Lecture 4: Hypothesis Testing in OLS Regression

Introduction to Econometrics,Spring 2025

Zhaopeng Qu

Nanjing University Business School

March 20 2025

Review of the Previous Lecture

Omitted Variable Bias and Multiple OLS Regression

- **Omitted Variable Bias(OVB)** violates the first Least Squares Assumption:

$$E(u_i|X_i) = 0$$

  - It renders Simple OLS estimation both **biased** and **inconsistent**.
  - If the omitted variable can be observed and measured, we can include it in the regression, thereby **controlling** for it to eliminate the bias.
- We extended **Simple OLS regression** to **Multiple OLS regression**.

Multiple OLS Regression

- The multiple regression model is expressed as:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + ... + \beta_k X_{k,i} + u_i, i = 1, ..., n$$

- where:
    - Y_i is the **dependent variable**
    - $X_1, X_2, ...X_k$ are the **independent variables** (including one treatment variable and several control variables)
    - $\beta_j, j = 1...k$ are the slope coefficients corresponding to each X_j
    - β_0 is the **intercept**, representing the value of Y when all $X_j = 0, j = 1...k$
    - u_i is the **error term** (unobserved factors that affect Y)

Multiple OLS Regression: Estimation

- Multiple OLS regression estimates the coefficients $\beta_0, \beta_1, ...\beta_k$ by minimizing the sum of squared residuals $\hat{u}_i^2$:

$$\arg \min_{b_0, b_1, ..., b_k} \sum (Y_i - b_0 - b_1 X_{1,i} - ... - b_k X_{k,i})^2$$

where $b_0 = \hat{\beta}_0, b_1 = \hat{\beta}_1, ..., b_k = \hat{\beta}_k$ are the Multiple OLS estimators.

Multiple Regression: Assumptions

If the four least squares assumptions in the multiple regression model hold:

- Assumption 1: The conditional distribution of u_i given $X_{1i}, \ldots, X_{ki}$ has zero mean, thus
$$E[u_i|X_{1i}, \ldots, X_{ki}] = 0$$
- Assumption 2: $(Y_i, X_{1i}, \ldots, X_{ki})$ are independently and identically distributed (i.i.d.)
- Assumption 3: Large outliers are unlikely
- Assumption 4: **No perfect multicollinearity**

Then:

- The OLS estimators $\hat{\beta}_0, \hat{\beta}_1 \ldots \hat{\beta}_k$ are *unbiased*
- The OLS estimators $\hat{\beta}_0, \hat{\beta}_1 \ldots \hat{\beta}_k$ are *consistent*
- The OLS estimators $\hat{\beta}_0, \hat{\beta}_1 \ldots \hat{\beta}_k$ are *normally distributed* in large samples

Multiple Regression and Causality

- OLS regression yields valid causal explanations only when all least squares assumptions are satisfied.
- The most critical assumption is the **Conditional Expectation Zero (CEZ)**:

$$E(u_i|D,C) = E(u_i|C)$$

  - where D is the treatment variable and C represents the control variable(s).
- In causal inference, our *primary focus* is ensuring that the coefficient of the treatment variable D, denoted as β_D, is *unbiased* and *consistent*, rather than concerning ourselves with all coefficients $\beta_j, j = 0, 1, ..., k$ in the model.
- In most cases, non-experimental data fails to satisfy these conditions. Therefore, the central challenge is establishing convincing causal inference when these assumptions are violated.
  - Solutions include: Instrumental Variables (IV), Regression Discontinuity (RD), Difference-in-Differences (DID), Synthetic Control Methods (SCM), etc.

Hypothesis Testing

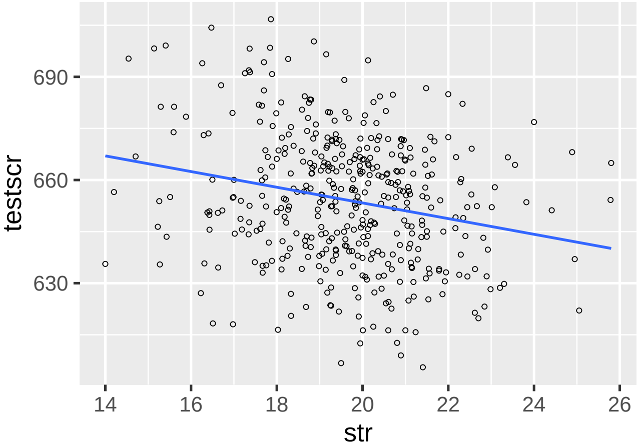
From Samples to the Population

- So far,we have learned how to estimate the OLS regression model and how to interpret the results.
- However, don't forget that our estimation is based on **a sample**, and the result may not be representative of the **population**.
- Therefore, we have to make sure that our estimation based on a sample is not **a coincidence**, but **a reliable** inference for the population.
    - **Hypothesis testing** is a tool to help us to make this inference.

Class size and Test Score

Recall our simple OLS regression mode is

$$TestScore_i = \beta_0 + \beta_1 STR_i + u_i \tag{4.3}$$



Class Size and Test Score

- Then we got the result of a simple OLS regression

$$\widehat{TestScore} = 698.9 - 2.28 \times STR, \;\; R^2 = 0.051, SER = 18.6, N = 420$$

- How can you be certain about the result in population as the one in a sample?
  - In other words, *how confident* you can believe the result from the sample inferring to the population?
- If someone believes that your results are not reliable but coincidental.
  - They states that cutting the class size will **NOT** help boost test scores.
- Can you dismiss the claim based your *scientific evidence-based* data analysis?
  - This is where **Hypothesis Testing in OLS regressions** comes into play.

Review: Hypothesis Testing

- A hypothesis is typically an **assertion** or **statement** about **unknown population parameters**,
  - Such as θ, which can be any **statistic** of interest including the *mean*, *variance*, *median*, etc.
- Suppose we want to test
  - *whether the parameter is significantly different from a specific value* μ_0
- Then we set two *mutually exclusive* competing hypotheses:
  - **null hypothesis**:
$$H_0 : \theta = \mu_0$$
  - **alternative hypothesis**:
$$H_1 : \theta \neq \mu_0$$

Review: Hypothesis Testing

- Our goal is to test **whether the null hypothesis or(the alternative) is true** based on the sample data.
- There are two strategies for testing hypotheses:
  - Prove **positively** by demonstrating the null hypothesis is **true**.
  - Prove **negatively** by demonstrating the null hypothesis is **false**.
- For example, in a simple case:
  - Null hypothesis: *All sheep are white*
  - Alternative hypothesis: *Not all sheep are white*
  - We can **reject the null hypothesis** if we find just *one sheep* that is not white.

The Principle of Falsification(证伪)

- Karl Popper (1902-1994), an Austrian philosopher of science renowned for his principle of falsification.
- From a philosophical and logical standpoint, it is significantly easier to prove something false than to prove it true.
- The principle of falsification serves as the standard for distinguishing scientific from non-scientific approaches in research methodology.

Review: Hypothesis Testing

- Now, in a world of uncertainty, we never know the true value of the parameter.
    - "Never say Never"
- Instead, we can say:
    - **reject the null hypothesis** in some level of confidence or
    - **fail to reject the null hypothesis** in some level of confidence.
- In econometrics, our goal is often **to reject the null hypothesis**, as this provides strong evidence in support of the alternative hypothesis.

Review: Two Type Errors

- A certain risk that our conclusion is wrong:

	H_0 is true(H_A is false)	H_0 is false(H_A is true)
Fail to reject H_0		
Reject H_0		

- **Type I error** : Rejecting the null hypothesis when it is actually true.
- **Type II error**: Failing to reject the null hypothesis when it is actually false.

- Both types of errors are **inversely** related - as you decrease the probability of one type of error, you typically increase the probability of the other.
- The **trade-off** between **Type I** and **Type II** errors cannot be *eliminated* simply by increasing sample size, though larger samples can help reduce both to some extent.

Review: Two Type Errors

- A certain risk that our conclusion is wrong:

	H_0 is true(H_A is false)	H_0 is false(H_A is true)
Fail to reject H_0		
Reject H_0	Type I error	

- **Type I error** : Rejecting the null hypothesis when it is actually true.
- **Type II error**: Failing to reject the null hypothesis when it is actually false.

- Both types of errors are **inversely** related - as you decrease the probability of one type of error, you typically increase the probability of the other.
- The **trade-off** between **Type I** and **Type II** errors cannot be *eliminated* simply by increasing sample size, though larger samples can help reduce both to some extent.

Review: Two Type Errors

- A certain risk that our conclusion is wrong:

	H_0 is true(H_A is false)	H_0 is false(H_A is true)
Fail to reject H_0		Type II error
Reject H_0	Type I error	

- **Type I error** : Rejecting the null hypothesis when it is actually true.
- **Type II error**: Failing to reject the null hypothesis when it is actually false.

- Both types of errors are **inversely** related - as you decrease the probability of one type of error, you typically increase the probability of the other.
- The **trade-off** between **Type I** and **Type II** errors cannot be *eliminated* simply by increasing sample size, though larger samples can help reduce both to some extent.

Review: Two Type Errors

- A certain risk that our conclusion is wrong:

	H_0 is true(H_A is false)	H_0 is false(H_A is true)
Fail to reject H_0	Correct(True Negative)	Type II error
Reject H_0	Type I error	

- Type I error : Rejecting the null hypothesis when it is actually true.
- Type II error: Failing to reject the null hypothesis when it is actually false.

- Both types of errors are inversely related - as you decrease the probability of one type of error, you typically increase the probability of the other.
- The **trade-off** between **Type I** and **Type II** errors cannot be *eliminated* simply by increasing sample size, though larger samples can help reduce both to some extent.

Review: Two Type Errors

- A certain risk that our conclusion is wrong:

	H_0 is true(H_A is false)	H_0 is false(H_A is true)
Fail to reject H_0	Correct(True Negative)	Type II error
Reject H_0	Type I error	Correct(True Positive)

- **Type I error** : Rejecting the null hypothesis when it is actually true.
- **Type II error**: Failing to reject the null hypothesis when it is actually false.

- Both types of errors are **inversely** related - as you decrease the probability of one type of error, you typically increase the probability of the other.
- The **trade-off** between **Type I** and **Type II** errors cannot be *eliminated* simply by increasing sample size, though larger samples can help reduce both to some extent.

Review: Hypothesis Testing in Justice Systems

- In the criminal justice system, the principle of **innocent until proven guilty(疑罪从无)** is applied.
  - The jury(陪审团) or judge(法官) begins with the **null hypothesis** that *the accused person is innocent*.
  - The prosecutor(检察官) asserts that the accused person is *guilty* and must present compelling evidence, which represents the **alternative hypothesis**.
  - The defendant’s lawyer(辩护律师) don’t need to prove the innocence, but to **disprove the guilt** or **cast doubt** on the evidence presented by the prosecutor.
  - The jury or judge must **reject the null hypothesis with substantial evidence** in order to convict the accused person.
- Why is the legal system structured this way?

Review: Hypothesis Testing in Justice Systems

- Every trial faces two types of potential errors:

Trial outcome	The defendant is innocent(H_0)	The defendant is guilty(H_A)
Guilty verdict (reject H_0)	Type I error	Correct(True Positive)
Not guilty verdict (fail to reject H_0)	Correct(True Negative)	Type II error

- Justice systems in most countries place **greater weight** on avoiding **Type I errors** than Type II errors:
  - *"Convicting an innocent person"* is considered much more detrimental to society than *"Allowing a guilty person to go free"*.

Review: Hypothesis Testing in Social Science

- Similarly, in social science we follow the presumption of insignificance until proven otherwise.
  - Initially, researchers must assume that the independent variable has zero impact on the dependent variable (the null hypothesis).
  - To establish a relationship, we need to provide compelling evidence that is strong enough to convince readers or policy makers to reject the null hypothesis of no effect.
- Therefore,we weight the two types of errors differently in social science,
  - **Type I error** is more serious than **Type II error**.
- "大胆假设，小心求证"——胡适 (1891-1962).

The Significance level(显著性水平)

- The **significance level** or **size of a test**, α, is the **maximum probability** of **the Type I Error** that we tolerate.

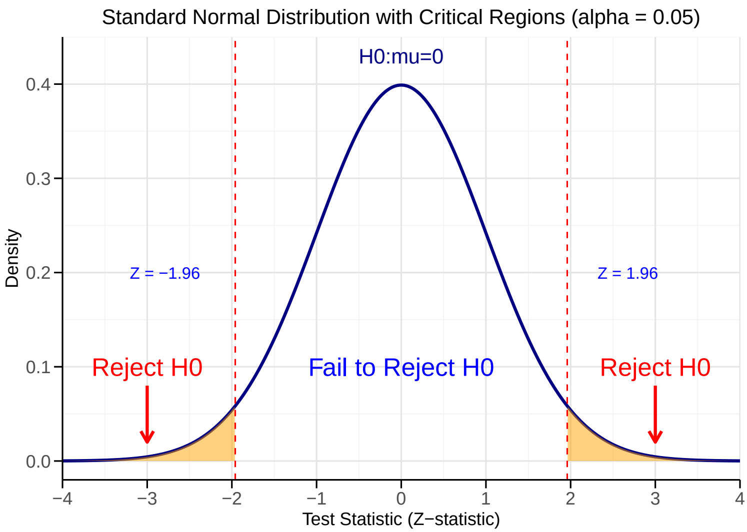
$$P(Type\ I\ error) = P(reject\ H_0 \mid H_0\ is\ true) = \alpha$$

- The usual significance level is set at 5% in social sciences. A less rigorous standard is 10%, whereas a more stringent one is 1%.

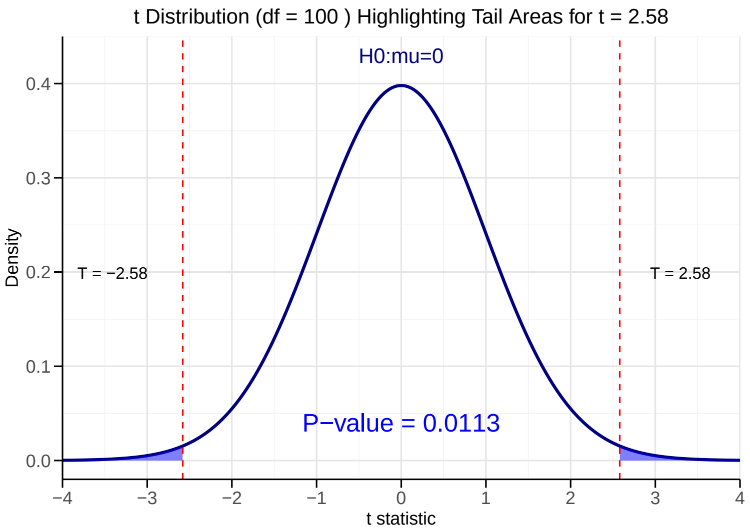
The Sampling Distribution in Hypothesis Testing

- How to calculate the likelihood of **Type-I error** for a given significance level?
  - We have to use the **sampling distribution** of the **test statistic** would be if the null were true.
- The **sampling distribution** of a **test statistic** is its distribution across repeated samples of the same size from the same population.
- Two methods to finish the hypothesis testing:
  - **critical value** is actually a **criteria** calculated by significance level and hypothesis value to make the judgement:
  - If the test statistic is **greater** than the critical value, we **reject** the null hypothesis.
  - **p-value** is the probability of observing a test statistic as **extreme** as the one computed from the sample data, assuming the null hypothesis is true.
  - If the p-value is **less than** the significance level, we **reject** the null hypothesis.

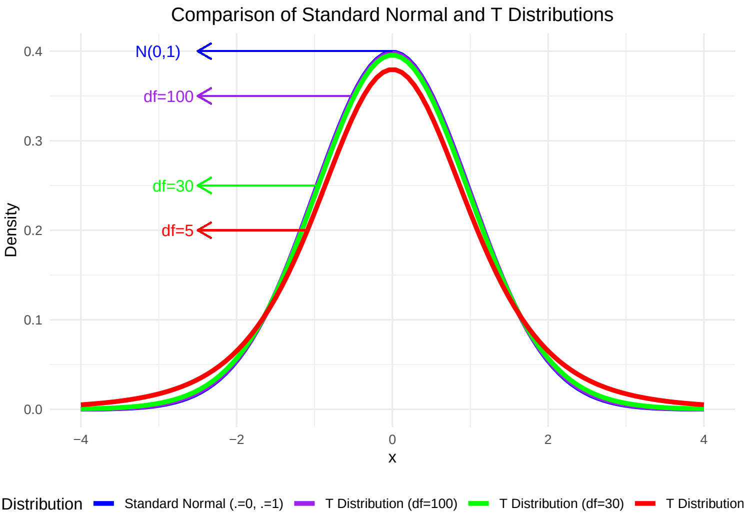
The Sampling Distribution and the Critical Value



The Sampling Distribution and the P Value



T and Standard Normal Distributions



Review: Hypothesis Testing of Population Mean

- **Question**: how to test the population mean of a random variable Y,thus $E(Y)$,by using a sample?
- Let $\mu_{Y,c}$ is a specific value to which the population mean equals(thus we suppose)
  - **the null hypothesis**:
$$H_0 : E(Y) = \mu_{Y,c}$$
  - **the alternative hypothesis(two-sided)**:
$$H_1 : E(Y) \neq \mu_{Y,c}$$

Review: Hypothesis Testing of Population Mean

- **Step 1** Compute the *sample mean* $\overline{Y}$
- **Step 2** Compute the *standard error* of $\overline{Y}$, recall

$$SE(\overline{Y}) = \frac{s_Y}{\sqrt{n}}$$

- **Step 3** Compute the *t-statistic* actually computed

$$t^{act} = \frac{\bar{Y}^{act} - \mu_{Y,c}}{SE(\bar{Y})}$$

- **Alternative Step 3** Compute the p-value

$$\text{p-value} = Pr_{H_0}(|t| > t^{act}) = 2\Phi(-|t^{act}|)$$

- **Step 4** See if we can **Reject the null hypothesis** at a certain significance level α,like 5%, or p-value is less than significance level.

$$|t^{act}| > critical\ value \text{ or } p - value < significance\ level$$

Hypotheses Testing in OLS Regressions

Hypotheses Testing in a Simple OLS

- A Simple OLS regression

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- The key **unknown population parameters** in the population regression equation is β_1.
- We then test whether β_1 equals to a specific value $\beta_{1,c}$ or not
  - **the null hypothesis**:

$$H_0 : \beta_1 = \beta_{1,c}$$

  - **the alternative hypothesis**:

$$H_1 : \beta_1 \neq \beta_{1,c}$$

Hypotheses Testing in a Simple OLS

- **Step1: Estimate $Y_i = \beta_0 + \beta_1 X_i + u_i$ by OLS to obtain $\hat{\beta}_1$**
- **Step2: Compute the *standard error* of $\hat{\beta}_1$**
- **Step3: Construct the *t-statistic***

$$t^{act} = \frac{\hat{\beta}_1 - \beta_{1,c}}{SE\left(\hat{\beta}_1\right)}$$

- **Step4: *Reject the null hypothesis* if**

$$\mid t^{act} \mid > critical\ value$$
$$or \quad p - value < significance\ level$$

The t-statistic v.s Z-statistic

- The statistic we use here is still the **t-statistic** rather than the **Z-statistic**. Why?
  - We can prove that

$$t^{act} = \frac{\hat{\beta}_1 - \beta_{1,c}}{SE\left(\hat{\beta}_1\right)} \sim t(n-2)$$

    given OLS assumptions plus one additional assumption: u_i is **normally distributed**. (If you're interested, you can prove this by yourself.)
- This means that when the sample size is **small**, there is a meaningful difference between the t-statistic and the Z-statistic.

The t-statistic v.s Z-statistic

- We have previously shown that the OLS estimator is **asymptotically normal** when the sample size is large,
  - This means we could theoretically use the Z-statistic instead of the t-statistic.
  - When **the sample size is large**, the difference between the t-statistic and the Z-statistic becomes negligible, as the t-distribution converges to the normal distribution.
- In practice, statisticians and econometricians typically use the t-statistic rather than the Z-statistic in most regression analyses, regardless of sample size.

The t-statistic in a Simple OLS

- The formula for the **t-statistic** is:

$$t^{act} = \frac{\hat{\beta}_1 - \beta_{1,c}}{SE\left(\hat{\beta}_1\right)}$$

or

$$t^{act} = \frac{estimator - hypothesized\ value}{standard\ error\ of\ the\ estimator}$$

- The key unknown component in this calculation is the **standard error** (S.E).

The Standard Error of $\hat{\beta}_1$

- **Recall** from the **Simple OLS Regression**
  - if the least squares assumptions hold, then in large samples $\hat{\beta}_0$ and $\hat{\beta}_1$ have a joint normal sampling distribution,thus $\hat{\beta}_1$

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1})$$

  - We also derived the form of the variance of the normal distribution, $\sigma^2_{\hat{\beta}_1}$ is

$$\sigma_{\hat{\beta}_1} = \sqrt{\frac{1}{n}\frac{Var[(X_i - \mu_X)u_i]}{[Var(X_i)]^2}} \tag{4.21}$$

- The value of $\sigma_{\hat{\beta}_1}$ is **unknown** and can not be obtained *directly* by the data.
  - $Var[(X_i - \mu_X)u_i]$ and $[Var(X_i)]^2$ are both unknown.

The Standard Error of $\hat{\beta}_1$

- However, we can use sample statistics to estimate $\sigma_{\hat{\beta}_1}$. For detailed derivation, see Appendix.
- The **standard error** of $\hat{\beta}_1$ is an **estimator** of the standard deviation of the sampling distribution $\sigma_{\hat{\beta}_1}$, thus

$$SE\left(\hat{\beta}_1\right) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}} = \sqrt{\frac{1}{n} \times \frac{\frac{1}{n-2}\sum(X_i - \bar{X})^2 \hat{u}_i^2}{\left[\frac{1}{n}\sum(X_i - \bar{X})^2\right]^2}} \tag{5.4}$$

- Everything in the equation (5.4) are known now or can be obtained by calculation.
- Now we can construct a *t-statistic* and then make a hypothesis test.

Application to Test Score and Class Size

```
. regress test_score class_size, robust

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

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

- the OLS regression line

$$\widehat{TestScore} = \underset{(10.4)}{698.9} - \underset{(0.52)}{2.28} \times STR, \; R^2 = 0.051, SER = 18.6$$

Testing a two-sided hypothesis concerning β_1

- **the null hypothesis** $H_0 : \beta_1 = 0$
  - It means that the class size will **not** affect the performance of students.
- **the alternative hypothesis** $H_1 : \beta_1 \neq 0$
  - It means that the class size **do** affect the performance of students (whatever positive or negative)
- Our primary goal is to **Reject the null**, and then make a conclusion:
  - Class Size **does matter** for the performance of students.

Testing a two-sided hypothesis concerning β_1

- **Step1: Estimate $\hat{\beta}_1 = -2.28$**
- **Step2: Compute the standard error: $SE(\hat{\beta}_1) = 0.52$**
- **Step3: Compute the *t-statistic***

$$t^{act} = \frac{\hat{\beta}_1 - \beta_{1,c}}{SE\left(\hat{\beta}_1\right)} = \frac{-2.28 - 0}{0.52} = -4.39$$

- **Step4: Reject the null hypothesis if**
  - $\mid t^{act} \mid = \mid -4.39 \mid > critical\ value = 1.96$
  - $p - value = 0 < significance\ level = 0.05$

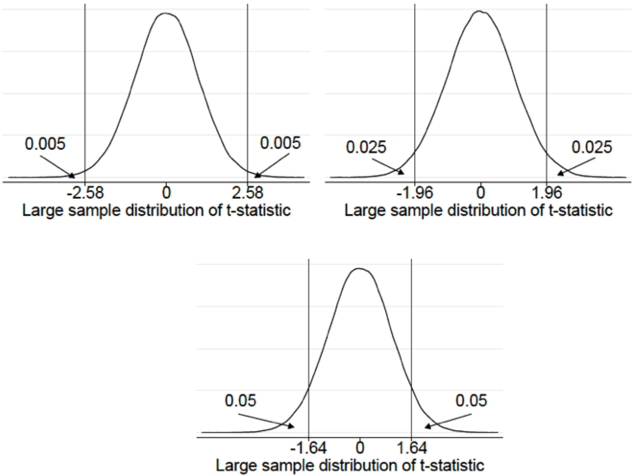
Application to Test Score and Class Size

```
. regress test_score class_size, 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
  test_score |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  class_size |  -2.279808   .5194892    -4.39   0.000    -3.300945   -1.258671
       _cons |    698.933   10.36436    67.44   0.000     678.5602    719.3057
------------------------------------------------------------------------------
```

- **We can reject the null hypothesis that $H_0 : \beta_1 = 0$, which means $\beta_1 \neq 0$ with a high probability(over 95%).**
- It suggests that Class size **matters** the students' performance in a very high chance.

Critical Values of the t-statistic

The critical value of t-statistic depends on significance level α



1% and 10% significant levels

- Step4: Reject the null hypothesis at a **10%** significance level
  - $\mid t^{act} \mid = \mid -4.39 \mid > critical\ value = 1.64$
  - $p - value = 0.00 < significance\ level = 0.1$
- Step4: Reject the null hypothesis at a **1%** significance level
  - $\mid t^{act} \mid = \mid -4.39 \mid > critical\ value = 2.58$
  - $p - value = 0.00 < significance\ level = 0.01$

Two-Sided Hypotheses: β_1 in a certain value

- **Step1: Estimate** $\hat{\beta}_1 = -2.28$
- **Step2: Compute the standard error:** $SE(\hat{\beta}_1) = 0.52$
- **Step3: Compute the t-statistic**

$$t^{act} = \frac{\hat{\beta}_1 - \beta_{1,c}}{SE\left(\hat{\beta}_1\right)} = \frac{-2.28 - (-2)}{0.52} = -0.54$$

- **Step4: can't reject the null hypothesis at 5% significant level because**
  - $\mid t^{act} \mid = \mid -0.54 \mid < critical\ value = 1.96$
  - $p - value = 0.59 > significance\ level = 0.05$

Two-Sided Hypotheses : β_1 in a certain value

```
. lincom class_size-(-2)

 ( 1)  class_size = -2
```

test_score	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	-.2798083	.5194892	-0.54	0.590	-1.300945	.7413286

- We cannot reject the null hypothesis that $H_0 : \beta_1 = -2$.
- It suggests that *there is no enough evidence* to support the statement:
  - cutting class size in one unit will boost the test score in 2 points.

One-sided Hypotheses Concerning β_1

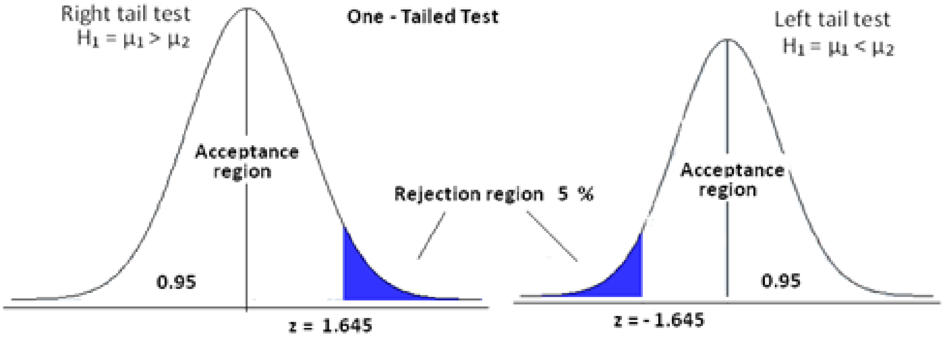
- Sometimes, we want to do a *one-sided Hypothesis testing*
- the null hypothesis is still unchanged $H_0 : \beta_1 = -2$
- **the alternative hypothesis** is $H_1 : \beta_1 < -2$
  - The statement is that reducing(or inversely increasing) class size will boost(or lower) student's performance.
  - More specifically,cutting class size in one unit will increase the test score in 2 points **at least**.
- Because the null hypothesis is the same for a one- and a two-sided hypothesis test, the construction of the t-statistic is the same.
- The difference between the two is the critical value and p-value.

One-sided Hypotheses Concerning β_1

- **Step1: Estimate** $\hat{\beta}_1 = -2.28$
- **Step2: Compute the standard error:** $SE(\hat{\beta}_1) = 0.52$
- **Step3: Compute the t-statistic**

$$t^{act} = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE\left(\hat{\beta}_1\right)} = \frac{-2.28 - (-2)}{0.52} = -0.54$$

One-sided Hypotheses Concerning β_1



One-sided Hypotheses Concerning β_1

- Step4: under the circumstance, the critical value is not the -1.96 but -1.645 at 5% significant level.
- We can't reject the null hypothesis because

$$t^{act} = -0.54 > critical\ value = -1.645$$

- The p-value is not the $2\Phi(-|t^{act}|)$ now but $Pr(Z < t^{act}) = \Phi(t^{act})$.
- It suggests that *there is NO enough evidence* to support the statement:cutting class size in one unit will increase the test score in **2 points at least**.

One-sided Hypotheses Concerning β_1

- One-sided alternative hypotheses should be used only when there is a clear reason for doing so.
- This reason could come from economic theory, prior empirical evidence, or both.
- However, even if it initially seems that the relevant alternative is one-sided, upon reflection this might not necessarily be so.
- In practice, one-sided test is used much less than two-sided test.

Wrap up

- Hypothesis tests are useful if you have a specific null hypothesis in mind.
- Being able to accept or reject this null hypothesis based on the statistical evidence provides a powerful tool for coping with the uncertainty inherent in **using a sample to learn about the population**.
- Yet, there are many times that no single hypothesis about a regression coefficient is dominant, and instead one would like to know a range of values of the coefficient that are consistent with the data.
- This calls for constructing a **confidence interval**.

Confidence Intervals

Introduction

- Because any statistical estimate of the slope β_1 necessarily has sampling uncertainty, we cannot determine the true value of β_1 exactly from a sample of data.
- It is possible, however, to use the OLS estimators and its standard error to construct a confidence interval for the slope β_1

CI for β_1

- Method for constructing a confidence interval for a population mean can be easily extended to constructing a confidence interval for a regression coefficient.
- Using a two-sided test, a hypothesized value for β_1 will be rejected at 5% significance level if

$$\mid t^{act} \mid > critical\ value = 1.96$$

- So $\hat{\beta}_1$ will be in the *confidence set* if $\mid t^{act} \mid \leq critical\ value = 1.96$
- Thus the 95% confidence interval for β_1 are within ± 1.96 standard errors of $\hat{\beta}_1$

$$\hat{\beta}_1 \pm 1.96 \cdot SE\left(\hat{\beta}_1\right)$$

CI for $\beta_{ClassSize}$

```
. regress test_score class_size, robust

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

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

CI for $\beta_{ClassSize}$

```
. regress test_score class_size, 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
  test_score |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  class_size |  -2.279808   .5194892    -4.39   0.000    -3.300945   -1.258671
       _cons |    698.933   10.36436    67.44   0.000     678.5602    719.3057
------------------------------------------------------------------------------
```

- Thus the 95% confidence interval for β_1 are within ± 1.96 standard errors of $\hat{\beta}_1$

$$\hat{\beta}_1 \pm 1.96 \cdot SE\left(\hat{\beta}_1\right) = -2.28 \pm (1.96 \times 0.519) = [-3.3, -1.26]$$

CI for predicted effets of changing X

- Consider changing X by a given amount,ΔX. The predicted change in Y associated with this change in X is $\beta_1 \Delta$.
- the 95% confidence interval for $\beta_1 \Delta X$ is

$$\hat{\beta}_1 \Delta X \pm 1.96 \cdot SE\left(\hat{\beta}_1\right) \times \Delta X$$

- eg reducing the student–teacher ratio by 2. then the 95% confidence interval is

$$[-3.3 \times 2, -1.34 \times 2] = [-6.6, -2.68]$$

Gauss-Markov theorem and Heteroskedasticity

Introduction

- Recall we discussed the properties of $\bar{Y}$ in Chapter 2.
  - an **unbiased** estimator of μ_Y
  - a **consistent** estimator of μ_Y
  - an **approximate normal sampling distribution** for large n

The Efficiency of $\bar{Y}$

- the fourth properties of $\bar{Y}$ in Chapter 3.
- the **Best Linear Unbiased Estimator(BLUE)**: $\bar{Y}$ is the most efficient estimator of μ_Y among all unbiased estimators that are weighted averages of $Y_1, ..., Y_n$, presented by $\hat{\mu}_Y = \frac{1}{n}\sum a_i Y_i$,thus,

$$Var(\overline{Y}) < Var(\hat{\mu}_Y)$$

Unnecessary Assumption for Simple OLS

- Three Simple OLS Regression Assumptions
  - Assumption 1
  - Assumption 2
  - Assumption 3
- Assumption 4: The error terms are **homoskedastic**

$$Var(u_i \mid X_i) = \sigma_u^2$$

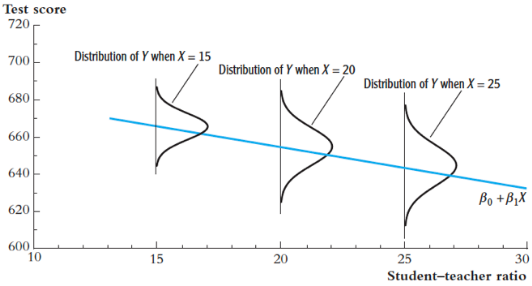
- Then $\hat{\beta}^{OLS}$ is the **Best Linear Unbiased Estimator(BLUE)**: it is the most efficient estimator of β_1 among all conditional unbiased estimators that are a linear function of $Y_1, Y_2, ..., Y_n$.

Heteroskedasticity & homoskedasticity

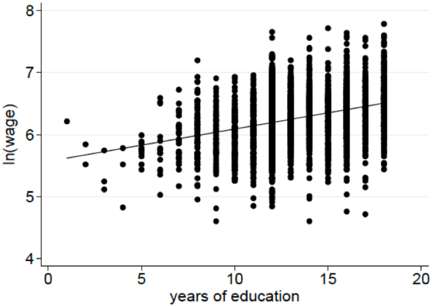
- The error term u_i is **homoskedastic** if the variance of the conditional distribution of u_i given X_i is constant for $i = 1, \ldots n$, in particular does not depend on X_i.
- Otherwise, the error term is **heteroskedastic**.

FIGURE 5.2 An Example of Heteroskedasticity

Like Figure 4.4, this shows the conditional distribution of test scores for three different class sizes. Unlike Figure 4.4, these distributions become more spread out (have a larger variance) for larger class sizes. Because the variance of the distribution of u given X, var(u|X), depends on X, u is heteroskedastic.



An Actual Example: the returns to schooling



- The spread of the dots around the line is clearly increasing with years of education X_i.
- Variation in (log) wages is higher at higher levels of education.
- This implies that

$$Var(u_i \mid X_i) \neq \sigma_u^2$$

Homoskedasticity: S.E.

- **Recall the standard deviation of β_1, $\sigma^2_{\hat{\beta}_1}$, is**

$$\sigma_{\hat{\beta}_1} = \sqrt{\frac{1}{n}\frac{Var[(X_i - \mu_X)u_i]}{[Var(X_i)]^2}} \tag{4.21}$$

- **If u_i is homoskedastic, thus**

$$Var(u_i|X_i) = \sigma_u^2$$

Homoskedasticity: S.E.

- The *numerator* in the square root in (4.21) can be transformed into

$$Var[(X_i - \mu_X)u_i]$$

Homoskedasticity: S.E.

- The *numerator* in the square root in (4.21) can be transformed into

$$Var[(X_i - \mu_X)u_i] = E[(X_i - \mu_X)u_i]^2 - (E[(X_i - \mu_X)u_i])^2$$

Homoskedasticity: S.E.

- The *numerator* in the square root in (4.21) can be transformed into

$$\begin{aligned} Var[(X_i - \mu_X)u_i] &= E[(X_i - \mu_X)u_i]^2 - (E[(X_i - \mu_X)u_i])^2 \\ &= E[(X_i - \mu_X)u_i]^2 \end{aligned}$$

Homoskedasticity: S.E.

- The *numerator* in the square root in (4.21) can be transformed into

$$\begin{aligned} Var[(X_i - \mu_X)u_i] &= E[(X_i - \mu_X)u_i]^2 - (E[(X_i - \mu_X)u_i])^2 \\ &= E[(X_i - \mu_X)u_i]^2 \\ &= E[(X_i - \mu_X)^2 E(u_i^2|X_i)] \end{aligned}$$

Homoskedasticity: S.E.

- The *numerator* in the square root in (4.21) can be transformed into

$$
\begin{aligned}
Var[(X_i - \mu_X)u_i] &= E[(X_i - \mu_X)u_i]^2 - (E[(X_i - \mu_X)u_i])^2 \\
&= E[(X_i - \mu_X)u_i]^2 \\
&= E[(X_i - \mu_X)^2 E(u_i^2|X_i)] \\
&= E[(X_i - \mu_X)^2 Var(u_i|X_i)]
\end{aligned}
$$

Homoskedasticity: S.E.

- The *numerator* in the square root in (4.21) can be transformed into

$$
\begin{aligned}
Var[(X_i-\mu_X)u_i] &= E[(X_i-\mu_X)u_i]^2 - (E[(X_i-\mu_X)u_i])^2 \\
&= E[(X_i-\mu_X)u_i]^2 \\
&= E[(X_i-\mu_X)^2 E(u_i^2|X_i)] \\
&= E[(X_i-\mu_X)^2 Var(u_i|X_i)] \\
&= \sigma_u^2 E[(X_i-\mu_X)^2]
\end{aligned}
$$

Homoskedasticity: S.E.

- **Then the equation (4.21) turns into**

$$\sigma_{\hat{\beta}_1}$$

Homoskedasticity: S.E.

- Then the equation (4.21) turns into

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

Homoskedasticity: S.E.

- **Then the equation (4.21) turns into**

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

Homoskedasticity: S.E.

- Then the equation (4.21) turns into

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

- So if we assume that the error terms are **homoskedastic**, then the **standard errors** of the OLS estimators β_1 simplify to

$$SE_{Homo}\left(\hat{\beta}_1\right) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}} = \sqrt{\frac{s^2_{\hat{u}}}{\sum(X_i - \bar{X})^2}}$$

The Standard Error of the Regression

- We would also like to know the standard deviation of u_i,thus σ_u^2. However, u_i are totally unobserved. We have to use the sample statistic to inference the population.
- The **standard error of the regression** (SER) is an *estimator* of the standard deviation of the regression error u_i.
- The **SER** is computed using their sample counterparts, the OLS residuals $\hat{u}_i$,thus

$$SER = s_{\hat{u}} = \sqrt{s_{\hat{u}}^2}$$

where $s_{\hat{u}}^2 = \frac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2$

- Think about it: why the denominator is $n-2$?

Homoskedasticity: S.E.

- However,in many applications homoskedasticity is **NOT a plausible assumption**.
- If the error terms are *heteroskedastic*, then you use the *homoskedastic* assumption to compute the S.E. of $\hat{\beta}_1$. It will leads to
  - The standard errors are wrong (often too small)
  - The t-statistic does NOT have a $N(0,1)$ distribution (also not in large samples).
  - But the estimating coefficients in OLS regression will not *change*.

Heteroskedasticity & homoskedasticity

- If the error terms are **heteroskedastic**, we should use the original equation of S.E.

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

- It is called as *heteroskedasticity robust-standard errors*,also referred to as Eicker-Huber-White standard errors,simply **Robust-Standard Errors**
- In the case, it is not difficult to find that *homoskedasticity* **is just a special case of** ***heteroskedasticity***.

Heteroskedasticity & homoskedasticity

- Since homoskedasticity is a special case of heteroskedasticity, these heteroskedasticity robust formulas are also **valid** if *the error terms are homoskedastic*.
- Hypothesis tests and confidence intervals based on above SE's are *valid* both in case of homoskedasticity and heteroskedasticity.
- In reality, since in many applications homoskedasticity is not a plausible assumption, *it is best to use heteroskedasticity robust standard errors*. Using **robust standard errors** rather than **standard errors with homoskedasticity** will lead us **lose nothing**.

Heteroskedasticity & homoskedasticity

- **It can be quite cumbersome to do this calculation by hand.Luckily,computer can help us do the job.**
  - **In `Stata`, the default option of regression is to assume homoskedasticity, to obtain heteroskedasticity robust standard errors use the option "robust":**

$$regress\ y\ x\ ,\ robust$$

  - **In `R`, many ways can finish the job. A convenient function named `vcovHC()` is part of the package `sandwich`.**

Test Scores and Class Size

```
. regress test_score class_size

      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

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

. regress test_score class_size, 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
  test_score |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  class_size |  -2.279808   .5194892    -4.39   0.000    -3.300945   -1.258671
       _cons |    698.933   10.36436    67.44   0.000     678.5602    719.3057
------------------------------------------------------------------------------
```

Test Scores and Class Size

```
. regress test_score class_size

      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

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

```
. regress test_score class_size, 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
  test_score |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  class_size |  -2.279808   .5194892    -4.39   0.000    -3.300945   -1.258671
       _cons |    698.933   10.36436    67.44   0.000     678.5602    719.3057
------------------------------------------------------------------------------
```

Wrap up: Heteroskedasticity in a Simple OLS

- If the error terms are heteroskedastic
  - The fourth simple OLS assumption is violated.
  - The Gauss-Markov conditions do not hold.
  - The OLS estimator is not BLUE (not the most efficient).
- But (given that the other OLS assumptions hold)
  - The OLS estimators are still *unbiased*.
  - The OLS estimators are still *consistent*.
  - The OLS estimators are *normally distributed* in large samples

OLS with Multiple Regressors: Hypotheses tests

Recall: the Multiple OLS Regression

- The multiple regression model is

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + ... + \beta_k X_{k,i} + u_i, i = 1, ..., n$$

- Four Basic Assumptions
  - Assumption 1 : $E[u_i \mid X_{1i}, X_{2i} ..., X_{ki}] = 0$
  - Assumption 2 : i.i.d sample
  - Assumption 3 : Large outliers are unlikely.
  - Assumption 4 : No perfect multicollinearity.
- The Sampling Distribution: the OLS estimators $\hat{\beta}_j$ for $j = 1, ..., k$ are approximately normally distributed in large samples.

Standard Errors for the Multiple OLS Estimators

- There is *nothing* conceptually different between the single- or multiple-regressor cases.
  - Standard Errors for a Simple OLS estimator β_1

$$SE\left(\hat{\beta}_1\right) = \hat{\sigma}_{\hat{\beta}_1}$$

  - Standard Errors for Mutiple OLS Regression estimators β_j

$$SE\left(\hat{\beta}_j\right) = \hat{\sigma}_{\hat{\beta}_j}$$

- Remind: since now the joint distribution is not only for (Y_i, X_i),but also for (X_{ij}, X_{ik}).
- The formula for the *standard errors* in Multiple OLS regression are related with a *matrix* named **Variance-Covariance matrix**.

Hypothesis Tests for a Single Coefficient

- **the *t-statistic* in Simple OLS Regression**

$$t_1^{act} = \frac{\hat{\beta}_1 - \beta_{1,c}}{SE\left(\hat{\beta}_1\right)} \sim N(0,1)$$

- **the *t-statistic* in Multiple OLS Regression**

$$t_j^{act} = \frac{\hat{\beta}_j - \beta_{j,c}}{SE\left(\hat{\beta}_j\right)} \sim N(0,1)$$

Hypothesis testing for single coefficient

- $H_0 : \beta_j = \beta_{j,c}$ $H_1 : \beta_1 \neq \beta_{j,c}$
- Step1: Estimate $\hat{\beta}_j$, by run a multiple OLS regression

$$Y_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_j X_{ji} + ... + \beta_k X_{ki} + u_i$$

- Step2: Compute the standard error of $\hat{\beta}_j$ (*requires matrix algebra*)
- Step3: Compute the t-statistic

$$t_j^{act} = \frac{\hat{\beta}_j - \beta_{j,c}}{SE\left(\hat{\beta}_j\right)}$$

- Step4: Reject the null hypothesis if
  - $\mid t^{act} \mid > critical\ value$
  - or if $p - value < significance\ level$

Confidence Intervals for a single coefficient

- Also the same as in a simple OLS Regression.
- $\hat{\beta}_j$ will be in the confidence set if $\mid t^{act} \mid \leq critical\ value = 1.96$ at the 95% confidence level.
- Thus the 95% confidence interval for β_j are within ± 1.96 standard errors of $\hat{\beta}_j$

$$\hat{\beta}_j \pm 1.96 \cdot SE\left(\hat{\beta}_j\right)$$

Test Scores and Class Size

```
. regress test_score class_size el_pct,robust

Linear regression                               Number of obs     =        420
                                                F(2, 417)         =     223.82
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4264
                                                Root MSE          =     14.464
```

test_score	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
class_size	-1.101296	.4328472	-2.54	0.011	-1.95213	-.2504616
el_pct	-.6497768	.0310318	-20.94	0.000	-.710775	-.5887786
_cons	686.0322	8.728224	78.60	0.000	668.8754	703.189

Case: Class Size and Test scores

- Does changing class size, while holding the percentage of English learners constant, have a statistically significant effect on test scores? (using a 5% significance level)
- $H_0 : \beta_{ClassSize} = 0 \; H_1 : \beta_{ClassSize} \neq 0$
- Step1: Estimate $\hat{\beta}_1 = -1.10$
- Step2: Compute the standard error: $SE(\hat{\beta}_1) = 0.43$
- Step3: Compute the t-statistic

$$t^{act} = \frac{\hat{\beta}_1 - \beta_{1,c}}{SE\left(\hat{\beta}_1\right)} = \frac{-1.10 - 0}{0.43} = -2.54$$

- Step4: Reject the null hypothesis if
  - $\mid t^{act} \mid = \mid -2.54 \mid > critical\ value.1.96$
  - $p - value = 0.011 < significance\ level = 0.05$

Tests of Joint Hypotheses: on 2 or more coefficients

- Can we just test one individual coefficient at a time?
- Suppose the angry taxpayer hypothesizes that neither the *student–teacher ratio* nor *expenditures per pupil* have an effect on test scores, once we control for the *percentage of English learners*.
- Therefore, we have to test a **joint null hypothesis** that both the coefficient on *student–teacher ratio* and the coefficient on *expenditures per pupil* are zero?

$$H_0 : \beta_{str} = 0 \;\&\; \beta_{expn} = 0,$$
$$H_1 : \beta_{str} \neq 0 \; and/or \; \beta_{expn} \neq 0$$

Testing 1 hypothesis on 2 or more coefficients

- If either t_{str} or t_{expn} exceeds 1.96, should we reject the null hypothesis?
- Assume that t_{str} and t_{expn} are *uncorrelated* at first:

$$
\begin{aligned}
& Pr(|t_{str}| > 1.96 \; and/or \; |t_{expn}| > 1.96) \\
& = 1 - Pr(|t_{str}| \leq 1.96 \; and \; |t_{expn}| \leq 1.96) \\
& = 1 - Pr(|t_{str}| \leq 1.96) * Pr\, |t_{expn}| \leq 1.96) \\
& = 1 - 0.95 \times 0.95 \\
& = 0.0975 > 0.05
\end{aligned}
$$

- We **cannot** reject the null hypothesis at 5% significant level now,even the single `t-test` for both variables can.

Testing 1 hypothesis on 2 or more coefficients

- If t_{str} and t_{expn} are correlated, then *it is more complicated* as simple t-statistic is not enough for hypothesis testing in Multiple OLS.
- In general, a joint hypothesis is a hypothesis that imposes two or more restrictions on the regression coefficients.

$$H_0 : \beta_j = \beta_{j,c}, \beta_k = \beta_{k,c}, ..., for\ a\ total\ of\ q\ restrictions$$

$$H_1 : one\ or\ more\ of\ q\ restrictions\ under\ H_0\ does\ not\ hold$$

  - where $\beta_j, \beta_k, ...$ refer to different regression coefficients.
- When the regressors are highly correlated, we use **F-statistic**to testing joint hypotheses.

Unrestricted v.s Restricted model

- **The unrestricted model**: the model without any of the restrictions imposed. It contains all the variables.
- **The restricted model**: the model on which the restrictions have been imposed.
- And we want to test that $H_0 : \beta_1 = 0 \; and \; \beta_2 = 0$,then $H_1 : \beta_1 \neq 0 \; and/or \; \beta_2 \neq 0$ for the regression model

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + u_i, i = 1, ..., n$$

- Then restricted model is

$$Y_i = \beta_0 + \beta_3 X_{3,i} + u_i$$

The F-statistic with q restrictions

- The F-statistic is computed using a simple formula based on the sum of squared residuals from two regressions.

$$F = \frac{(SSR_{\text{restricted}} - SSR_{\text{unrestricted}})/q}{SSR_{\text{unrestricted}}/(n-k-1)}$$

- $SSR_{restricted}$ is the sum of squared residuals from the **restricted** regression.
- $SSR_{unrestricted}$ is the sum of squared residuals from the **full** model.
- q is the number of restrictions under the null.
- k is the number of regressors in the unrestricted regression.

The F-statistic and R^2

- An alternative equivalent formula for the **homoskedasticity-only F-statistic** is based on the R^2 of the two regressions:

$$F = \frac{(R^2_{\text{restricted}} - R^2_{\text{unrestricted}})/q}{1 - R^2_{\text{unrestricted}}/(n-k-1)}$$

- Only if the error terms are **homoskedastic**

$$Var(u_i \mid X_i) = \sigma_u^2$$

Testing 1 hypothesis on 2 or more coefficients

- Suppose we want to test

$$H_0 : \beta_1 = 0 \ \& \ \beta_2 = 0 \quad H_1 : \beta_1 \neq 0 \ and/or \ \beta_2 \neq 0$$

- Then the *F-statistic* can also combine the two *t-statistics*t_1 and t_2 as follows

$$F = \frac{1}{2}\left(\frac{t_1^2 + t_2^2 - 2\hat{\rho}_{t_1 t_2} t_1 t_2}{1 - \hat{\rho}_{t_1 t_2}^2}\right)$$

where $\hat{\rho}_{t_1 t_2}$ is an estimator of the correlation between the two t-statistics.

Heteroskedasticity-Robust F-statistic

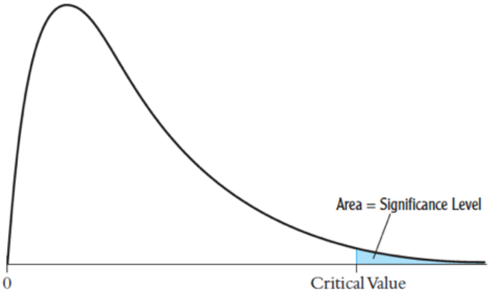
- Using matrix to show the form of the **heteroskedasticity-robust F-statistic** which is *beyond the scope of our class*.
- While,under the null hypothesis,regardless of whether the errors are homoskedastic or heteroskedastic, the F-statistic with q has a sampling distribution in large samples,

$$F - statistic \sim F_{q,\infty}$$

  - where q is the number of restrictions
- Then we can compute the F-statistic, the critical values from the table of the $F_{q,\infty}$ and obtain the p-value.

F-Distribution

TABLE 4 Critical Values for the $F_{m,\infty}$ Distribution



Degrees of Freedom	10%	5%	1%
	Significance Level		
1	2.71	3.84	6.63
2	2.30	3.00	4.61
3	2.08	2.60	3.78

Testing joint hypothesis with q restrictions

- $H_0 : \beta_j = \beta_{j,0}, ..., \beta_m = \beta_{m,0}$ for a total of q restrictions.
- H_1:at least one of q restrictions under H_0 does not hold.
- Step1: Estimate

$$Y_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_j X_{ji} + ... + \beta_k X_{ki} + u_i$$

  by OLS
- Step2: Compute the **F-statistic**
- Step3 : Reject the null hypothesis if

$$F - Statistic > F^{act}_{q,\infty}$$

  or

$$p - value = Pr[F_{q,\infty} > F^{act}] <= significant\ level$$

Case: Class Size and Test Scores

- **We want to test hypothesis that both the coefficient on *student–teacher ratio* and the coefficient on *expenditures per pupil* are zero?**
  - $H_0 : \beta_{str} = 0 \; \& \beta_{expn} = 0$
  - $H_1 : \beta_{str} \neq 0 \; and/or \; \beta_{expn} \neq 0$
- **The null hypothesis consists of two restrictions $q = 2$**

Case: Class Size and Test Scores

```
. regress test_score class_size expn_stu el_pct,robust

Linear regression                               Number of obs     =        420
                                                F(3, 416)         =     147.20
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4366
                                                Root MSE          =     14.353

------------------------------------------------------------------------------
             |               Robust
  test_score |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  class_size |  -.2863992   .4820728    -0.59   0.553    -1.234002     .661203
    expn_stu |   .0038679   .0015807     2.45   0.015     .0007607    .0069751
      el_pct |  -.6560227   .0317844   -20.64   0.000    -.7185008   -.5935446
       _cons |   649.5779   15.45834    42.02   0.000     619.1917    679.9641
------------------------------------------------------------------------------

. test class_size expn_stu

 ( 1)  class_size = 0
 ( 2)  expn_stu = 0

       F(  2,   416) =    5.43
            Prob > F =    0.0047
```

- It can be shown that the F-statistic with two restrictions has an approximate $F_{2,\infty}$ distribution in large samples

$$F_{act} = 5.43 > F_{2,\infty} = 4.61 \ at\ 1\%\ significant\ level$$

The “overall” regression F-statistic

- The “overall” F-statistic test the joint hypothesis that all the k slope coefficients are zero
  - $H_0 : \beta_j = \beta_{j,0}, ..., \beta_m = \beta_{m,0}$ for a total of $q = k$ restrictions.
  - H_1: at least one of $q = k$ restrictions under H_0 does not hold.

The "overall" regression F-statistic

```
. regress test_score class_size expn_stu el_pct,robust

Linear regression                               Number of obs     =        420
                                                F(3, 416)         =     147.20
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4366
                                                Root MSE          =     14.353

------------------------------------------------------------------------------
             |               Robust
  test_score |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  class_size |  -.2863992   .4820728    -0.59   0.553    -1.234002     .661203
    expn_stu |   .0038679   .0015807     2.45   0.015     .0007607    .0069751
      el_pct |  -.6560227   .0317844   -20.64   0.000    -.7185008   -.5935446
       _cons |   649.5779   15.45834    42.02   0.000     619.1917    679.9641
------------------------------------------------------------------------------

. test class_size expn_stu el_pct

 ( 1)  class_size = 0
 ( 2)  expn_stu = 0
 ( 3)  el_pct = 0

       F(  3,   416) =  147.20
            Prob > F =    0.0000
```

- The overall $F - Statistics = 147.2$ which indicates at least one coefficient can not be **ZERO**.

Case: Analysis of the Test Score Data Set

Introduction

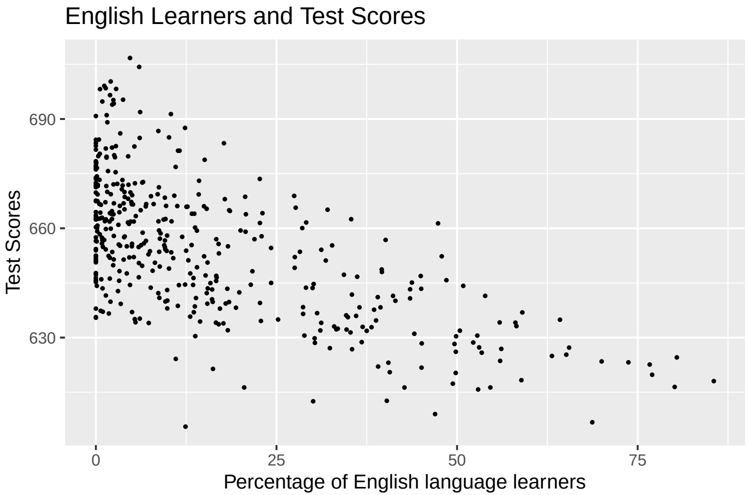
- How to use multiple regression in order to alleviate omitted variable bias and demonstrate how to report results.
- Considering **three variables** that control for unobservable student characteristics which correlate with the student-teacher ratio *and* are assumed to have an impact on test scores:
- $English$, the percentage of English learning students
- $lunch$, the share of students that qualify for a subsidized or even a free lunch at school
- $calworks$,the percentage of students that qualify for a income assistance program

Five different model equations:

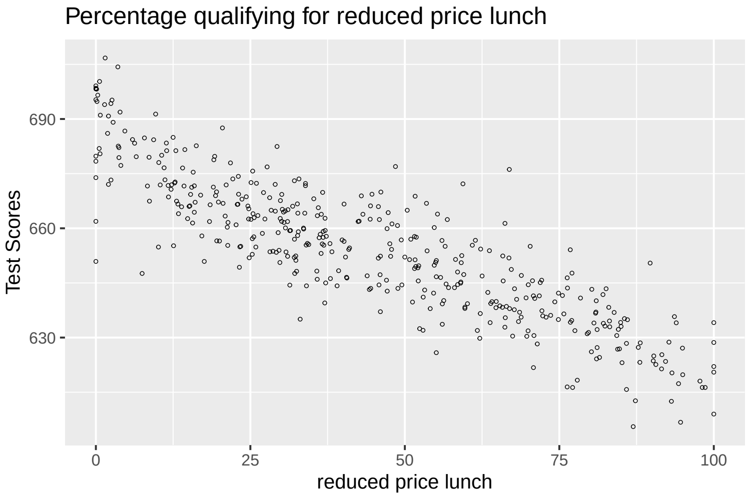
- **We shall consider five different model equations:**

$$
\begin{aligned}
&(1) \quad TestScore = \beta_0 + \beta_1 STR + u, \\
&(2) \quad TestScore = \beta_0 + \beta_1 STR + \beta_2 english + u, \\
&(3) \quad TestScore = \beta_0 + \beta_1 STR + \beta_2 english + \beta_3 lunch + u, \\
&(4) \quad TestScore = \beta_0 + \beta_1 STR + \beta_2 english + \beta_4 calworks + u, \\
&(5) \quad TestScore = \beta_0 + \beta_1 STR + \beta_2 english + \beta_3 lunch + \beta_4 calworks + u
\end{aligned}
$$

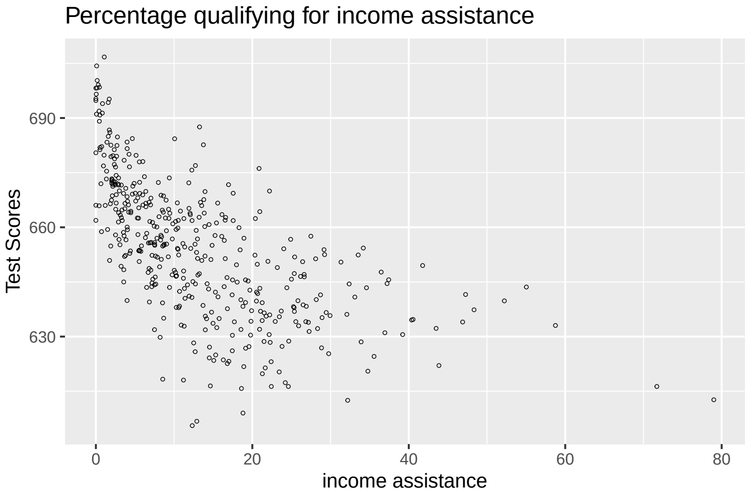
Scatter Plot: English learners and Test Scores



Scatter Plot: Free lunch and Test Scores



Scatter Plot: Income assistant and Test Scores



Correlations between Variables

- **The correlation coefficients are following**

```r
# estimate correlation between student characteristics and test scores
cor(CASchools$testscr, CASchools$el_pct)
```

```
#> [1] -0.6441237
```

```r
cor(CASchools$testscr, CASchools$meal_pct)
```

```
#> [1] -0.868772
```

```r
cor(CASchools$testscr, CASchools$calw_pct)
```

```
#> [1] -0.6268534
```

```r
cor(CASchools$meal_pct, CASchools$calw_pct)
```

```
#> [1] 0.7394218
```

Table 3

	Dependent Variable: Test Score	
	(1)	(2)
str	-2.280^{***}	-1.101^{**}
	(0.519)	(0.433)
el_pct		-0.650^{***}
		(0.031)
Constant	698.933^{***}	686.032^{***}
	(10.364)	(8.728)
Observations	420	420
R^2	0.051	0.426
Adjusted R^2	0.049	0.424
F Statistic	22.575^{***}	155.014^{***}
Note:		$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
		Robust S.E. are shown in the parentheses

Table 4

	Dependent Variable: Test Score			
	(1)	(2)	(3)	(4)
str	-2.280^{***}	-1.101^{**}	-0.998^{***}	-1.308^{***}
	(0.519)	(0.433)	(0.270)	(0.339)
el_pct		-0.650^{***}	-0.122^{***}	-0.488^{***}
		(0.031)	(0.033)	(0.030)
meal_pct			-0.547^{***}	
			(0.024)	
calw_pct				-0.790^{***}
				(0.068)
Constant	698.933^{***}	686.032^{***}	700.150^{***}	697.999^{***}
	(10.364)	(8.728)	(5.568)	(6.920)
Observations	420	420	420	420
R^2	0.051	0.426	0.775	0.629
Adjusted R^2	0.049	0.424	0.773	0.626

Table 5

	Dependent Variable: Test Score				
	(1)	(2)	(3)	(4)	(5)
str	-2.280^{***}	-1.101^{**}	-0.998^{***}	-1.308^{***}	-1.014^{***}
	(0.519)	(0.433)	(0.270)	(0.339)	(0.269)
el_pct		-0.650^{***}	-0.122^{***}	-0.488^{***}	-0.130^{***}
		(0.031)	(0.033)	(0.030)	(0.036)
meal_pct			-0.547^{***}		-0.529^{***}
			(0.024)		(0.038)
calw_pct				-0.790^{***}	-0.048
				(0.068)	(0.059)
Constant	698.933^{***}	686.032^{***}	700.150^{***}	697.999^{***}	700.392^{***}
	(10.364)	(8.728)	(5.568)	(6.920)	(5.537)
Observations	420	420	420	420	420
R^2	0.051	0.426	0.775	0.629	0.775
Adjusted R^2	0.049	0.424	0.773	0.626	0.773

The "Star War" and Regression Table

Dependent variable: average test score in the district.

Regressor	(1)	(2)	(3)	(4)	(5)
Student–teacher ratio (X_1)	−2.28** (0.52)	−1.10* (0.43)	−1.00** (0.27)	−1.31* (0.34)	−1.01* (0.27)
Percent English learners (X_2)		−0.650** (0.031)	−0.122** (0.033)	−0.488** (0.030)	−0.130** (0.036)
Percent eligible for subsidized lunch (X_3)			−0.547* (0.024)		−0.529* (0.038)
Percent on public income assistance (X_4)				−0.790** (0.068)	0.048 (0.059)
Intercept	698.9** (10.4)	686.0** (8.7)	700.2** (5.6)	698.0** (6.9)	700.4** (5.5)
Summary Statistics					
SER	18.58	14.46	9.08	11.65	9.08
$\bar{R}^2$	0.049	0.424	0.773	0.626	0.773
n	420	420	420	420	420

These regressions were estimated using the data on K–8 school districts in California, described in Appendix (4.1). Heteroskedasticity-robust standard errors are given in parentheses under coefficients. The individual coefficient is statistically significant at the *5% level or **1% significance level using a two-sided test.

Discussion of the empirical results

- We should focus on the coefficient of our main interest, the student-teacher ratio (STR), in the regression table.
  - Though we estimate the effect of STR on test scores in different specifications, the coefficient of STR is consistently negative and statistically significant at around from -1 to -1.3.
- We should explain the results in the context of the research question.
  1. The sign of the coefficient
  2. The magnitude of the coefficient
  3. The statistical significance of the coefficient
  4. The economic significance of the coefficient

Warp Up

- Therefore, we have to build a framework to test the hypothesis about the population parameters based on the sample given a certain level of confidence.
- Using the hypothesis testing and confidence interval in OLS regression, we could make a more reliable judgment about the relationship between the treatment and the outcomes.
- The analysis in this and the preceding lectures has presumed that the population regression function is linear in the regressor which might not be true.
  - We will extend the model into nonlinearity in the next lectures.

Appendix

The Standard Error of $\hat{\beta}_1$

- if the least squares assumptions hold, then in large samples $\hat{\beta}_0$ and $\hat{\beta}_1$ have a joint normal sampling distribution,thus $\hat{\beta}_1$

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1})$$

- We also derived the form of the variance of the normal distribution, $\sigma^2_{\hat{\beta}_1}$ is

$$\sigma_{\hat{\beta}_1} = \sqrt{\frac{1}{n} \frac{Var[(X_i - \mu_X)u_i]}{[Var(X_i)]^2}} \tag{4.21}$$

The Standard Error of $\hat{\beta}_1$

- Because $Var(X) = EX^2 - (EX)^2$, then the *numerator* in the square root in (4.21) is

$$Var[(X_i - \mu_X)u_i] = E[(X_i - \mu_X)u_i]^2 - (E[(X_i - \mu_X)u_i])^2$$

- Based on the Law of Iterated Expectation(L.I.E), we have

$$E[(X_i - \mu_X)u_i] = E(E[(X_i - \mu_X)u_i]|X_i)$$

- Again by the 1st OLS assumption, thus $E(u_i|X_i) = 0$,

$$E[(X_i - \mu_X)u_i] = 0$$

- Then the second term in the equation above

$$Var[(X_i - \mu_X)u_i] = E[(X_i - \mu_X)u_i]^2$$

The Standard Error of $\hat{\beta}_1$

- **Because $plim(\overline{X}) = \mu_X$, then we use $\overline{X}$ and $\hat{\mu}_i$ to replace μ_X and μ_i in (4.21)(in large sample), then the *numerator* is**

$$Var[(X_i - \mu_X)u_i]$$

The Standard Error of $\hat{\beta}_1$

- Because $plim(\overline{X}) = \mu_X$, then we use $\overline{X}$ and $\hat{\mu}_i$ to replace μ_X and μ_i in (4.21)(in large sample), then the *numerator* is

$$Var[(X_i - \mu_X)u_i] = E[(X_i - \mu_X)u_i]^2$$

The Standard Error of $\hat{\beta}_1$

- Because $plim(\overline{X}) = \mu_X$, then we use $\overline{X}$ and $\hat{\mu}_i$ to replace μ_X and μ_i in (4.21)(in large sample), then the *numerator* is

$$\begin{aligned} Var[(X_i - \mu_X)u_i] &= E[(X_i - \mu_X)u_i]^2 \\ &= E[(X_i - \mu_X)^2 u_i^2] \end{aligned}$$

The Standard Error of $\hat{\beta}_1$

- Because $plim(\overline{X}) = \mu_X$, then we use $\overline{X}$ and $\hat{\mu}_i$ to replace μ_X and μ_i in (4.21)(in large sample), then the *numerator* is

$$
\begin{aligned}
Var[(X_i - \mu_X)u_i] &= E[(X_i - \mu_X)u_i]^2 \\
&= E[(X_i - \mu_X)^2 u_i^2] \\
&= plim\Big(\frac{1}{n-2}\sum_{i=1}^{n}(X_i - \overline{X})^2 \hat{u}_i^2\Big)
\end{aligned}
$$

where $n-2$ is the *freedom of degree*. Because when we calculate u_i, we have estimated two coefficiencts,β_0 and β_1.

The Standard Error of $\hat{\beta}_1$

- **Because** $plim(s_x) = \sigma_x^2 = Var(X_i)$**, then**

$$Var(X_i)$$

The Standard Error of $\hat{\beta}_1$

- **Because** $plim(s_x) = \sigma_x^2 = Var(X_i)$**, then**

$$Var(X_i) = plim(s_x)$$

The Standard Error of $\hat{\beta}_1$

- **Because** $plim(s_x) = \sigma_x^2 = Var(X_i)$**, then**

$$\begin{aligned} Var(X_i) &= plim(s_x) \\ &= plim(\frac{n-1}{n}(s_x)) \end{aligned}$$

The Standard Error of $\hat{\beta}_1$

- **Because $plim(s_x) = \sigma_x^2 = Var(X_i)$, then**

$$\begin{aligned} Var(X_i) &= plim(s_x) \\ &= plim(\frac{n-1}{n}(s_x)) \\ &= plim[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2] \end{aligned}$$

- **Then the *denominator* in the square root in (4.21) is**

$$[Var(X_i)]^2 = plim[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2]^2$$

The Standard Error of $\hat{\beta}_1$

- Then the **standard errors** of $\hat{\beta}_1$ is an **estimator** of the standard deviation of the sampling distribution $\sigma_{\hat{\beta}_1}$, thus

$$SE\left(\hat{\beta}_1\right) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}} = \sqrt{\frac{1}{n} \times \frac{\frac{1}{n-2}\sum(X_i - \bar{X})^2 \hat{u}_i^2}{\left[\frac{1}{n}\sum(X_i - \bar{X})^2\right]^2}} \tag{5.4}$$