Applied Statistics and Econometrics

Lecture 5

Saul Lach

September 2017

Outline of Lecture 5

- Now that we know the sampling distribution of the OLS estimator, we are ready
  1. to perform hypothesis tests about β_1 (SW 5.1)
  2. to construct confidence intervals about β_1 (SW 5.2)

     In addition, we will cover some loose ends about regression:
  3. Empirical application (Italian LFS)
  4. Regression when X is binary (0/1) (SW 5.3)
  5. Heteroskedasticity and homoskedasticity (SW 5.4)
  6. The theoretical foundations of OLS – Gauss-Markov Theorem (SW 5.5)

Testing hypotheses (SW 5.1)

- Suppose a skeptic suggests that reducing the number of students in a class has no effect on learning or, specifically, on test scores.
- Recall the model $Testscore = \beta_0 + \beta_1 STR + u$. The skeptic thus asserts the hypothesis,

$$H_0 : \beta_1 = 0$$

- We wish to test this hypothesis using data, and reach a tentative conclusion whether it is correct or incorrect.

Null and alternative hypotheses

- The null hypothesis and **two-sided alternative**:

$$H_0 : \beta_1 = \beta_{1,0} \qquad vs. \qquad H_1 : \beta_1 \neq \beta_{1,0}$$

  where $\beta_{1,0}$ is the hypothesized value of β_1 under the null.
- Null hypothesis and **one-sided alternative**:

$$H_0 : \beta_1 = \beta_{1,0} \qquad vs. \qquad H_1 : \beta_1 < (>) \beta_{1,0}$$

- In Economics, it is almost always possible to come up with stories in which an effect could "go either way," so it is standard to focus on two-sided alternatives.

General approach to testing

- In general, the t-statistic has the form

$$t = \frac{estimator - hypothesized\ value\ under\ H_0}{s\tan dard\ error\ of\ estimator}$$

- For testing the mean of Y this simplifies to

$$t = \frac{\bar{Y} - \mu_{Y,0}}{s_Y / \sqrt{n}}$$

- And for testing β_1 this becomes

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)}$$

  where $SE(\hat{\beta}_1)$ is the square root of an estimator of the variance of the sampling distribution of $\hat{\beta}_1$.

Formula for SE(betahat1)

- Recall the expression for the variance of $\hat{\beta}_1$ (large n):

$$\sigma^2_{\hat{\beta}_1} = \mathrm{Var}(\hat{\beta}_1) = \frac{1}{n} \times \frac{Var((X_i - \mu_X)u_i)}{(\sigma^2_X)^2}$$

- The **estimator** of the variance of $\hat{\beta}_1$ replaces the unknown population values by estimators constructed from the data:

$$\begin{aligned} \hat{\sigma}^2_{\hat{\beta}_1} &= \widehat{\mathrm{Var}}(\hat{\beta}_1) = \frac{1}{n} \times \frac{\text{estimator of } Var((X_i - \mu_X)u_i)}{(\text{estimator of } \sigma^2_X)^2} \\ &= \frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^n ((X_i - \bar{X})^2 \hat{u}_i^2)}{\left[\frac{1}{n}\sum_i^n (X_i - \bar{X})^2\right]^2} \end{aligned}$$

Formula for SE(betahat1)

Thus

$$
\begin{aligned}
SE(\hat{\beta}_1) &= \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}} = \sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_1)} \\
&= \sqrt{\frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n}((X_i - \bar{X})^2 \hat{u}_i^2)}{\left[\frac{1}{n}\sum_i^n (X_i - \bar{X})^2\right]^2}}
\end{aligned}
$$

- It looks complicated but the software does it for us. In Stata's regression output it appears in the "Std. Err." column.

Coefficient estimates and their variance in Stata: test scores and STR

```
. reg testsc str

      Source |       SS           df       MS      Number of obs   =       420
-------------+----------------------------------   F(1, 418)       =     22.58
       Model |  7794.11004         1  7794.11004   Prob > F        =    0.0000
    Residual |  144315.484       418  345.252353   R-squared       =    0.0512
-------------+----------------------------------   Adj R-squared   =    0.0490
       Total |  152109.594       419  363.030056   Root MSE        =    18.581

------------------------------------------------------------------------------
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -2.279808   .4798256    -4.75   0.000     -3.22298   -1.336637
       _cons |    698.933   9.467491    73.82   0.000     680.3231    717.5428
------------------------------------------------------------------------------
```

Summary for testing null against 2 sided alternative

- Construct the t-statistic:

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{\sigma}^2_{\hat{\beta}_1}}}$$

- Reject at 5% significance level if $|t| > 1.96$
- The p-value is $p = Pr[|t| > |t^{act}|]$ = probability in tails of standard normal distribution above $|t^{act}|$
  - Reject H_0 at the 5% significance level if the p-value < 0.05.
  - In general, reject H_0 at the $\alpha \times 100\%$ significance level if the p-value is $< \alpha$.
- This procedure relies on the large-n approximation; typically $n = 50$ is large enough for the approximation to be excellent.

Testing whether class size affects performance

- We test whether STR has any effect on testscores,

$$H_0 : \beta_1 = 0 \ vs. \ H_1 : \beta_1 \neq 0$$

$$testscore = \underset{(9.47)}{698.93} - \underset{(0.48)}{2.28} \ str$$

- The t-statistic is

$$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = -4.75$$

- Check the "t" column in Stata's regression output (automatically checks hypothesis that coefficient is zero).
- Since $|t| > 1.64$, we **reject** the null hypothesis at 10%;
- Since $|t| > 1.96$, we **reject** the null hypothesis at 5%;
- Can we reject at 1.35%? Check and see whether the p-value is less than 0.0135 ....
  - The p-value – given in column $P > |t|$ – is essentially zero, thus we **reject** at any (relevant) significance level.

Test for the significance of a regressor

- The t-statistc reported by Stata is for the null hypothesis $H_0 : \beta_1 = 0$.
- Thus, the p-value reported by Stata is the p-value for $H_0 : \beta_1 = 0$ *vs.* $H_1 : \beta_1 \neq 0$.
- If your test is different, e.g., $H_0 : \beta_1 = 1$, you can't use the reported t-statistic and p-value!
- However, the test $H_0 : \beta_1 = 0$ *vs.* $H_1 : \beta_1 \neq 0$ is perhaps the most popular one, we call it a **test for the significance of a regressor** (or its parameter).
  - If we reject the null hypothesis, we say that the coefficient β_1 is "statistically significant".

Where are we?

1. Hypothesis tests about β_1 (SW 5.1)
2. **Confidence intervals about β_1 (SW 5.2)**
3. Empirical application
4. Regression when X is binary (0/1) (SW 5.3)
5. Heteroskedasticity and homoskedasticity (SW 5.4)
6. The theoretical foundations of OLS – Gauss-Markov Theorem (SW 5.5)

Confidence interval for beta1 (SW 5.2)

- Recall that a 95% confidence interval can be expressed, equivalently:
  - as the set of points that cannot be rejected at the 5% significance level;
  - as an interval (that is a function of the data) that contains the true parameter value 95% of the time in repeated samples.
- In general, if the sampling distribution of an estimator is normal for large n, then a 95% confidence interval can be constructed as the estimator $\pm$ 1.96$\times$ standard error.
- So, a 95% confidence interval for β_1 is,

$$\begin{aligned} & \{\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\} \\ = & \{\hat{\beta}_1 - 1.96 \times SE(\hat{\beta}_1)\ ,\ \hat{\beta}_1 + 1.96 \times SE(\hat{\beta}_1)\} \end{aligned}$$

  - The format of the confidence interval for β_1 is similar to that for the mean μ_Y and for the difference in group means (Lecture 3).

Confidence interval

$$testscore = \underset{(9.47)}{698.93} - \underset{(0.48)}{2.28}\, str$$

$$\hat{\beta}_1 = -2.28, \quad SE(\hat{\beta}_1) = 0.48$$

- 95% Confidence interval

$$\left(\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\right) = (-2.28 \pm 1.96 \times 0.48) = (-3.22, -1.34)$$

- Check last two columns in Stata's regression output.
- 90% Confidence interval

$$\left(\hat{\beta}_1 \pm 1.64 \times SE(\hat{\beta}_1)\right) = (-2.28 \pm 1.64 \times 0.48) = (-3.22, -1.49)$$

Where are we?

1. Hypothesis tests about β_1 (SW 5.1)
2. Confidence intervals about β_1 (SW 5.2)
3. **Empirical application**
4. Regression when X is binary (0/1) (SW 5.3)
5. Heteroskedasticity and homoskedasticity (SW 5.4)
6. The theoretical foundations of OLS – Gauss-Markov Theorem (SW 5.5)

Empirical application: wages and education

Italian Labour Force Survey - ISTAT, 2015 Q3

RETRIC: wages of full-time employees

```
. sum retric,d

                  retribuzione netta del mese scorso
-------------------------------------------------------------
      Percentiles      Smallest
 1%          250            250
 5%          500            250
10%          680            250       Obs              26,127
25%         1000            250       Sum of Wgt.      26,127

50%         1300                      Mean           1306.757
                        Largest       Std. Dev.      522.5101
75%         1550           3000
90%         1950           3000       Variance       273016.8
95%         2290           3000       Skewness       .7246319
99%         3000           3000       Kurtosis       4.273194
```

Empirical application: wages and education

Italian Labour Force Survey - ISTAT, 2015 Q3

Education: education level attained

```
. tab educ_lev if retric~=.
```

educ_lev	Freq.	Percent	Cum.
Nessun titolo	142	0.54	0.54
Licenza elementare	700	2.68	3.22
Licenza media	7,510	28.74	31.97
Diploma 2-3	2,289	8.76	40.73
Diploma 4-5	10,530	40.30	81.03
Laurea	4,956	18.97	100.00
Total	26,127	100.00	

Recoding education

- We recode educ_lev in terms of years of education (and call it educ_years)

```
. //recoding education
. recode educ_lev (1=0) (2=5) (3=8) (4=11) (5=13) (6=18), gen(educ_years)
(89794 differences between educ_lev and educ_years)

. tab educ_lev educ_years if retric~=.
```

	RECODE of educ_lev						
educ_lev	0	5	8	11	13	18	Tot
Nessun titolo	142	0	0	0	0	0	1
Licenza elementare	0	700	0	0	0	0	7
Licenza media	0	0	7,510	0	0	0	7,5
Diploma 2-3	0	0	0	2,289	0	0	2,2
Diploma 4-5	0	0	0	0	10,530	0	10,5
Laurea	0	0	0	0	0	4,956	4,9
Total	142	700	7,510	2,289	10,530	4,956	26,1

OLS regression of wages on education in Stata

```
. reg retric educ_years

      Source |       SS           df       MS      Number of obs   =    26,127
-------------+----------------------------------   F(1, 26125)     =   2740.72
       Model |   677244129         1   677244129   Prob > F        =    0.0000
    Residual |  6.4556e+09    26,125  247104.039   R-squared       =    0.0949
-------------+----------------------------------   Adj R-squared   =    0.0949
       Total |  7.1328e+09    26,126  273016.809   Root MSE        =     497.1

------------------------------------------------------------------------------
      retric |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  educ_years |    43.0118   .8215895    52.35   0.000     41.40144    44.62216
       _cons |   788.4206   10.36762    76.05   0.000     768.0995    808.7417
------------------------------------------------------------------------------
```


OLS regression of wages on education in Stata

- We usually report results in a table but sometimes we write

$$\widehat{wages} = \underset{(10.37))}{788.42} + \underset{(0.822)}{43.01}\, educ$$

- We reject $H_0 =$ "no effect of education" at the 5% significance level.
- An additional year of education is associated with an increase of 43 euros in average monthly salary.
- The average wage for individuals with no education is 788.4 euros per month.

Where are we?

1. Hypothesis tests about β_1 (SW 5.1)
2. Confidence intervals about β_1 (SW 5.2)
3. Empirical application
4. **Regression when X is binary (0/1) (SW 5.3)**
5. Heteroskedasticity and homoskedasticity (SW 5.4)
6. The theoretical foundations of OLS – Gauss-Markov Theorem (SW 5.5)

Regression when X is binary (SW 5.3)

- Sometimes a regressor is binary:
    - X = 1 if female, = 0 if male.
    - X = 1 if treated (experimental drug), = 0 if not treated.
    - X = 1 if small class size, = 0 if not a small class size.
- Binary regressors are often called "dummy" variables or regressors.
- So far, β_1 has been called a "slope," but that doesn't make much sense if X is binary.
- How do we interpret regression with a binary regressor?

Interpretation of slope parameter when X is binary

- Consider
$$Y_i = \beta_0 + \beta_1 X_i + u_i,$$
  when X is binary.
- Then,

$$E[Y_i|X_i = 0] = \beta_0$$
$$E[Y_i|X_i = 1] = \beta_0 + \beta_1$$

Thus,

$$\begin{aligned} \beta_1 &= E[Y_i|X_i = 1] - E[Y_i|X_i = 0] \\ &= \text{population difference in group means} \end{aligned}$$

Example of binary X

- When X is binary we usually call it D (for dummy).
- For example,
$$Testscore = \beta_0 + \beta_1 D + u$$
  where
$$D_i = \begin{cases} 1 & \text{if STR}_i \leq 20 \\ 0 & \text{if STR}_i > 20 \end{cases}$$

Generating a dummy variable in Stata

- There are several options.
- For example, `g D=(str<=20)`

```
. g D=(str<=20)

. reg testscr D

      Source |       SS           df       MS      Number of obs   =       420
-------------+----------------------------------   F(1, 418)       =     15.05
       Model |  5286.87866         1  5286.87866   Prob > F        =    0.0001
    Residual |  146822.715       418  351.250514   R-squared       =    0.0348
-------------+----------------------------------   Adj R-squared   =    0.0324
       Total |  152109.594       419  363.030056   Root MSE        =    18.742

------------------------------------------------------------------------------
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           D |   7.185129    1.85201     3.88   0.000     3.544715    10.82554
       _cons |   649.9994   1.408711   461.41   0.000     647.2304    652.7685
------------------------------------------------------------------------------
```


Comparison with difference in group means

- In Lecture 1 we computed

	Test score		
STR	n	mean	sd
Small	243	657.19	19.29
Large	177	650.00	17.97
All	420	654.16	19.05

- And the difference in test score means is

$$\bar{Y}_{small} - \bar{Y}_{large} = 657.19 - 650.0 = 7.19$$

exactly as the OLS estimator of β_1 ...as expected.

Comparison with difference in group means

- We also showed in Lecture 1 that

$$SE(\bar{Y}_{small} - \bar{Y}_{large}) = \sqrt{\frac{s^2_{small}}{n_{small}} + \frac{s^2_{large}}{n_{large}}} = \sqrt{\frac{19.29^2}{243} + \frac{17.97^2}{177}} = 1.832$$

  difference in SE is due to rounding.
- Given the OLS estimates, what are the mean testscores in both groups of districts?

Gender difference in wages

- Variable sg11 denotes gender of individual.
- sg11 coded as: 1 for male; 2 for female
- Recode and create new variable `female`
- What does -291 mean?

```
. recode sg11 (1=0) (2=1), g(female)
(101916 differences between sg11 and female)

. reg retric female

      Source |       SS           df       MS      Number of obs   =    26,127
-------------+----------------------------------   F(1, 26125)     =   2195.02
       Model |   552850657         1   552850657   Prob > F        =    0.0000
    Residual |  6.5800e+09    26,125  251865.512   R-squared       =    0.0775
-------------+----------------------------------   Adj R-squared   =    0.0775
       Total |  7.1328e+09    26,126  273016.809   Root MSE        =    501.86

------------------------------------------------------------------------------
      retric |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -291.3592   6.218838   -46.85   0.000    -303.5485     -279.17
       _cons |   1444.535   4.276474   337.79   0.000     1436.153    1452.917
------------------------------------------------------------------------------
```

Where are we?

1. Hypothesis tests about β_1 (SW 5.1)
2. Confidence intervals about β_1 (SW 5.2)
3. Empirical application
4. Regression when X is binary (0/1) (SW 5.3)
5. **Heteroskedasticity and homoskedasticity (SW 5.4)**
6. The theoretical foundations of OLS – Gauss-Markov Theorem (SW 5.5)

Heteroskedasticity and homoskedasticity (SW 5.4)

- What do these two terms mean?
- If $Var(u|X = x)$ is constant — that is, if the variance of the conditional distribution of u given X does not depend on X then u is said to be **homoskedastic**. Otherwise, u is **heteroskedastic**.
- What, if any, is the impact on the OLS estimator of having homoskedastic or heteroskedastic errors?

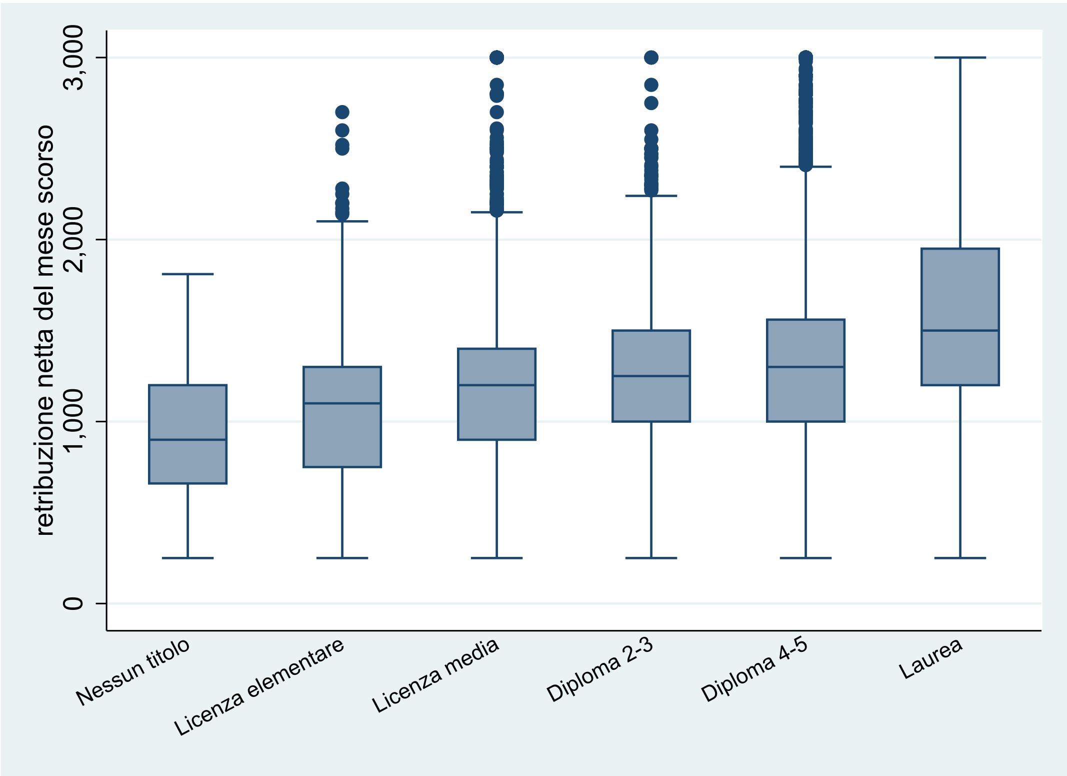
Heteroskedasticity and homoskedasticity

- Consider the example

$$wage_i = \beta_0 + \beta_1 educ_i + u_i$$

- Homoskedasticity means that the variance of u_i does not change with the education level.
- Of course, we do not know anything about $Var(u_i|educ_i)$, but we can use data to get an idea.
- On option is to plot the boxplot of wages for each educational category — if u_i is homoskedastic, the boxes should approximately be of the same size

Homoskedasticity in a picture



Homoskedasticity in a table

Another option is to check the variances directly

```
. bysort educ_lev: sum retric
```

-> educ_lev = Nessun titolo

Variable	Obs	Mean	Std. Dev.	Min	Max
retric	142	904.2254	322.8163	250	1810

-> educ_lev = Licenza elementare

Variable	Obs	Mean	Std. Dev.	Min	Max
retric	700	1037.114	423.5996	250	2700

-> educ_lev = Licenza media

Variable	Obs	Mean	Std. Dev.	Min	Max
retric	7,510	1156.618	429.893	250	3000

-> educ_lev = Diploma 2-3

Variable	Obs	Mean	Std. Dev.	Min	Max
retric	2,289	1243.128	429.0504	250	3000

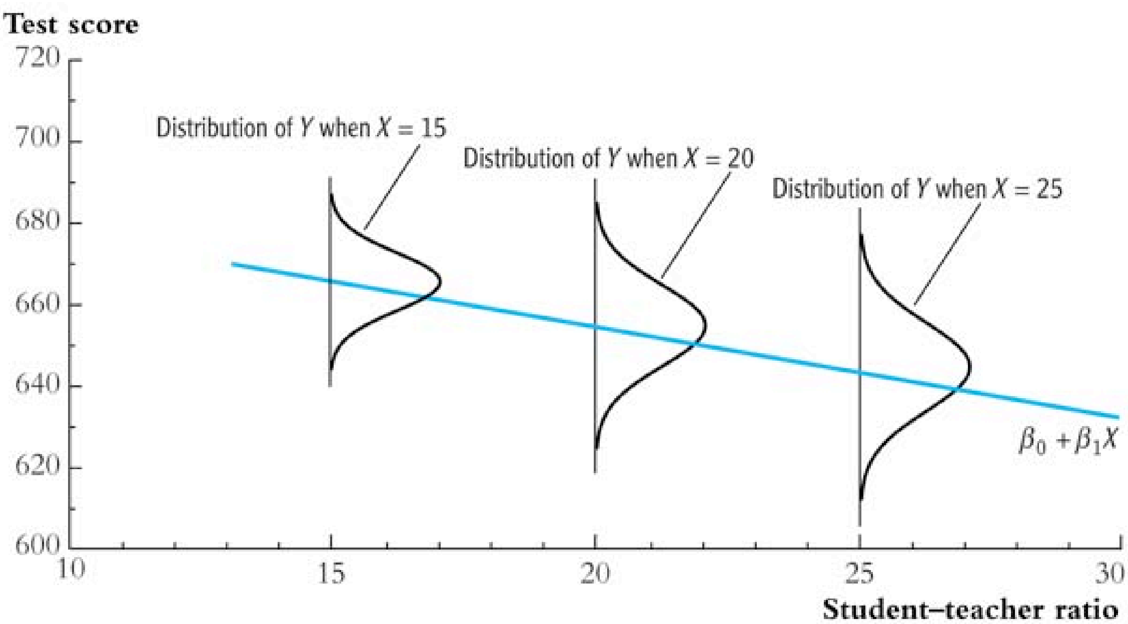
-> educ_lev = Diploma 4-5

Variable	Obs	Mean	Std. Dev.	Min	Max
retric	10,530	1307.365	485.4817	250	3000

-> educ_lev = Laurea

Variable	Obs	Mean	Std. Dev.	Min	Max
retric	4,956	1611.981	633.4167	250	3000

More pictures...



Impact of Heteroskedasticity/homoskedasticity on bias

- Whether the error is heteroskedastic or homoskedastic does not matter for the unbiasedness and consistency of the OLS estimator.
- Recall that the OLS estimator is unbiased and consistent under the three Least Squares Assumptions...
- And the heteroskedasticity/homoskedasticity distinction does not appear in these assumptions.....so it does not affect the unbiasedness/consistency of the OLS estimator.
- The heteroskedasticity/homoskedasticity distinction also does not affect the large sample (asymptotic) normality of the OLS estimator.

Impact of Heteroskedasticity/homoskedasticity on variance

- The only place where this distinction matters is in the formula for computing the variance of the OLS estimator $\hat{\beta}_1$.
- The formula presented in Lecture 4 is valid **always**, irrespective of whether there is heteroskedasticity or homoskedasticity.
- This formula delivers **heteroskedastic-robust standard errors** (which are also correct if there is homoskedasticity).
- There is **another** formula – not presented here – which is valid **only** when there is homoskedasticity. This is a simpler formula and it is often the default setting for many software programs.
  - We sometimes call the S.E. based on this simpler estimator: "homoskedasticity-only SEs".
  - To get the heteroskedastic-robust std. errs. you must override the default.
- If you don't override the default and there is in fact heteroskedasticity, your standard errors (and t-statistics and confidence intervals) will be wrong.
  - Typically, homoskedasticity-only SEs are somewhat smaller.

Summary on heteroskedasticity/homoskedasticity

- If the errors are either homoskedastic or heteroskedastic and you use heteroskedastic-robust standard errors, you are OK.
- If the errors are heteroskedastic and you use the homoskedasticity-only formula for standard errors, your standard errors will be wrong (the homoskedasticity-only estimator of the variance of $\hat{\beta}_1$ is inconsistent if there is heteroskedasticity).
- The two formulas coincide (when n is large) in the special case of homoskedastic errors.
- So, you should **always** use heteroskedasticity-robust standard errors.

Heteroskedastic-robust standard errors in Stata

Use option `robust`

```
. reg testscr str,robust

Linear regression                               Number of obs     =        420
                                                F(1, 418)         =      19.26
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0512
                                                Root MSE          =     18.581

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -2.279808   .5194892    -4.39   0.000    -3.300945   -1.258671
       _cons |    698.933   10.36436    67.44   0.000     678.5602    719.3057
------------------------------------------------------------------------------

. reg testscr str

      Source |       SS           df       MS      Number of obs   =       420
-------------+----------------------------------   F(1, 418)       =     22.58
       Model |  7794.11004         1  7794.11004   Prob > F        =    0.0000
    Residual |  144315.484       418  345.252353   R-squared       =    0.0512
-------------+----------------------------------   Adj R-squared   =    0.0490
       Total |  152109.594       419  363.030056   Root MSE        =    18.581

------------------------------------------------------------------------------
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -2.279808   .4798256    -4.75   0.000     -3.22298   -1.336637
       _cons |    698.933   9.467491    73.82   0.000     680.3231    717.5428
------------------------------------------------------------------------------
```

Where are we?

1. Hypothesis tests about β_1 (SW 5.1)
2. Confidence intervals about β_1 (SW 5.2)
3. Empirical application
4. Regression when X is binary (0/1) (SW 5.3)
5. Heteroskedasticity and homoskedasticity (SW 5.4)
6. **The theoretical foundations of OLS – Gauss-Markov Theorem (SW 5.5)**

The theoretical foundations of OLS – Gauss-Markov Theorem

- We have already learned a very great deal about OLS:
  1. OLS is unbiased and consistent (under the three LS assumptions);
  2. We have a formula for heteroskedasticity-robust standard errors;
  3. We can use them to construct confidence intervals and test statistics.
- These are good properties of OLS. In addition, a very good reason to use OLS is that everyone else does, so by using it, others will understand what you are doing. In fact, OLS is the language of regression analysis, and if you use a different estimator, you will be speaking a different language.
- The natural question to ask is whether there other estimators that might have a smaller variance than OLS?

The theoretical foundations of OLS – Gauss-Markov Theorem

- We can always find an estimator than has a very small variance (e.g., take the estimator to be the number 1.1).
- So, when we ask wether there are estimators with a smaller variance than OLS we need to be precise about which type of estimators we want to compare OLS to.
- We take all estimators that are linear functions of $Y_1, \ldots Y_n$ and unbiased. Just as OLS is.
    - From the proof of unbiasedness in Lecture 4 we can deduce that we can write $\hat{\beta}_1 = \sum_{i=1}^{n} \omega_i Y_i$.
- We then have the following important result:

Theorem (Gauss-Markov Theorem)

If the three Least Squares assumptions hold and if the errors are homoskedastic, then the OLS estimator is the **B***est* **L***inear* **U***nbiased* **E***stimator (***BLUE***).*

The Gauss Markov Assumptions

- This is a pretty amazing result: it says that, if in addition to LSA 1-3 the errors are homoskedastic, then OLS is the **best choice** among all other linear and unbiased estimators.
- An estimator with the smallest variance is called an **efficient** estimator. The GM theorems says that OLS is the **efficient** estimator among the linear unbiased estimators of β_1.
- You could choose a biased estimator with zero variance, e.g. 1.1....but this is surely a poor choice!
- The set of four (4) assumptions necessary for the GM Theorem to hold are sometimes called the "Gauss-Markov Assumptions"

Limitations of the Gauss-Markov Theorem

- For the GM theorem to hold we need homoskedasticity which is often a not very realistic assumption.
- If there is heteroskedasticity – and the GM theorem does not hold – there may be be more efficient estimators than OLS. But OLS is still unbiased and consistent under the three LS assumptions...it just may not have the smallest variance possible.
- The result is only for linear estimators. There may be other non-linear estimators with lower variance.
- OLS is more sensitive to outliers than some other estimators.

A stronger Gauss-Markov Theorem

Theorem (Gauss-Markov Theorem)

If the three Least Squares assumptions hold and if the errors are homoskedastic and normally distributed, then the OLS estimator has the smallest variance among all consistent estimators.