Introductory Econometrics
Ch4 Multiple Regression Analysis: Inference

LIU Chenyuan

Spring 2024

Motivation

- In the previous chapters, we study the linear regression model, its interpretation, and how to construct the OLS estimator.
- We derived the expectation and variance of $\hat{\beta}_j$ under certain conditions.
- This chapter: statistical inference.
- We are interested in the following question: given our data, can we infer that β_j is of some particular value?

Motivation

- Example: we are interested in whether people with more years of education earn a higher wage.

  $$wage = \beta_0 + \beta_1 educ + u.$$

  We care about whether $\beta_1 = 0$.
- β_1 is unknown, so we collect data and calculate $\hat{\beta}_1 = 3.5$.
- Given this evidence, can we infer $\beta_1 \neq 0$? What if we estimate that $\hat{\beta}_1 = 0.35$, or $\hat{\beta}_1 = 0.0035$? Would our answer change if the sample size is 5, 500, or 5 million?
- We need a systematic way to guide the decision.

Learning Objectives

- In this chapter, we learn how to test hypotheses about the parameters in the population regression model.
- The building blocks:

1. Sampling distribution of the OLS estimators and the standardized estimators
2. Testing hypotheses about a single population parameter
3. Testing hypotheses about multiple population parameters

Outline

The Distribution of $\hat{\beta}_j$

- In the last lecture, we have derived the expectation and variance of $\hat{\beta}_j$ under certain conditions.
- These conditions are:
  1. MLR.1 Linear in Parameters
  2. MLR.2 Random Sampling
  3. MLR.3 No Perfect Collinearity
  4. MLR.4 Zero Conditional Mean: $E(u|x) = 0$
  5. MLR.5 Homoskedasticity: $Var(u|x) = \sigma^2$

The Distribution of $\hat{\beta}_j$

$$E[\hat{\beta}_j] = \beta_j,$$
$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}.$$

Suppose we are also interested in deriving the full distribution of $\hat{\beta}_j$. We need more assumptions.

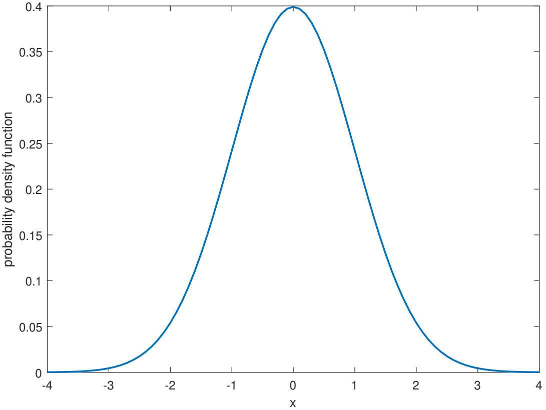
MLR.6 Normality

The population error u is independent of the explanatory variables $x_1, x_2, ..., x_k$ and is normally distributed with zero mean and variance σ^2: $u \sim Normal(0, \sigma^2)$.

- MLR.6 is much stronger than the previous assumptions.
- MLR.6 implies MLR.4 and MLR.5.

Recap: Normal Distribution

- $x \sim Normal(\mu, \sigma^2)$ or $x \sim N(\mu, \sigma^2)$: x follows a normal distribution with a mean of μ, and variance of σ^2.
- The PDF has a bell shape.
- Standard normal distribution: $Z \sim N(0, 1)$. The PDF:



Recap: Normal Distribution

- The PDF of the standard normal distribution: $\phi(z)$
- The CDF of the standard normal distribution:$\Phi(z)$. It represents the area left to z, below the PDF curve.
- We have the following properties for normal distributions:
  - $P(Z > z) = 1 - \Phi(z)$
  - $P(Z < -z) = P(Z > z)$
  - $P(a \leq Z \leq b) = \Phi(b) - \Phi(a)$
  - $P(|Z| > c) = P(Z < -c) + P(Z > c) = 2P(Z > c) = 2(1 - \Phi(c))$

Recap: Normal Distribution

- If $X \sim N(\mu, \sigma^2)$, then $aX + b \sim N(a\mu + b, a^2\sigma^2)$.
  - We can always transform a normally distributed random variable to a random variable following the standard normal distribution.
  - If $X \sim N(\mu, \sigma^2)$, then $\frac{X-\mu}{\sigma} \sim N(0, 1)$.
- If random variables $(X_i : i = 1, 2, ..., N)$ are independent, and follow the same normal distribution, then any linear combination of them also follows a normal distribution.

Classical Linear Model

1. MLR.1 Linear in Parameters
2. MLR.2 Random Sampling
3. MLR.3 No Perfect Collinearity
4. MLR.4 Zero Conditional Mean: $E(u|x) = 0$
5. MLR.5 Homoskedasticity: $Var(u|x) = \sigma^2$
6. MLR.6 Normality: $u \sim N(0, \sigma^2)$

- Assumptions MLR.1 through MLR.6 are called the classical linear model (CLM) assumptions.
- We refer to the model under these six assumptions as the **classical linear model**.

Classical Linear Model

- It can be shown that under the CLM assumptions (MLR.1 - MLR.6), the OLS estimators have the minimum variance among all the unbiased estimators.
- Compared with the Gauss-Markov Theorem: Under MLR.1 - MLR.5, the OLS estimators have the minimum variance among all the linear unbiased estimators.
- A succinct way to summarize the population assumptions of the CLM is

$$y|\mathbf{x} \sim Normal(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k, \sigma^2),$$

where $\mathbf{x}$ stands for $(x_1, x_2, ..., x_k)$.

- In reality, it is often hard to believe that the error term is normally distributed.

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u.$$

  Conditional on education, years of labor market experience, and years with the current employer, is wage normally distributed?
- Most likely, it is not true because the wage is always non-negative, so strictly speaking, the wage cannot be normally distributed.
- In Chapter 5, we will discuss that the nonnormality of the errors is not a serious problem with large sample sizes.

Normal Sampling Distributions

Normal Sampling Distributions

Under the CLM assumptions MLR.1 through MLR.6, conditional on the sample values of the independent variables,

$$\hat{\beta}_j \sim Normal(\beta_j, Var(\hat{\beta}_j)).$$

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \sim Normal(0, 1),$$

where $Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1-R_j^2)}$.

Proof

Recall that in a simple linear regression model,

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{N}(x_i - \bar{x})u_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2}.$$

When we conditional on x, then functions of x can be viewed as constants. As a result, $\hat{\beta}_1$ can be viewed as a linear combination of u_is, which are identically and independently normally distributed random variables, and thus has a normal distribution.

t Distribution for the Standardized Estimators

- In practice, σ^2 is not observed, so we need to estimate it using s^2:

$$se(\hat{\beta}_j) = \frac{\hat{\sigma}}{[SST_j(1 - R_j^2)]^{1/2}},$$

  where $\hat{\sigma}^2 = \frac{\sum_i \hat{u}_i^2}{N-k-1}$.
- We care about the distribution of $\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)}$.

t Distribution for the Standardized Estimators

t Distribution for the Standardized Estimators

Under the CLM assumptions MLR.1 through MLR.6,

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{N-k-1} = t_{df},$$

where $k+1$ is the number of unknown parameters in the population model $y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u$ and the $N-k-1$ is the degrees of freedom (df).

Normal Sampling Distribution vs t Distribution

- Compare the two results:

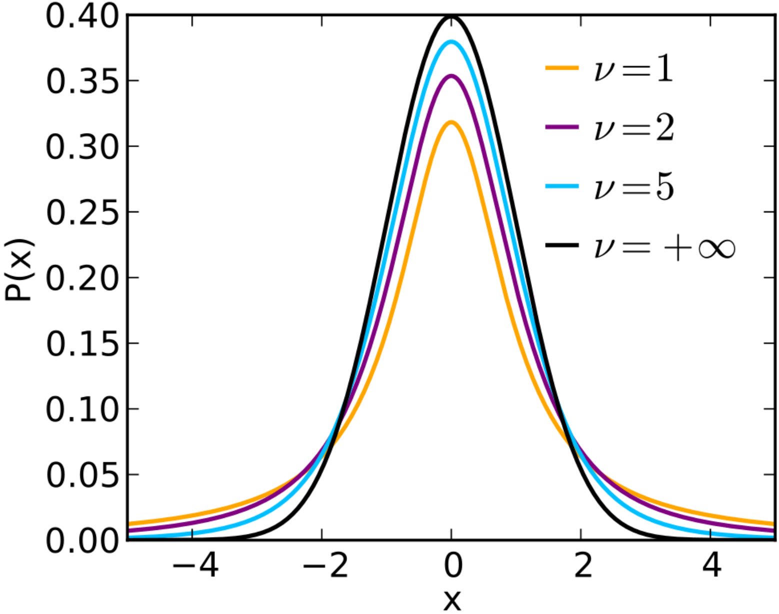
$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \sim Normal(0,1)$$

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{N-k-1} = t_{df}.$$

- The t distribution comes from the fact that the constant σ in $sd(\hat{\beta}_j)$ has been replaced with the random variable $\hat{\sigma}$.
- Because of this replacement, the resulting distribution follows a t distribution (proof not required.)

Recap: t distribution

- t distribution has one parameter: degree of freedom d. If a random variable X follows a t distribution with a degree of freedom of d, then we write $X \sim t_d$.
- The PDF of the t distribution is similar to the normal distribution but more dispersed, so there are more areas in the tail.
- When the degree of freedom gets larger, the t distribution becomes closer to the normal distribution.
- The degree of freedom of $\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)}$ is $N - k - 1$. When the sample size is large enough, we can approximate the distribution of $\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)}$ as normal.



0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
P(x)
−4
−2
0
2
4
x
$\nu = 1$
$\nu = 2$
$\nu = 5$
$\nu = +\infty$

Outline

Classical Linear Regression Model

The t Test

Confidence Interval

Testing Hypotheses about Multiple Parameters

Hypothesis Testing

- The reason why we care about the distribution of the OLS estimators is that we often want to test hypotheses about the parameters in the population regression model.
- Example: we are interested in whether obtaining an extra year of education improves wage, after accounting for labor market experience (*exper*) and years with the current employer (*tenure*).

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u.$$

  We care about whether $\beta_1 = 0$.
- Our hypothesis is $\beta_1 = 0$, and we want to test whether it is true.

Hypothesis Testing

- Let H_0 be the null hypothesis that we want to test. Let H_1 be the alternative hypothesis.
- Suppose that H_0 is true. We then ask whether the world seems consistent with it.
- Specifically: we perform some experiment and see if the results of the experiment are consistent with the hypothesis.

Hypothesis Testing

- In statistics, we often never know the answer for sure.
- Consider a random variable X. The hypothesis is that $E(X) = 0$.
- Suppose we collect a sample of X, and calculate the average $\bar{X} = \frac{1}{N}\sum_i X_i$. Even if $\bar{X} = 10000$, it is still possible that $E(X) = 0$ is true, though very unlikely.
- Hypothesis testing helps formalize the idea of "unlikeliness".

Test Statistic

What is a test statistic?

1. Depend on the data
2. We know its distribution under the null hypothesis

How to use the test statistic to conduct hypothesis testing?

- We reject the null hypothesis when the test statistic falls in the **rejection region**.

Rejection Region

How to decide the rejection region?

- We care about the type I error:

$$\text{significance level } \equiv \alpha = \Pr(\text{Rejecting } H_0 | H_0 \text{ is true}).$$

- We care about the type II error:

$$\Pr(\text{Not rejecting } H_0 | H_1 \text{ is true}).$$

When deciding the rejection region, classical hypothesis testing requires that

- we initially specify a significance level (quantifying the tolerance for Type I error)
- then minimize the probability of a Type II error

Testing Against One-Sided Alternatives

Suppose we are interested in testing

$$H_0: \quad \beta_j = 0.$$
$$H_1: \quad \beta_j > 0.$$

Consider a test statistic:

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)}$$

When H_0 is true, the test statistic is:

$$\frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \sim t_{N-k-1} = t_{df}.$$

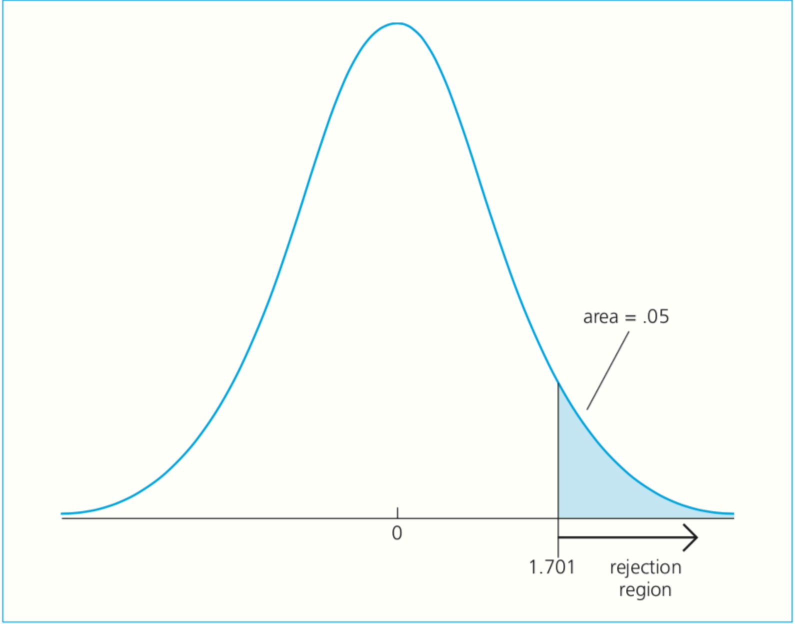
1. It depends on the data.
2. We know its distribution under H_0.

$$t_{\hat{\beta}_j} \equiv \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

- We often call $t_{\hat{\beta}_j}$ t-statistic or t-ratio of $\hat{\beta}_j$.
- $t_{\hat{\beta}_j}$ has the same sign as $\hat{\beta}_j$, because $se(\hat{\beta}_j) > 0$.
- Intuitively, we reject H_0 when $t_{\hat{\beta}_j}$ is large enough: the larger the $t_{\hat{\beta}_j}$, the less likely that H_0 is true, the more likely that H_1 is true.
- How large is "large enough"?
- Fix a **significance level** of 5%. The **critical value**, c is the 95th percentile when H_0 is true. It means when H_0 is true, the probability of getting a value as large as c is 5%.
- Rejection rule:

$$t_{\hat{\beta}_j} > c.$$

FIGURE 4.2 5% rejection rule for the alternative H_1: $\beta_j > 0$ with 28 *df*.



- We know that under H_0, the probability of getting a value as extreme as c is 5%.
- Rejecting H_0 when $t_{\hat{\beta}_j} > c$ means the probability of making a type I error, that is, the probability of rejecting H_0 when H_0 is true, is 5%.
- The idea of test
  1. Fix a significance level α. That is, decide our level of "tolerence" for the type I error.
  2. Find the critical value associated with α. For $H_1 : \beta_j > 0$, this means finding the $(1 - \alpha)$-th percentile of the t distribution with $df = N - k - 1$.
  3. Reject H_0 if
  $$t_{\hat{\beta}_j} > c.$$

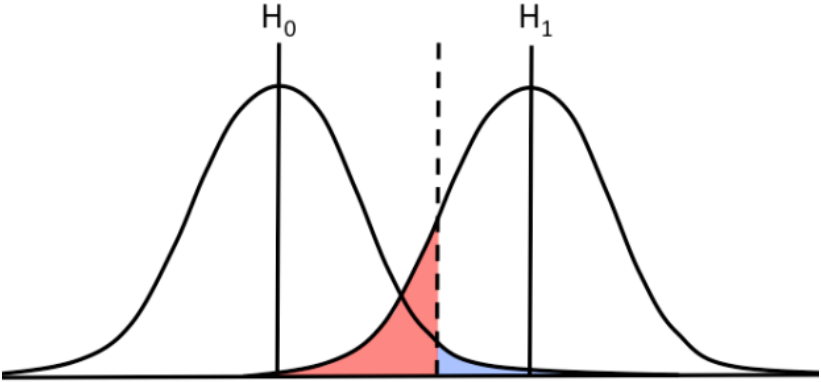
- The critical value c depends on the significance level α, and the degree of freedom $N-k-1$.
- When N is large enough, we can use normal distribution to approximate t-distribution.
- Use `STATA` to find the critical value:

```
display invt(N-k-1, 1-α).
```

Trade-off between Type I and Type II Error

	H_0 is true	H_1 is true
Reject H_0	Type I error	Correct decision
Not reject H_0	Correct decision	Type II error

- We want our rejection rule to minimize both the type I error and the type II error.
- However, it is often the case that when we reduce the type I error, we increase the type II error.



H_0
H_1

Example: Return on Education

We want to test $H_0 : \beta_{educ} = 0$ against $H_1 : \beta_{educ} > 0$.

$$\widehat{\log(wage)} = \underset{(.104)}{.284} + \underset{(.007)}{.092}\, educ + \underset{(.0017)}{.0041}\, exper + \underset{(.003)}{.022}\, tenure$$

$$n = 526, R^2 = .316,$$

- We usually put standard errors in parentheses.
- The t-stat for $\beta_{educ} = \frac{0.092}{0.007} \approx 13.143$.
- The degree of freedom is 526-3-1 $= 522$. The 5% critical value is 1.645, and the 1% critical value is 2.326.
- Rule of thumb: when degree of freedom is larger than 120, we can use the standard normal critical values.
- We conclude that β_{educ} is statistically greater than zero at the 1% significance level.

Similarly, if we want to test

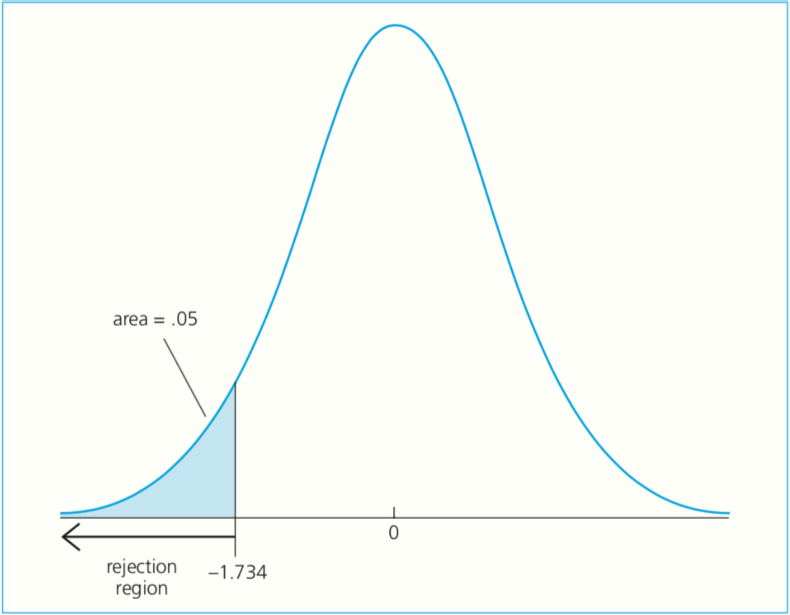
$$
\begin{aligned}
H_0 : &\quad \beta_j = 0. \\
H_1 : &\quad \beta_j < 0.
\end{aligned}
$$

Then we could:

1. Fix a significance level α. That is, decide our level of "tolerence" for the type I error.
2. Find the critical value associated with α. For $H_1 : \beta_j < 0$, this means finding the α-th percentile of the t distribution with $df = N - k - 1$.
3. Reject H_0 if

$$t_{\hat{\beta}_j} < c.$$

FIGURE 4.3 5% rejection rule for the alternative H_1: $\beta_j < 0$ with 18 df.



Two-sided Alternatives

We want to test:

$$H_0: \quad \beta_j = 0.$$
$$H_1: \quad \beta_j \neq 0.$$

- This is the relevant alternative when the sign of β_j is not well determined by theory (or common sense).
- Even when we know whether β_j is positive or negative under the alternative, a two-sided test is often prudent.

Two-sided Alternatives

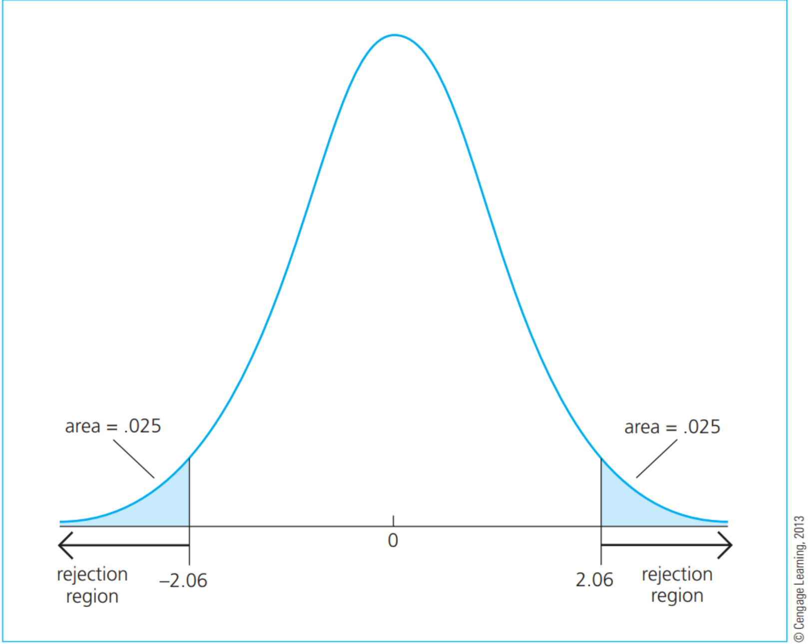
$$
\begin{aligned}
H_0: &\quad \beta_j = 0. \\
H_1: &\quad \beta_j \neq 0.
\end{aligned}
$$

Then we could:

1. Fix a significance level α. That is, decide our level of "tolerence" for the type I error.
2. Find the critical value associated with α. For $H_1 : \beta_j \neq 0$, this means finding the (1-$\alpha/2$)-th percentile of the t distribution with $df = N - k - 1$.
3. Reject H_0 if

$$|t_{\hat{\beta}_j}| > c.$$

FIGURE 4.4 5% rejection rule for the alternative H_1: $\beta_j \neq 0$ with 25 *df*.



- When a specific alternative is not stated, it is usually considered to be two-sided.
- If H_0 is rejected in favor of $H_1 : \beta_j \neq 0$ at the 5% level, we usually say that "x_j is statistically significant, or statistically different from zero, at the 5% level."
- If H_0 is not rejected, we say that "x_j is statistically insignificant at the 5% level."

Example: School Size and Test Scores

We regress students' math score (*math10*) on school size (*enroll*), controlled for average annual teacher compensation (*totcomp*) and the number of staff per one thousand students (*staff*).

$$\underset{(6.113)}{\widehat{math10}} = 2.274 + \underset{(.00010)}{.00046}\, totcomp + \underset{(.040)}{.048}\, staff - \underset{(.00022)}{.00020}\, enroll$$

$$n = 408,\ R^2 = .0541.$$

- The t statistic on enroll is $-0.00020/.00022 \approx -0.91$.
- Since $df = 408 - 3 - 1 = 404$, we use the standard normal critical value. At the 5% level, the critical value is 1.96.
- We conclude that *enroll* is not statistically significant at the 5% level.

Testing Other Hypothesis about β_j

If the null is stated as:

$$H_0 : \beta_j = a.$$

Then the t-statistic is

$$\frac{\hat{\beta}_j - a}{se(\hat{\beta}_j)} \sim t_{N-k-1}.$$

We can use the general t statistic to test against one-sided or two-sided alternatives.

Computing p-Values for t Tests

- Given the observed value of the t statistic, what is the smallest significance level at which the null hypothesis would be rejected?
- We call this "smallest significance level" **p-value**.
- p-value represents the probability of observing a value as extreme as $t_{\hat{\beta}_j}$ under the H_0.

The p-value for two-sided alternatives

$$H_0: \quad \beta_j = 0.$$
$$H_1: \quad \beta_j \neq 0.$$

The p-value, in this case, is

$$P(|T| > |t|) = 2P(T > |t|),$$

where we let T denote a t-distributed random variable with $N - k - 1$ degrees of freedom and let t denote the numerical value of the test statistic. $P(T > |t|)$ is the probability that random variable T is larger than the value $|t|$.

The p-value

- The p-value nicely summarizes the strength or weakness of the empirical evidence against the null hypothesis.
- The p-value is the probability of observing a t statistic as extreme as we did if the null hypothesis is true.
- Small p-values are evidence against H_0; large p-values provide little evidence against H_0.

t statistics vs p-value

- Significance level and critical value have a one-one-mapping relationship
- We can either compare $t_{\hat{\beta}_j}$ with c, or equivalently, compare the significance level and the p-value.

1. Fix a significance level α, calculate the critical value c, and then reject H_0 if $|t_{\hat{\beta}_j}| > c$.
2. Fix a significance level α, calculate the p-value, reject H_0 if $p < \alpha$.

Economic versus Statistical Significance

- The statistical significance of a variable x_j is determined entirely by the size of $t_{\hat{\beta}_j}$, whereas the economic significance or practical significance of a variable is related to the size (and sign) of $\hat{\beta}_j$.
- We often care about both statistical significance and economic significance.

$$
\begin{aligned}
\widehat{prate} &= \underset{(0.78)}{80.29} + \underset{(0.52)}{5.44}\, mrate + \underset{(.045)}{.269}\, age - \underset{(.00004)}{.00013}\, totemp \\
n &= 1{,}534,\ R^2 = .100.
\end{aligned}
$$

- We are interested in whether the number of employees ($totemp$) in a firm has an impact on the participation rate in retirement savings plan ($prate$), controlled for firm match rate $mrate$ and age of a plan age.
- The t-stat for β_{totemp} is -3.25, which is statistically significant at 1% level.
- Holding mrate and age fixed, if a firm grows by 10,000 employees, the participation rate falls by $10,000(.00013) = 1.3$ percentage points.

Guidelines

1. Check for statistical significance. If the variable is statistically significant, discuss the magnitude and its economic importance.
2. If a variable is not statistically significant at the usual levels (10%, 5%, or 1%), you might still ask if the variable has the expected effect on y and whether that effect is practically large.
3. Variables with small t stat that have the "wrong" sign - we conclude that the variables are statistically insignificant.
4. A significant variable that has an unexpected sign and a practically large effect usually indicates a problem with the model and the nature of the data: Chapters 9 and 15.

Outline

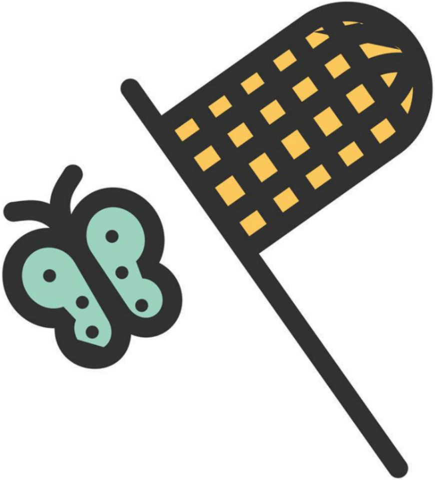
Confidence Interval

- Sometimes, besides calculating a point estimate, we can also calculate an interval estimate, which provides a range of likely values for the population parameter.
- We can construct a confidence level depending on α. We call it a $(1-\alpha)$ confidence interval:
$$[\hat{\beta}_j - c \cdot se(\hat{\beta}_j), \hat{\beta}_j + c \cdot se(\hat{\beta}_j)]$$
- The critical value c is the $(1-\alpha/2)$ percentile in a t distribution with $df = N-k-1$.

Interpretation of the Confidence Interval

- The upper and lower bounds of a confidence interval are random variables.
- The true parameter is fixed and unknown.
- For every sample, we can calculate a confidence interval.
- The meaning of a 95% confidence interval: if we sample repeatedly many times, then the true β_j will appear in 95% of the confidence intervals.
- Unfortunately, for the single sample that we use to construct the CI, we do not know whether β_j is actually contained in the interval.

We want to catch the butterfly (β_j) with our net (confidence interval). The butterfly is fixed while we move our net.



Conduct Two-Tailed Hypotheses Tests Using Confidence Interval

$$
\begin{aligned}
H_0 : \quad & \beta_j = 0. \\
H_1 : \quad & \beta_j \neq 0.
\end{aligned}
$$

- Fix a significance level. Calculate the critical value and construct the confidence interval.
- Reject H_0 if 0 is not inside the confidence interval.

$$
\begin{aligned}
H_0: &\quad \beta_j = 0. \\
H_1: &\quad \beta_j \neq 0.
\end{aligned}
$$

Three equivalent rejection rules:

1. $|t_{\hat{\beta}_j}| > c$
2. $p < \alpha$
3. 0 is not inside the confidence interval.

Stata Example

```
. reg lwage educ exper tenure married sibs

      Source |       SS           df       MS      Number of obs   =       935
-------------+----------------------------------   F(5, 929)       =     40.85
       Model |  29.8570149         5  5.97140298   Prob > F        =    0.0000
    Residual |  135.799279       929  .146177911   R-squared       =    0.1802
-------------+----------------------------------   Adj R-squared   =    0.1758
       Total |  165.656294       934  .177362199   Root MSE        =    .38233

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0720685   .0066042    10.91   0.000     .0591076    .0850293
       exper |   .0138376   .0033345     4.15   0.000     .0072936    .0203816
      tenure |   .0125549    .002556     4.91   0.000     .0075386    .0175712
     married |   .1980791   .0407447     4.86   0.000     .1181168    .2780413
        sibs |  -.0119699   .0055986    -2.14   0.033    -.0229574   -.0009825
       _cons |   5.415827   .1209113    44.79   0.000     5.178536    5.653118
------------------------------------------------------------------------------
```

Outline

Classical Linear Regression Model

The t Test

Confidence Interval

Testing Hypotheses about Multiple Parameters

Testing Hypotheses about a Single Linear Combination of the Parameters

- Consider the model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u.$$

- We want to test

$$\begin{aligned} H_0: &\quad \beta_1 = \beta_2. \\ H_1: &\quad \beta_1 \neq \beta_2. \end{aligned}$$

- How to do that?

- Method 1: construct the t-stat for $\hat{\beta}_1 - \hat{\beta}_2$

$$t_{\hat{\beta}_1 - \hat{\beta}_2} = \frac{\hat{\beta}_1 - \hat{\beta}_2}{se(\hat{\beta}_1 - \hat{\beta}_2)}$$

- The challenge: calculate $se(\hat{\beta}_1 - \hat{\beta}_2)$.

$$se(\hat{\beta}_1 - \hat{\beta}_2) = \{[se(\hat{\beta}_1)]^2 + [se(\hat{\beta}_2)]^2 - 2s_{12}\}^{1/2}.$$

  where s_{12} is the estimate of $Cov(\hat{\beta}_1, \hat{\beta}_2)$.

Method 2: estimating a new regression model:

- Define $\theta = \beta_1 - \beta_2$. So $\beta_1 = \beta_2 + \theta$.

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \\ &= \beta_0 + (\beta_2 + \theta) x_1 + \beta_2 x_2 + u \\ &= \beta_0 + \theta x_1 + \beta_2 (x_1 + x_2) + u. \end{aligned}$$

- We can construct a new variable, $z = x_1 + x_2$, and then regress y on x_1 and z.
- Test in the new regression whether the coefficient for x_1 is 0.

Testing Multiple Linear Restrictions: The F Test

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

We want to test

$$H_0: \quad \beta_1 = 0 \text{ and } \beta_2 = 0.$$
$$H_1: H_0 \text{ is not true.}$$

- The null hypothesis constitutes two **exclusion restrictions**: if H_0 is true, then x_1 and x_2 have no partial effect on y.
- A test of multiple restrictions is called a **multiple hypotheses test** or a **joint hypotheses test**.

- The **unrestricted model** is

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

- Consider the **restricted model** when H_0 is true

$$y = \beta_0 + \beta_3 x_3 + u.$$

- If H_0 is true, the two models are the same. That means when we include x_1 and x_2 into the model, the sum of squared residuals (**SSR**) should not change much.
- However, if H_0 is false, that means that at least one of β_1, β_2 is nonzero, then the SSR should fall when we include these new variables.
- We should reject H_0 if the two SSR are very different.

F Test

- Write the unrestricted model with k independent variables as:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u.$$

- The null hypothesis

$$H_0 : \beta_{k-q+1} = 0, ..., \beta_k = 0.$$

- The restricted model is:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_{k-q} x_{k-q} + u.$$

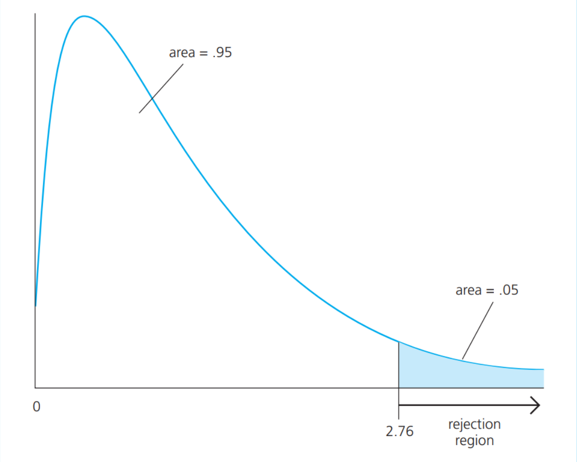
- The test statistic, the **F statistic** is

$$F \equiv \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(N-k-1)}.$$

$$F \equiv \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(N-k-1)}.$$

- SSR_r is the sum of squared residuals from the restricted model, and SSR_{ur} is the SSR from the unrestricted model.
- The F statistic is always non-negative: when we include more explanatory variables, SSR would decrease.
- q is the number of restrictions
- We can show that the sampling distribution of the F-stat: $F \sim F_{q,N-k-1}$. We call this an F distribution with q degrees of freedom in the numerator and $N-k-1$ degrees of freedom in the denominator.

FIGURE 4.7 The 5% critical value and rejection region in an $F_{3,60}$ distribution.



- Fix the significance level α, and then calculate the critical value c. Reject H_0 if $F > c$.

Relationship between F and t Stat

- We can also apply the F statistic to the case of testing the significance of a single independent variable
- In fact, the t-statistic squared has an F distribution with 1 degree of freedom in the numerator.

- The R^2 version of the F statistic

$$F \equiv \frac{(R^2_{ur} - R^2_r)/q}{(1 - R^2_{ur})/(N - k - 1)}.$$

- In the F testing context, the p-value is defined as

$$P(\mathcal{F} > F).$$

  where $\mathcal{F}$ denotes an F random variable with $(q, N - k - 1)$ degrees of freedom, and F is the actual value of the test statistic.

- p-value is the probability of observing a value of F at least as large as we did, given that the null hypothesis is true.
- Reject H_0 if $p < \alpha$.

Example

Suppose we have a regression model:

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

We want to test

$$H_0 : \beta_1 = 1, \beta_2 = 0.$$
$$H_1 : H_0 \text{ is not true.}$$

We can use the same idea of the F test. Suppose H_0 is true, then the restricted model is:

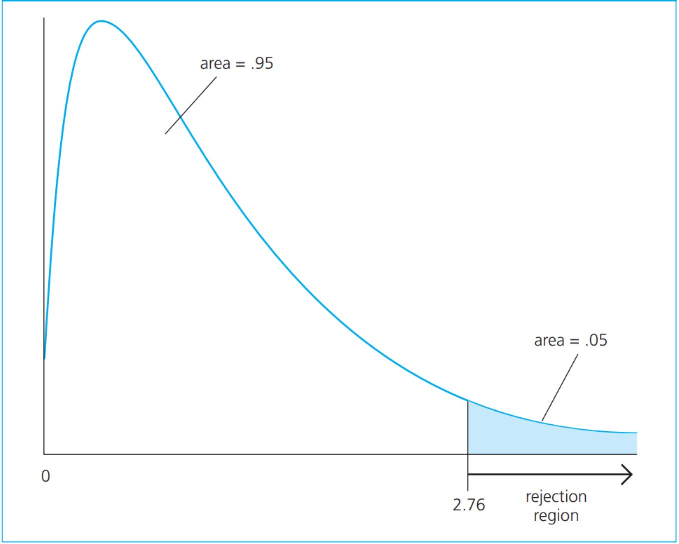
$$y - x_1 = \beta_0 + \beta_3 x_3 + u.$$

We are testing $q = 2$ restrictions, and there are $N - 4$ df in the unrestricted model. The F statistic is simply $[(SSR_r - SSR_{ur})/SSR_{ur}][(N - 4)/2]$

- In general, use the F test to test multiple linear hypotheses.
- Assume H_0 is true, write out the restricted model. We then estimate both the unrestricted model and the restricted model to get SSR and then calculate the F statistic.
- We cannot use the R-squared form of the F statistic for this example because the dependent variables of the two models are different.
- More generally, the SSR form of the F statistic should be used if a different dependent variable is needed in running the restricted regression.

In `Stata`, use `test` to calculate F stat and its p-value.

FIGURE 4.7 The 5% critical value and rejection region in an $F_{3,60}$ distribution.



```
reg y x1 x2 x3

      Source |       SS           df       MS      Number of obs   =     2,539
-------------+----------------------------------   F(3, 2535)      =     44.33
       Model |  2.0942e+11         3  6.9807e+10   Prob > F        =    0.0000
    Residual |  3.9923e+12     2,535  1.5749e+09   R-squared       =    0.0498
-------------+----------------------------------   Adj R-squared   =    0.0487
       Total |  4.2017e+12     2,538  1.6555e+09   Root MSE        =     39685

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   2054.521   274.6959     7.48   0.000     1515.869    2593.172
          x2 |   2079.292   1748.835     1.19   0.235    -1349.999    5508.583
          x3 |   13280.51   1823.447     7.28   0.000     9704.914    16856.11
       _cons |    17969.6   3530.166     5.09   0.000      11047.3    24891.91
------------------------------------------------------------------------------

test x2 = 0

( 1)  x2 = 0

      F(  1,  2535) =    1.41
           Prob > F =    0.2346
```

```
reg y x1 x2 x3

      Source |       SS           df       MS      Number of obs   =     2,539
-------------+----------------------------------   F(3, 2535)      =     44.33
       Model |  2.0942e+11         3  6.9807e+10   Prob > F        =    0.0000
    Residual |  3.9923e+12     2,535  1.5749e+09   R-squared       =    0.0498
-------------+----------------------------------   Adj R-squared   =    0.0487
       Total |  4.2017e+12     2,538  1.6555e+09   Root MSE        =     39685

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   2054.521   274.6959     7.48   0.000     1515.869    2593.172
          x2 |   2079.292   1748.835     1.19   0.235    -1349.999    5508.583
          x3 |   13280.51   1823.447     7.28   0.000     9704.914    16856.11
       _cons |    17969.6   3530.166     5.09   0.000      11047.3    24891.91
------------------------------------------------------------------------------
```

```
test (x1 = 0) (x2 = 0) (x3 = 0)

 ( 1)  x1 = 0
 ( 2)  x2 = 0
 ( 3)  x3 = 0

       F(  3,  2535) =   44.33
            Prob > F =    0.0000
```

Summary

- Classical Linear Regression Model
- Sampling distribution of $t_{\hat{\beta}_j}$.
- Test a single linear restriction: $H_0 : \beta = 0$
  - t test
  - p-value
  - confidence interval
- Test multiple linear restrictions:
  - F test