\ln y = \beta x$
$y = \beta x$) cannot be compared

② R^2 ↳ [0,1] : drawback: always tend to recommend more x, even x has low explanatory power to y

higher → better fit

③ $\bar{R}^2$

↳ most popular

$(-\infty, 1]$

* First, try to use logic / theories to determine whether it make sense.

Rescale

STATA session:

- open wage.dta

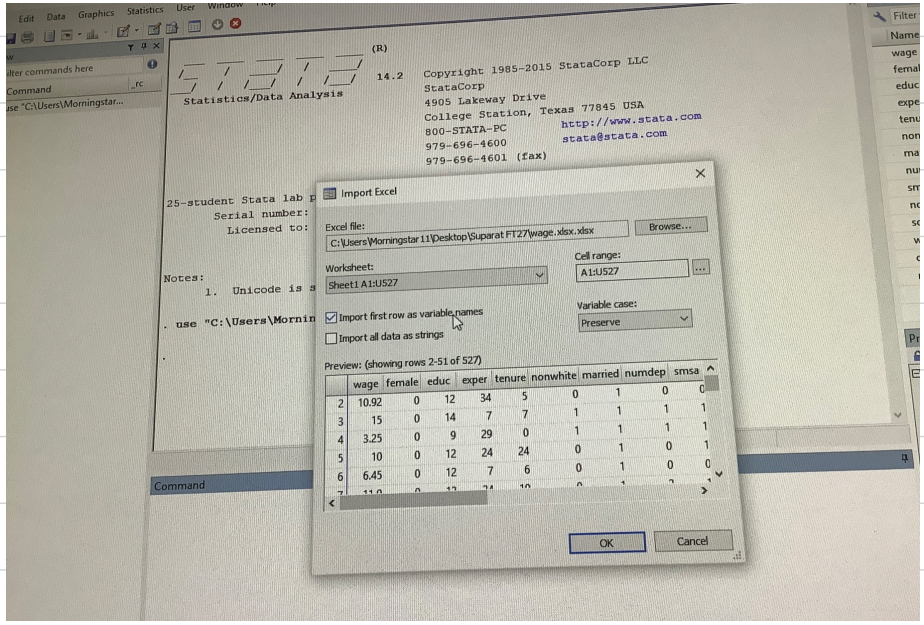
Excel

↳ first row is name of Y and X (random variable)

file : import excel spreadsheet

strings → means word/text ← don't click this

* !!! format of r.v. should not be text → ไม่งั้น STATA จะ: treat ทุกตัวเป็น text → cannot do calculation



* don't do editor edit!!

* If you want to save code → use "do·find editor"

summarize

↳ command to get obs, mean, sd, min, max of each variable

/* *\ text

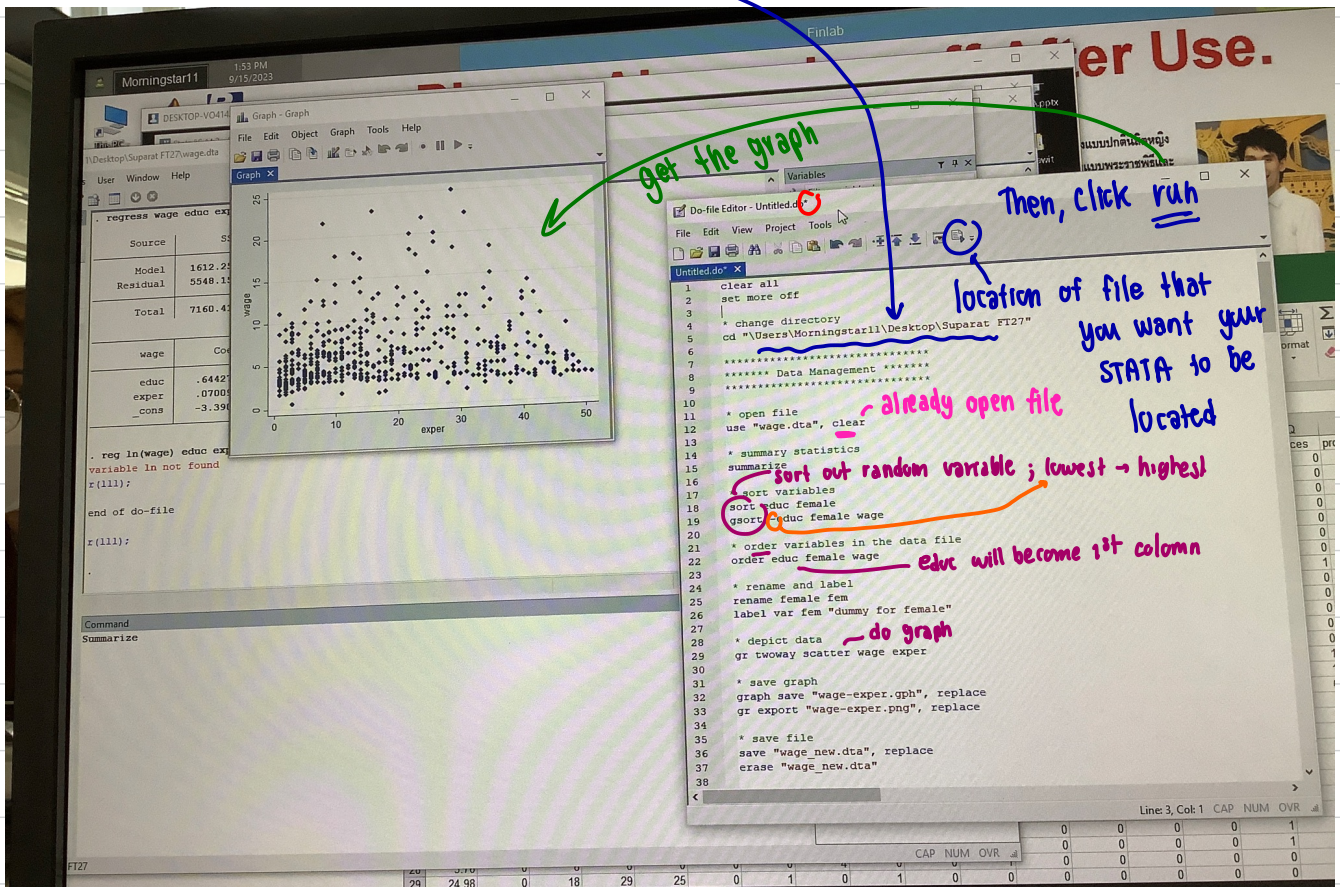
Command blue is command

---

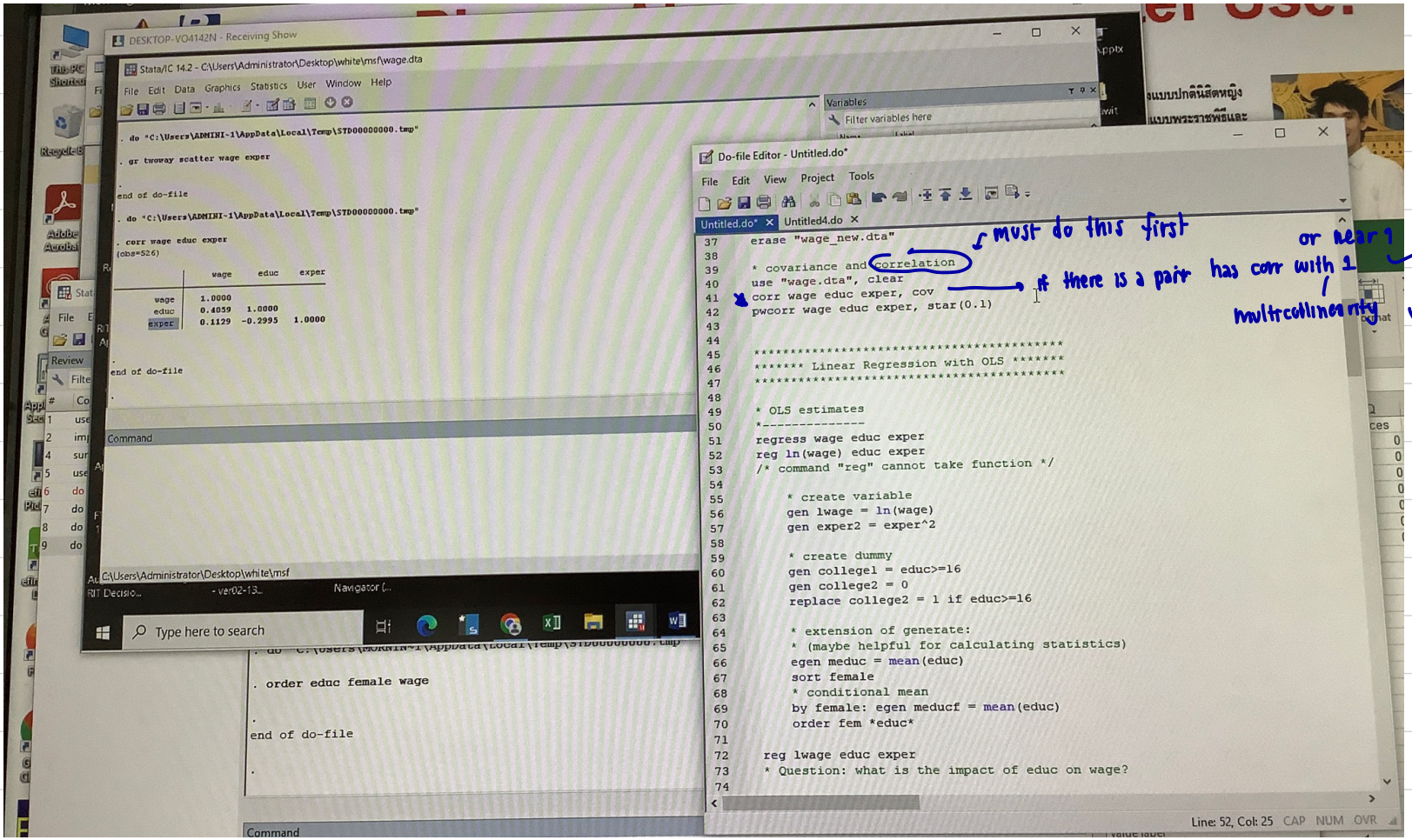
* copy text from Finometrics FT27.do to the " do-file editor"

then change

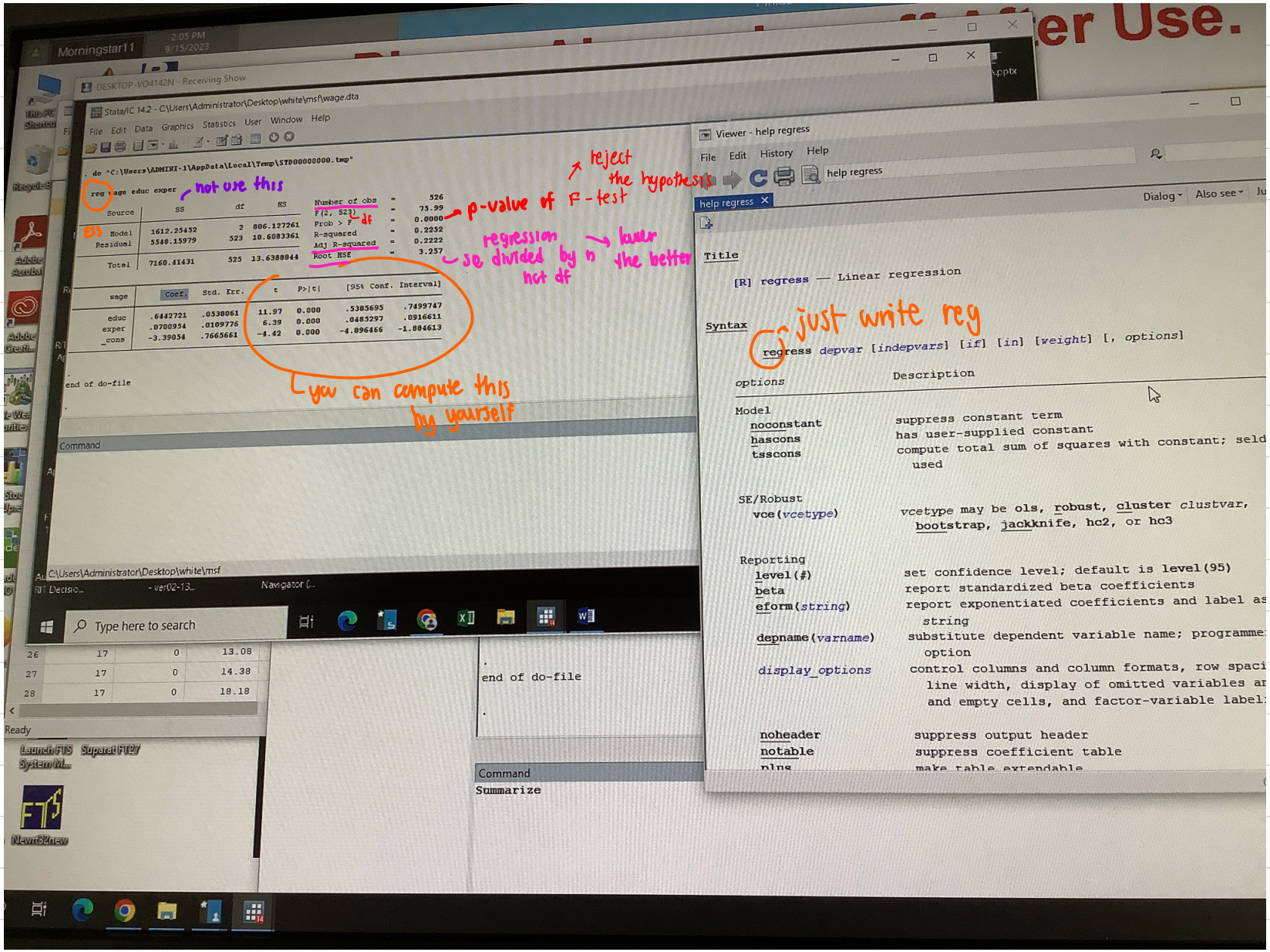
* means you haven't save it



Use graphics command can create types of graph



Help → search ...



* cannot put function in reg command!

need to create random variable by use gen command

Do-file Editor - Untitled.do*

File Edit View Project Tools

Untitled.do* | Untitled4.do

```
**********************************************

* OLS estimates
*--------------
reg wage educ exper
reg ln(wage) educ exper
/* command "reg" cannot take function */

    * create variable
    gen lwage = ln(wage)
    gen exper2 = exper^2

    * create dummy
    gen college1 = educ>=16
    gen college2 = 0
    replace college2 = 1 if educ>=16

    * extension of generate:
    * (maybe helpful for calculating statistics)
    egen meduc = mean(educ)
    sort female
    * conditional mean
    by female: egen meducf = mean(educ)
    order fem *educ*

reg lwage educ exper
* Question: what is the impact of educ on wage?

reg lwage educ exper exper2

* Dummy variable trap
gen male = 1-female
reg wage educ exper male female
reg wage educ exper male
reg wage educ exper female
reg wage educ exper male female, nocon
```

to see the diff effect btw group

exactly the same

just diff way to write code

ดูจาก Help ว่าใช้กับอะไรได้บ้าง

9.79%

Line: 63, Col: 1 CAP NUM OVR

Data

Line: 63, Col: 37 CAP NUM OVR

ENG 2:11 PM 9/15/2023

Finlab

Do-file Editor - FinometricsFX27.do*

```
66      * (maybe helpful for calculating statistics)
67      egen meduc = mean(educ)
68      sort female
69      * conditional mean
```

Do-file Editor - Untitled.do*

```
47    ****************************************
48
49    * OLS estimates
50    *--------------
51    reg wage educ exper
52    reg ln(wage) educ exper
53    /* command "reg" cannot take function */
54
55        * create variable
56        gen lwage = ln(wage)
57        gen exper2 = exper^2
58
59        * create dummy
60        gen college1 = educ>=16
61        gen college2 = 0
62        replace college2 = 1 if educ>=16
63
64        * extension of generate:
65        * (maybe helpful for calculating statistics)
66        egen meduc = mean(educ)
67        sort female
68        * conditional mean
69        by female: egen meducf = mean(educ)
70        order fem *educ*
71
72    reg lwage educ exper
73    * Question: what is the impact of educ on wage?
74
75    reg lwage educ exper exper2
76
77    * Dummy variable trap
78    gen male = 1-female
79    reg wage educ exper male female
80    reg wage educ exper male
81    reg wage educ exper female
82    reg wage educ exper male female, nocon
83
84
```

Line: 63, Col: 1

```
.
end of do-file

. do "C:\Users\MORNIN~1\AppData\Local\Temp\STD00000000.tmp"

. reg lwage educ exper

      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(2, 523)       =     86.86
       Model |  36.9850401         2  18.4925201   Prob > F        =    0.0000
    Residual |   111.34471       523  .212896195   R-squared       =    0.2493
-------------+----------------------------------   Adj R-squared   =    0.2465
       Total |   148.32975       525  .282532857   Root MSE        =    .46141

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0979356   .0076224    12.85   0.000     .0829613    .1129099
       exper |   .0103469   .0015551     6.65   0.000     .0072919     .013402
       _cons |   .2168544    .108595     2.00   0.046     .0035183    .4301904
------------------------------------------------------------------------------

.
end of do-file

.
```

9.79%

because of ln

Command

Summarize

Stata/IC 14.2 - C:\Users\Administrator\Desktop\white\msf\wage.dta

```
. reg wage educ exper male female
note: female omitted because of collinearity
```

Stata/IC 14.2 - C:\Users\Morningstar11\Desktop\Suparat FT27\wage.dta

```
. do "C:\Users\MORNIN~1\AppData\Local\Temp\STD00000000.tmp"

. gen male = 1-female

. reg wage educ exper male female
note: female omitted because of collinearity

      Source |       SS           df       MS      Number of obs   =       526
-------------+----------------------------------   F(3, 522)       =     77.92
       Model |  2214.74207         3  738.247358   Prob > F        =    0.0000
    Residual |  4945.67224       522  9.47446789   R-squared       =    0.3093
-------------+----------------------------------   Adj R-squared   =    0.3053
       Total |  7160.41431       525  13.6388844   Root MSE        =    3.0781

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .6025802   .0511174    11.79   0.000     .5021591    .7030012
       exper |   .0642417   .0104003     6.18   0.000     .0438101    .0846734
        male |   2.155517   .2703055     7.97   0.000     1.624497    2.686537
      female |          0  (omitted)
       _cons |  -3.889998   .7271441    -5.35   0.000    -5.318487   -2.46151
------------------------------------------------------------------------------

. reg wage educ exper male
```

drop this!!!

Command

```
Summarize
```

Do-file Editor - Untitled.do*

Untitled.do* | Untitled4.do

```
gen lwage = ln(wage)
gen exper2 = exper^2

* create dummy
gen college1 = educ>=16
gen college2 = 0
replace college2 = 1 if educ>=16

* extension of generate:
* (maybe helpful for calculating statistics)
egen meduc = mean(educ)
sort female
* conditional mean
by female: egen meducf = mean(educ)
order fem *educ*

reg lwage educ exper
* Question: what is the impact of educ on wage?

reg lwage educ exper exper2

* Dummy variable trap
gen male = 1-female
reg wage educ exper male female
reg wage educ exper male
reg wage educ exper female
reg wage educ exper male female, nocon

* outreg2
*--------
*ssc install outreg2

reg wage educ exper
outreg2 using "wagereg.doc", replace ctitle(benchmark)

reg wage educ exper exper2
outreg2 using "wagereg.doc", append ctitle(non-linear) ///
```

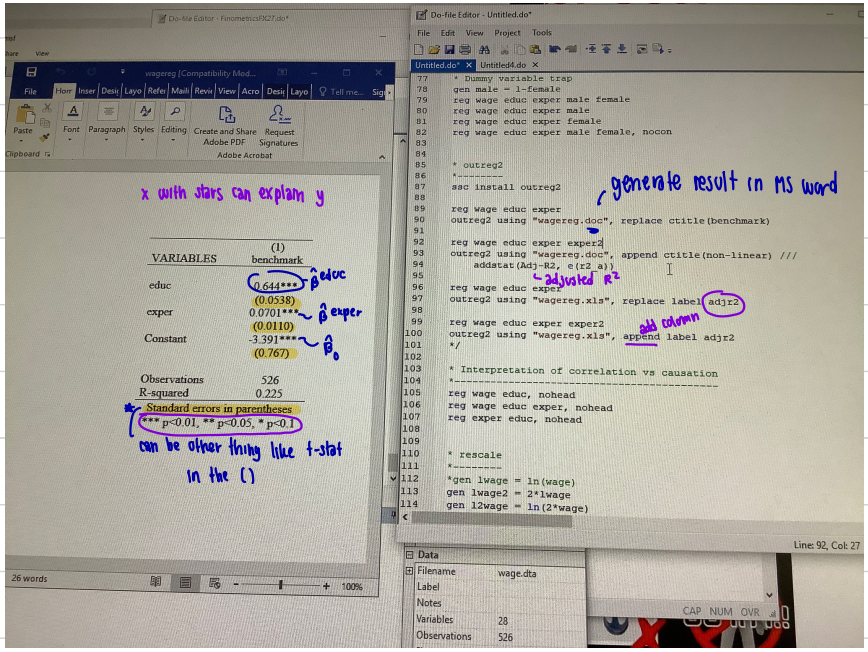
don't estimate constant

Line: 82 Col: 39

add-on!!!

Outreg 2 : professional way to do regression

reinstall it by ssc install outreg 2



age $\geq$ 16 → college

```
/* Exercise:
*-----------
run a regression which allows impact of educ on wage
to be different between those who have and don't have a college degree.

1. what is the estimated impact of educ on wage for non-college graduates?
2. what is the estimated impact of educ on wage for college graduates?
3. is there an evidence that the impacts for the two groups are different?
*/
```

edu ⇒ wage

create dummy variable

grad | non-grad

college: 1, non-college: 0

Model construction

$$\ln[wage] = \beta_0 + \beta_1 edu + \beta_2 college + \beta_3 edu \cdot college$$

diff by edu : $\beta_1 + \beta_3 college$ ⇒ impact

① non-college : β_1

② college : $\beta_1 + \beta_3$

③ diff btw non- and college β_3

don't forget to add control

$H_0: \beta_3 = 0$

$H_1: \beta_3 \neq 0$

Hypothesis test

1 equation : t-test
2 or more equations : F-test

Rescaling Variables

```
. gen lwage2 = 2*lwage
. gen l2wage = ln(2*wage)
. gen educ2 = 2*educ


. quietly reg lwage educ exper exper2
. outreg2 using "wagereg.xls", replace ctitle(lwage) adjr2
wagereg.xls
dir : seeout

. quietly reg lwage educ2 exper exper2
. outreg2 using "wagereg.xls", append ctitle(2*educ) adjr2
wagereg.xls
dir : seeout

. quietly reg lwage2 educ exper exper2
. outreg2 using "wagereg.xls", append ctitle(2ln(wage)) adjr2
wagereg.xls
dir : seeout

. quietly reg l2wage educ exper exper2
. outreg2 using "wagereg.xls", append ctitle(ln(2wage)) adjr2
wagereg.xls
dir : seeout
```

Rescaling Variables

	(1)	(2)	(3)	(4)
VARIABLES	lwage	2*educ	2ln(wage)	ln(2wage)
educ	0.0904***		0.181***	0.0904***
	(0.00747)		(0.0149)	(0.00747)
exper	0.0410***	0.0410***	0.0820***	0.0410***
	(0.00520)	(0.00520)	(0.0104)	(0.00520)
exper2	-0.000714***	-0.000714***	-0.00143***	-0.000714***
	(0.000116)	(0.000116)	(0.000232)	(0.000116)
educ2		0.0452***		
		(0.00373)		
Constant	0.128	0.128	0.256	0.821***
	(0.106)	(0.106)	(0.212)	(0.106)
Observations	526	526	526	526
Adjusted R-squared	0.296	0.296	0.296	0.296
Standard errors in parentheses				
*** p<0.01, ** p<0.05, * p<0.1				

Rescaling Variables

Consider the multivariate model

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + U$$

Rescaling the regressand to $\boldsymbol{cY}$:

- Since the LHS of the equation changes, to keep the equation satisfy, we need to multiply c to the RHS as well:

$$\boldsymbol{c}Y = \boldsymbol{c\beta_0} + \boldsymbol{c\beta_1} X_1 + \cdots + \boldsymbol{c\beta_k} X_k + cU$$

- Each coefficient has to be rescaled by multiplying c as well.
- The OLS estimator will also gives the same estimates multiplied by the rescaling factor c, i.e.

$$\hat{\beta}_j \Rightarrow c\hat{\beta}_j \text{ and } se(\hat{\beta}_j) \Rightarrow se(c\hat{\beta}_j) = |c| se(\hat{\beta}_j)$$

Rescaling a regressor X_1 to cX_1:

- The LHS of the equation is not change.
- So, we need to keep the RHS the same, the coefficient has to be rescaled by dividing by c:

$$Y = \beta_0 + \left(\frac{\boldsymbol{\beta_1}}{\boldsymbol{c}}\right)(\boldsymbol{c}X_1) + \cdots + \beta_k X_k + U$$

- The OLS estimator will also gives the same estimates divided by the rescaling factor c, i.e.

$$\hat{\beta}_1 \Rightarrow \frac{\widehat{\beta}_1}{c} \text{ and } se(\hat{\beta}_1) \Rightarrow se\left(\frac{\widehat{\beta}_1}{c}\right) = \left|\frac{1}{c}\right| se(\hat{\beta}_1)$$

Rescaling does not affect goodness of fit. Both R^2 and $\bar{R}^2$ will remain the same.