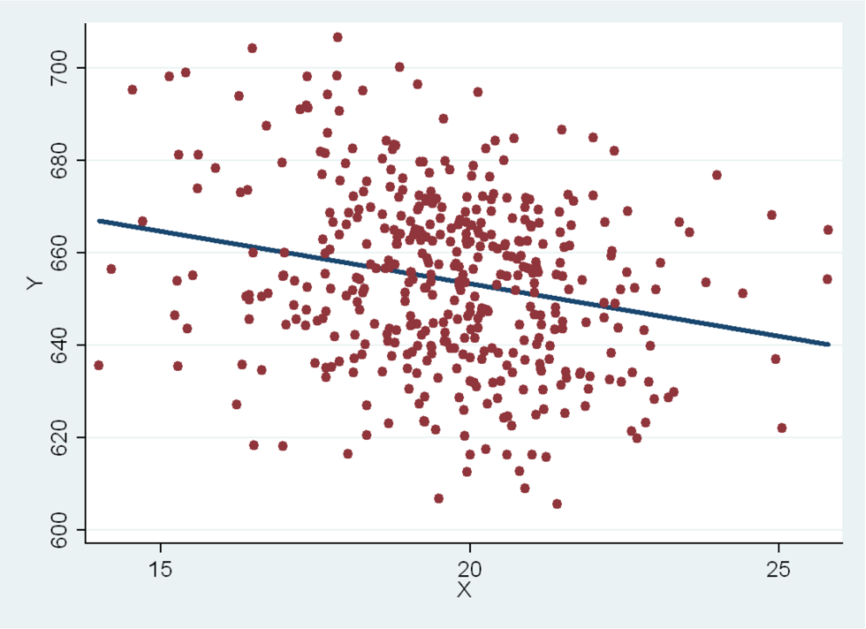
EC 3303: Econometrics I

Linear Regression with One Regressor
(Part 3)



Kelvin Seah

AY 2022/2023, Semester 2

Testing Hypotheses about β_1

- Does a change in one variable affect another?

  - is the population slope $\beta_1 = \frac{\Delta Y}{\Delta X} = 0$?

- E.g.: Do changes in class size affect test scores?

- Hypothesis testing allows us to test if some claim about the population (like $\beta_1 = 0$) can be rejected using evidence from a sample of data.

- You previously obtained an estimate $\hat{\beta}_{ClassSize} = -2.28$.

- Does this provide good evidence that $\beta_{ClassSize} \neq 0$?

- $\hat{\beta}_{ClassSize} = -2.28$ was estimated using a sample of data. Another sample would likely give a different estimate. Can we say more about our result?

- Specify the null & alternative hypotheses:

$$\boldsymbol{H_0: \beta_1 = \beta_{1,0}} \text{ vs. } \boldsymbol{H_1: \beta_1 \neq \beta_{1,0}} \quad \textbf{(2-sided alternative)}$$

$$\boldsymbol{H_0: \beta_1 = \beta_{1,0}} \text{ vs. } \boldsymbol{H_1: \beta_1 < \beta_{1,0}} \quad \textbf{(1-sided alternative)}$$

$$\boldsymbol{H_0: \beta_1 = \beta_{1,0}} \text{ vs. } \boldsymbol{H_1: \beta_1 > \beta_{1,0}} \quad \textbf{(1-sided alternative)}$$

where $\beta_{1,0}$ is the hypothesized value under the null.

- Depending on what we want to show, we pick the appropriate alternative hypothesis:
  - In class size e.g., if superintendent just wants to know whether class size has an effect on test score (whether positive or negative), then use a 2-sided alternative $H_1: \beta_1 \neq 0$.
  - But, if superintendent wants to know whether cutting class sizes ***improves*** test scores, then can use a 1-sided alternative $H_1: \beta_1 < 0$.

Two-Sided Alternative Hypotheses Concerning β_1

- Approach to testing hypothesis about the population slope β_1 is ***same as*** for the population mean $E(Y)$

$$H_0\!: \beta_1 = \beta_{1,0} \text{ vs. } H_1\!: \beta_1 \neq \beta_{1,0} \text{ (2-sided alternative)}$$

1) Compute the standard error of $\hat{\beta}_1$, $SE(\hat{\beta}_1)$

2) Compute the t-statistic: $t^{act} = \frac{\widehat{\beta^{act}}_1 - \beta_{1,0}}{SE(\widehat{\beta}_1)}$

3) Calculate the p-value: $Pr(|Z| > |t^{act}|) = 2\Phi(-|t^{act}|)$

p-value is the area in the tails of a standard normal outside $|t^{act}|$

- Reject H_0 at the 5% significance level if the p-value<0.05.
- or, reject H_0 at the 5% significance level if $|t^{act}| > 1.96$.

Two-Sided Alternative Hypotheses Concerning $\boldsymbol{\beta_1}$

- above procedure relies on the large n approximation. Recall that the distribution of $\hat{\beta}_1$ is approximately normal if the sample size is large.

- typically, $n \geq 100$ is sufficient for the distribution of $\hat{\beta}_1$ to be well approximated by a normal distribution.

Formula for $SE(\hat{\beta}_1)$

- $SE(\hat{\beta}_1)$ is the estimator of the standard deviation of the sampling distribution of $\hat{\beta}_1$.

$$SE(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}}$$

where

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n}(X_i - \bar{X})^2 \hat{u}^2_i}{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]^2}$$

$SE(\hat{\beta}_1)$ is computed by Stata, no need to memorize this formula.

More good news: the t-statistic and p-value are also computed automatically by Stata.

One-Sided Alternative Hypotheses Concerning β_1

- Sometimes, we might have reason to believe that β_1 is strictly smaller (or larger) than $\beta_{1,0}$.

- In such cases, can use a one-sided test.

  - e.g.: reasoning might lead you to believe that smaller classes should ***improve*** test scores.

  - So test the null that $\beta_1 = 0$ (no effect) against the one-sided alternative that $\beta_1 < 0$ (smaller classes are associated with higher test scores).

- Computation of $SE(\hat{\beta}_1)$ and t-statistic under a two-sided and one-sided test are the same.
    - ***calculation of p-values*** & ***critical values is different***.
- Under a one-sided test of the form:

$$H_0\text{: } \beta_1 = \beta_{1,0} \text{ vs. } H_1\text{: } \beta_1 < \beta_{1,0} \text{ (1-sided alternative)}$$

$$p - value = Pr(Z < t^{act})$$

    - p-value under a one-sided test is the area under the standard normal distribution to the ***left*** of the calculated t-statistic.
    - only values of the estimate smaller than $\beta_{1,0}$ count as evidence against H_0.
    - only ***large negative values*** of the t-statistic provide evidence against H_0.

$H_0\colon \beta_1 = \beta_{1,0}$ vs. $H_1\colon \beta_1 < \beta_{1,0}$ (1-sided alternative)

- Critical values for a one-sided alternative are:
  - If $\alpha = 0.05$, then for the 1-sided alternative above, reject H_0 if $t^{act} < -1.64$.
    - cuts off 5% of the area under the **lower tail** of the $N(0,1)$ distribution.
  - If $\alpha = 0.01$, then for the 1-sided alternative above, reject H_0 if $t^{act} < -2.33$.
- If alternative hypothesis is $H_1\colon \beta_1 > \beta_{1,0}$, everything discussed applies except that signs are switched.

Example: *Test Scores* & *STR*, California Data

Estimated (i.e. OLS) regression line:

$$\widehat{TestScore} = 698.9 - 2.28 \times STR$$

Regression software reports the standard errors:

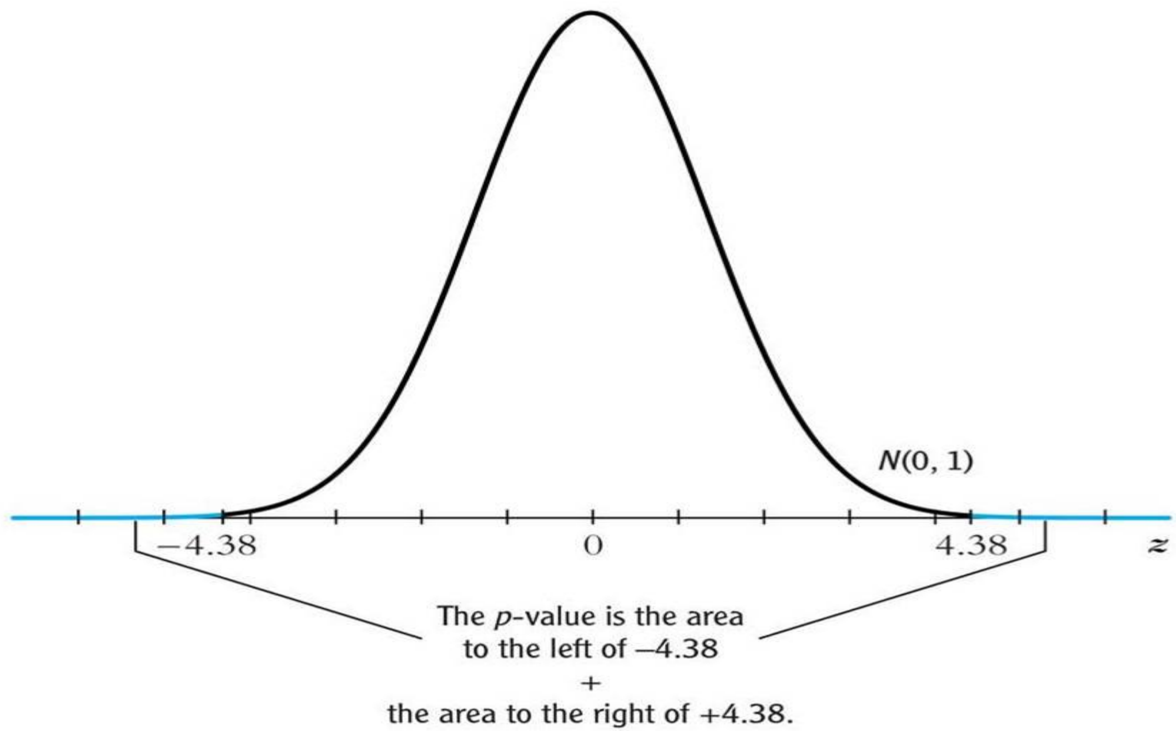
$$SE\left(\hat{\beta}_0\right) = 10.4 \text{ and } SE\left(\hat{\beta}_1\right) = 0.52$$

Suppose want to test whether β_1 is significantly different from zero at the 5% level:

1) Formulate hypothesis: $H_0 : \beta_1 = 0$; $H_1 : \beta_1 \neq 0$

2) Compute the t-statistic: $t^{act} = \frac{\widehat{\beta^{act}}_1 - \beta_{1,0}}{SE(\widehat{\beta}_1)} = \frac{-2.28-0}{0.52} = -4.38$

3) Calculate the p-value: $2\Phi(-|t^{act}|) = 2\Phi(-4.38) \approx 0.00001$



- probability of observing a value of $\hat{\beta}_1$ (or of observing a t-statistic) as extreme or more extreme than -2.28 (or -4.38) is very small, assuming that the null is true.

- since this event is so unlikely, conclude the null hypothesis is false.

- alternatively, since $|t^{act}| = 4.38$ >1.96 (5% two-sided critical value), reject H_0 in favour of H_1 at the 5% significance level.

Concise Way to Report Regression Results

Put standard errors in ***parentheses below the estimated coefficients*** to which they apply.

$$\widehat{TestScore} = 698.9 - 2.28 \times STR, \;\; R^2 = 0.05, SER = 18.6$$

(10.4)   (0.52)

expression gives a lot of information:

- estimated regression line is
$$\widehat{TestScore} = 698.9 - 2.28 \times STR$$
- standard error of $\hat{\beta}_0$ is 10.4
- standard error of $\hat{\beta}_1$ is 0.52
- R^2 is 0.05; SER is 18.6

OLS Regression: Reading STATA Output

```
regress testscr str, robust

Regression with robust standard errors                 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    -4.38    0.000    -3.300945   -1.258671
  _cons |    698.933   10.36436    67.44    0.000     678.5602    719.3057
-------------------------------------------------------------------------
```

so:

$$\widehat{TestScore} = 698.9 - 2.28 \times STR, \; , \; R^2 = \mathbf{0.05}, \; SER = \mathbf{18.6}$$

$$(10.4) \quad (0.52)$$

t-stat = **–4.38**, *p*-value = **0.000** (2-sided)

- p-value presented in Stata is for a ***two-sided test.***
- to get the p-value for a ***one-sided test***, ***divide*** the p-value shown in the output, by ***2***.

Confidence Intervals for β_1

- A confidence interval is an interval which contains the true value of a parameter with a certain prespecified probability.
  - e.g. A 95% **confidence interval** for β_1 is an interval that contains the true value of β_1 in 95% of repeated samples.
- Construction of a confidence interval for β_1 is ***same*** as for the population mean

When n is large,

$$95\% \text{ CI for } \beta_1 = \left[\hat{\beta}_1 - 1.96SE\left(\hat{\beta}_1\right), \hat{\beta}_1 + 1.96SE\left(\hat{\beta}_1\right)\right]$$

$$90\% \text{ CI for } \beta_1 = \left[\hat{\beta}_1 - 1.64SE\left(\hat{\beta}_1\right), \hat{\beta}_1 + 1.64SE\left(\hat{\beta}_1\right)\right]$$

$$99\% \text{ CI for } \beta_1 = \left[\hat{\beta}_1 - 2.58SE\left(\hat{\beta}_1\right), \hat{\beta}_1 + 2.58SE\left(\hat{\beta}_1\right)\right]$$

OLS Regression: Reading STATA Output

```
regress testscr str, robust

Regression with robust standard errors                 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     -4.38   0.000     -3.300945   -1.258671
  _cons |    698.933   10.36436     67.44   0.000      678.5602    719.3057
-------------------------------------------------------------------------------
```

so:

$$\widehat{TestScore} = 698.9 - 2.28 \times STR, \ , R^2 = .05, \ SER = 18.6$$

$$(10.4) \quad (0.52)$$

$t\ (\beta_1 = 0) = -4.38$, p-value $= 0.000$ (2-sided)

95% confidence interval for β_1 is $(-3.30, -1.26)$

Regression when X is a Binary Variable

- So far, regressor considered is a continuous one.
- But regressors can also be ***binary variables***; binary variables are those which take on only two values: 0 or 1.
  - aka ***indicator variables*** or ***dummy variables***.
- X might be a variable indicating:
  - gender (=1 if female, =0 if male)
  - class size (=1 if small, =0 if large)
  - marital status (=1 if married, =0 if not)

Interpretation of β_1 when X is a Binary Variable

- Interpretation of β_1 when X is a binary variable is different:

Suppose,

$$D_i = \begin{cases} 1 \; if \; the \; STR \; in \; the \; ith \; district < 20 \\ 0 \; if \; the \; STR \; in \; the \; ith \; district \geq 20 \end{cases}$$

- Population regression model with D_i as regressor is

$$Y_i = \beta_0 + \beta_1 D_i + u_i \quad (1)$$

- Because D_i can take on only 2 values, cannot think of β_1 as a "slope". Refer to β_1 as "***coefficient*** on D_i"

- If β_1 is not a slope, what is it?

$$Y_i = \beta_0 + \beta_1 D_i + u_i \qquad (1)$$

Consider the two cases, where $D_i = 0$ & $D_i = 1$:

- If class size is large ($D_i = 0$), (1) becomes

$$Y_i = \beta_0 + u_i \qquad (2)$$

- If class size is small ($D_i = 1$), (1) becomes

$$Y_i = \beta_0 + \beta_1 + u_i \qquad (3)$$

- population mean value of test scores when the class size is large is:

$$E(Y_i | D_i = 0) = \beta_0 \qquad (4)$$

[follows because, from LSA#1, $E(u_i | D_i) = 0$]

- population mean value of test scores when the class size is small is:

$$E(Y_i | D_i = 1) = \beta_0 + \beta_1 \qquad (5)$$

(5) − (4) yields:

$$\beta_1 = E(Y_i|D_i = 1) - E(Y_i|D_i = 0)$$

- β_1 is the difference between mean test score in districts with small classes and mean test score in districts with large classes.

- In general, if X is binary, β_1 is the difference between the mean value of Y_i when $X_i = 1$ and the mean value of Y_i when $X_i = 0$:

$$\beta_1 = E(Y_i|X_i = 1) - E(Y_i|X_i = 0)$$

- Hypothesis tests & confidence intervals are conducted as before.

Example: *Test Scores & Class Size*, California Data

- Using OLS to estimate the relationship between test scores and the class size binary variable $SmallSTR_i$ using the 420 observations, we get:

$$\widehat{TestScore} = 650.0 + 7.4 SmallSTR, \;\; R^2 = 0.037, SER = 18.7$$

$$(1.3) \qquad (1.8)$$

- average test score for the subsample of districts with large classes ($SmallSTR_i = 0$) is 650.0; average test score for the subsample of districts with small classes ($SmallSTR_i = 1$) is $650.0 + 7.4 = 657.4$.

- "estimated difference in mean test scores between districts with small classes and large classes in this sample is 7.4".

Are the population mean test scores for the two groups different?

- if the mean test score for the two groups, in the population, are the same, then $\beta_1 = 0$.

$$H_0\colon \beta_1 = 0 \text{ vs. } H_1\colon \beta_1 \neq 0$$

1) Compute the t-statistic: $t^{act} = \frac{\widehat{\beta^{act}}_1 - \beta_{1,0}}{SE(\widehat{\beta}_1)} = \frac{7.4-0}{1.8} = 4.04$

2) $|t^{act}| = 4.04 > 1.96$, so we reject H_0 at the 5% significance level

Conclude: population mean test scores for the two groups are statistically different at the 5% level.

Heteroskedasticity & Homoskedasticity

- What the…?!!!
- Consequences
- Implication for computing standard errors

Heteroskedasticity & Homoskedasticity

What do these two terms mean?

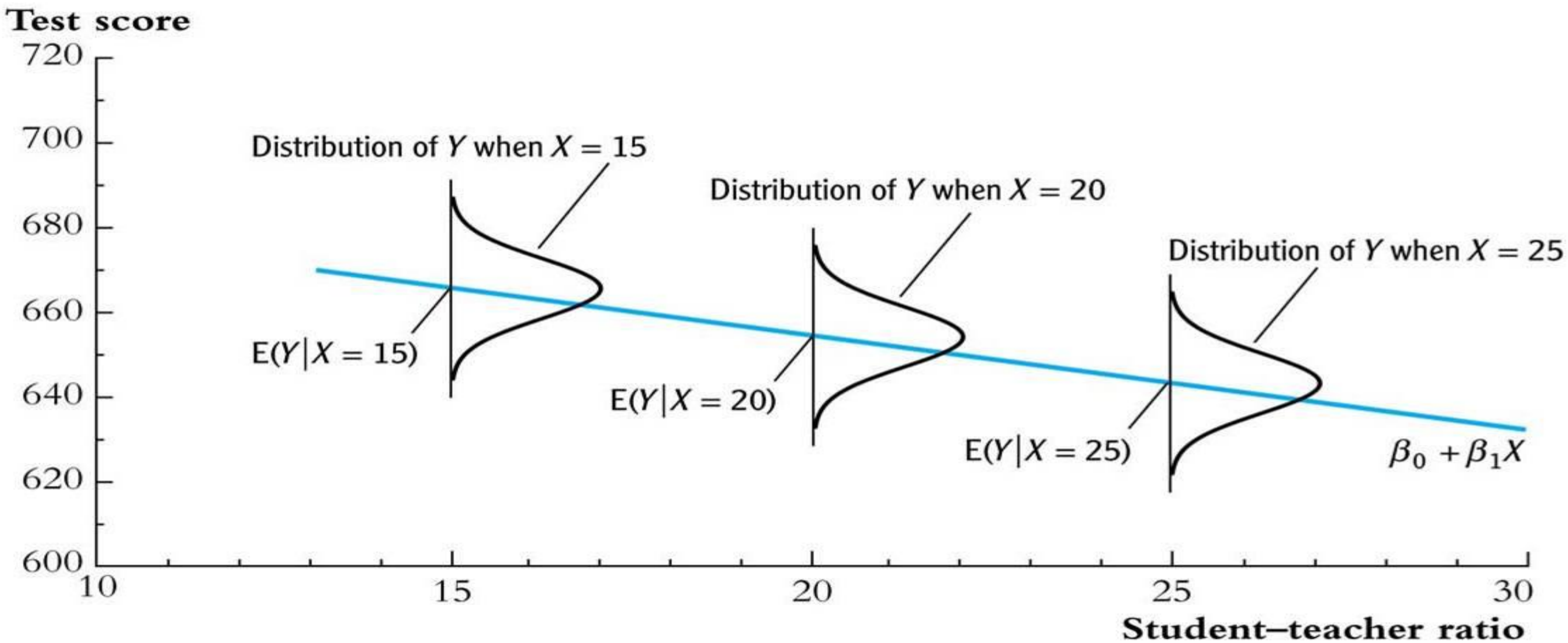
- If $Var(u_i|X_i = x) = \sigma^2{}_u$ (a constant) – that is, if it does not depend on the value of X_i, then the population errors are homoskedastic.

- Otherwise they are heteroskedastic.

Words have Greek origins

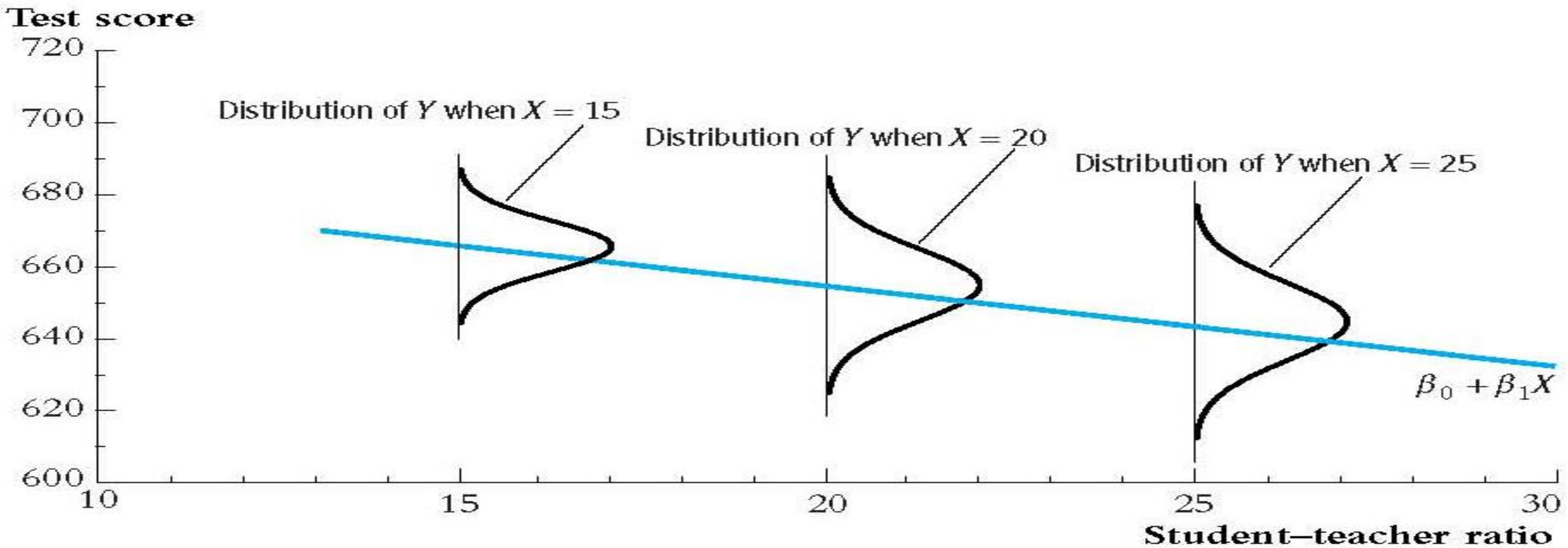
- ***Homo***=same, ***Hetero***=different, ***Skedanime***=disperse, scatter

E.g. of Homoskedasticity in a Picture



- Distribution of the population errors u_i is shown for $X_i = 15, 20,$ and 25.
- $E(u_i|X_i = x) = 0$ (distribution of errors satisfies LSA#1).
- All the conditional distributions have the same spread; conditional variance of u_i given $X_i = x$ does not depend on x, so errors are ***homoskedastic.***

E.g. of Heteroskedasticity in a picture

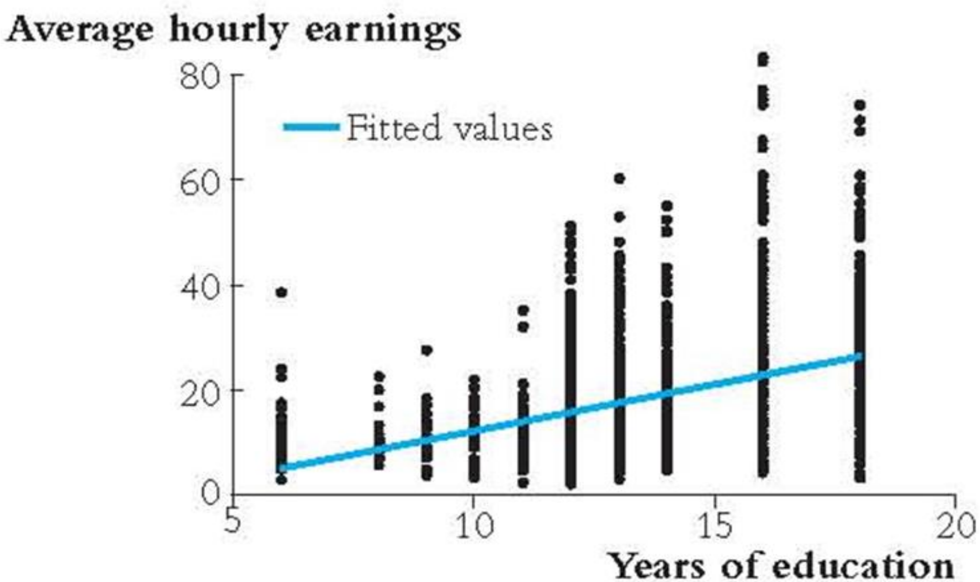


- $E(u_i|X_i = x) = 0$ (here, distribution of errors also satisfies LSA#1).

- But conditional variance of u_i here depends on the value of X_i; conditional distribution of u_i spreads out more as the value of X_i increases so the errors are ***heteroskedastic.***

Errors are Heteroskedastic in Many Econometric Applications…

- OLS residuals $\hat{u}_i$ are sample counterparts of population errors u_i.
- can examine the distributions of residuals to learn about the distributions of population errors.



Are the errors here likely to be homoskedastic or heteroskedastic?

Heteroskedasticity & Homoskedasticity

Recall the three least squares assumptions:

1. $E(u_i \mid X_i = x) = 0$, for $i = 1,\ldots,n$.
2. (X_i, Y_i), $i = 1,\ldots,n$, are i.i.d.
3. Large outliers are rare

They say nothing about the conditional variance of u_i

So whether the error terms u_i are homoskedastic or heteroskedastic, the properties we learnt earlier about the OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_0$ remain – the OLS estimators are ***unbiased***, ***consistent***, and have an asymptotically ***normal distribution*** in large samples.

What is Different Then?

- One important difference in whether the errors are homoskedastic or heteroskedastic is in the computation of the variance and standard error of $\hat{\beta}_1$.

- formulas for the variance and standard errors of $\hat{\beta}_1$ ***differ*** based on whether the errors are homoskedastic or heteroskedastic.

- without explicitly saying so, we have actually been assuming that u_i might be heteroskedastic.

- recall the formulas for the variance & standard error of $\hat{\beta}_1$:

$$Var(\hat{\beta}_1) = \frac{1}{n} \times \frac{Var[(X_i - \mu_X)u_i]}{[Var(X_i)]^2}$$

$$SE(\hat{\beta}_1) = \sqrt{\frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n}(X_i - \bar{X})^2 \hat{u}^2{}_i}{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]^2}}$$

$$Var(\hat{\beta}_1) = \frac{1}{n} \times \frac{Var[(X_i - \mu_X)u_i]}{[Var(X_i)]^2}$$

$$SE(\hat{\beta}_1) = \sqrt{\frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n}(X_i - \bar{X})^2 \hat{u}^2{}_i}{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]^2}}$$

- These variance and standard error formulas are general – they are valid whether the errors are heteroskedastic or homoskedastic.

- They are called the ***heteroskedasticity-robust variance*** and the ***heteroskedasticity-robust standard error.***

What if the Errors are in fact Homoskedastic?

Then formula for the variance and the standard error of $\hat{\beta}_1$ simplifies:

If var($u_i|X_i=x$) = σ_u^2 (i.e. some constant), then

$$\text{var}(\hat{\beta}_1) = \frac{\sigma_u^2}{n\sigma_X^2}$$ (homoskedasticity-only variance formula)

where $\sigma^2{}_u$ is the conditional variance of the error term; $\sigma^2{}_X$ is the variance of X; and n is the sample size

and

$$SE(\hat{\beta}_1) = \sqrt{\frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}}$$

(homoskedasticity-only standard error)

We now have 2 Formulas for the Standard Error of $\widehat{\beta}_1$

$$SE(\hat{\beta}_1) = \sqrt{\frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}}$$

- ***Homoskedasticity-only standard errors*** – these are valid only if the errors are homoskedastic.

$$SE\left(\hat{\beta}_1\right) = \sqrt{\frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n}(X_i - \bar{X})^2 \hat{u}^2{}_i}{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]^2}}$$

- The usual standard errors – also called ***heteroskedasticity- robust standard errors*** – are valid whether the errors are heteroskedastic or homoskedastic.
- The two formulas will coincide (when *n* is large) if the population errors are homoskedastic.

Practical Implications

- Homoskedasticity-only formula and heteroskedasticity-robust formula for the standard error of $\hat{\beta}_1$ differ:
  - So, in general, ***you will get different values of the standard errors using the different formulas***
- Homoskedasticity-only standard errors are the default setting in statistical software. To get the general heteroskedasticity-robust standard errors, you have to override the default.
- If you do not override the default (by specifying "robust" in STATA) and if the errors are in fact heteroskedastic, your standard errors will be ***wrong.***
  - Typically, homoskedasticity-only standard errors are too small.

Heteroskedasticity-Robust Standard Errors in STATA

```
regress testscr str, robust

Regression with robust standard errors                 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
```

- If you use the "**robust**" option, STATA computes heteroskedasticity-robust standard errors.

- Otherwise, STATA computes homoskedasticity-only standard errors.

The Bottom Line

- If the errors are either homoskedastic or heteroskedastic and you use heteroskedasticity-robust standard errors, you are fine.
- But if the errors are heteroskedastic and you use the homoskedasticity-only formula, your standard errors will be wrong.
- So, in practice, ***always*** use ***heteroskedasticity-robust standard errors***.