Lecture 9 Multiple Regression Analysis - Inference

Recap

Previously, we have shown that the OLS estimator is unbiased

$$\mathbb{E}[\hat{\beta} \mid \mathbf{X}] = \beta.$$

We also found the variance of the OLS estimator under *Gauss-Markov* assumptions

$$\text{Var}(\hat{\beta}_j \mid \mathbf{X}) = \frac{\sigma^2}{SST_j(1-R_j^2)},$$

as well as the full covariance matrix

$$\text{Var}(\hat{\beta} \mid \mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

We saw that the *Gauss-Markov theorem* guarantees that OLS has the smallest variance among all linear unbiased estimators. We call OLS a *linear estimator* because the coefficient estimates can be written as a linear combination of outcome values

$$\hat{\beta}_j = \sum_{i=1}^{n} w_{ij} y_i, \quad w_{ij} \equiv \frac{\hat{u}_{ij}}{\sum_{i=1}^{n} \hat{u}_{ij}^2}.$$

In this section, we will talk about how these properties can help us make *statistical inference* and test hypotheses. However, first we need to make one more assumption.

Sampling distribution of the OLS

Even though we know the expected value and variance of OLS, it is not enough to make statistical inference. We need to know the full *sampling distribution* of OLS. To get that distribution, we have to add one more assumption to our set of *Gauss-Markov assumptions*. We have to assume that the error term in the population model has a *normal distribution*.

> ⚠ **Assumption: Normality**
>
> The population error term U is independent of predictors and is normally distributed with zero mean and variance σ^2.
>
> $$U \sim \mathcal{N}(0, \sigma^2)$$

Notice that this assumption implies two of our previous assumptions: that the error term is mean-independent of predictors and homoskedasticity. The normality assumption together with the Gauss-Markov assumption forms a set of *classical linear model* (CLM) assumptions.

1. Linear CEF
    1. Linear model

    2. Error term is mean-independent of predictors
2. Random Sampling
3. No Perfect Collinearity
4. Homoskedasticity
5. Normality

A model that satisfies these assumptions is called a *classical linear model*.

If the error term is normally distributed, then the conditional distribution of the outcome Y given the predictors $\mathbf{x} = (X_1, X_2, \ldots, X_k)$ is also normal:

$$Y \mid \mathbf{x} \sim \mathcal{N}(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k, \sigma^2).$$

The normality assumption allows us to derive the sampling distribution of the OLS estimator

Theorem: Normal sampling distribution of OLS

Under the CLM assumptions the distribution of the OLS estimator (conditional on the values of predictors) is normal.

$$\hat{\beta}_j \sim \mathcal{N}(\beta_j, \mathrm{Var}(\hat{\beta}_j)), \quad j = 1, \ldots, k$$

Proof

From the regression anatomy formula, we know that OLS is a linear combination of outcomes:

$$\hat{\beta}_j = \sum_{i=1}^{n} w_{ij} y_i.$$

A linear combination of normally distributed random variables is also normally distributed. Since the outcome is normally distributed, each $\hat{\beta}_j$ is also normally distributed.

Testing hypotheses about a single coefficient

Let's begin with our population model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + U.$$

The assumptions we made allows us to test hypotheses about the coefficients of the model. We begin with testing hypotheses about a single coefficient, β_j. One of the most common hypotheses we would like to test is whether the coefficient is zero or not:

$$\mathrm{H}_0 : \beta_j = 0.$$

This hypothesis means that the predictor X_j does not have an effect on the conditional mean of the outcome, after we controlled for other predictors. If our statistical test rejects this null hypothesis, we say that β_j is *statistically significant*. If we fail to reject the null, we say that β_j is *not statistically significant*.

Note

The hypotheses we formulate are about population parameters. We do not test hypotheses about the estimates of population parameters. E.g., writing

$$H_0 : \hat{\beta}_j = 0$$

does not make sense.

To test this hypothesis, we first need to form a *test statistic*. In our case, this will be the so-called t *statistic*

$$t \equiv \frac{\hat{\beta}_j}{\mathrm{se}(\hat{\beta}_j)},$$

where $\mathrm{se}(\hat{\beta}_j)$ is the standard error of $\hat{\beta}_j$ (the estimate of the standard deviation of $\hat{\beta}_j$).

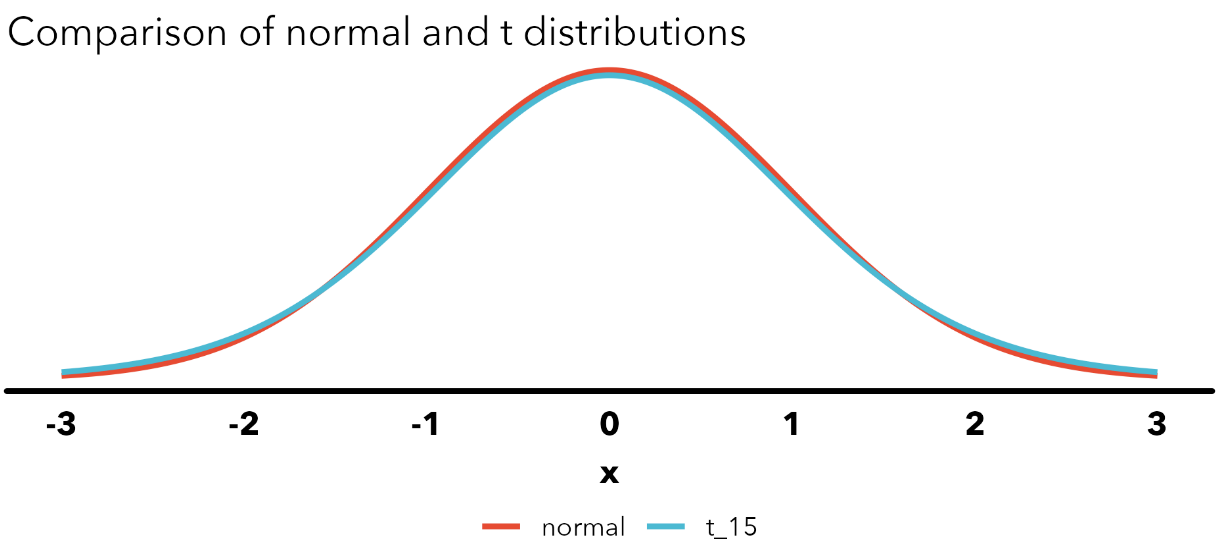
Second, we need to know the *distribution* of the test statistic under the null hypothesis (H_0), i.e., when the null hypothesis is true. The following result characterizes this distribution.

Theorem: t distribution for the standardized estimators

Under the CLM assumptions, the standardized OLS estimator follows a Student's t distribution with $n - k - 1$ degrees of freedom (df):

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

The graph below compares the standard normal distribution with a t distribution with 15 *df*.



Third, we need to form an *alternative* hypothesis. One option would be to formulate a one-sided alternative, e.g.,

$$\mathrm{H}_1 : \beta_j > 0.$$

Fourth, we need to decide on the *significance level* α of our test. The significance level is the probability of rejecting H_0 when it is true. A common choice for the significance level is $\alpha = 5\%$.

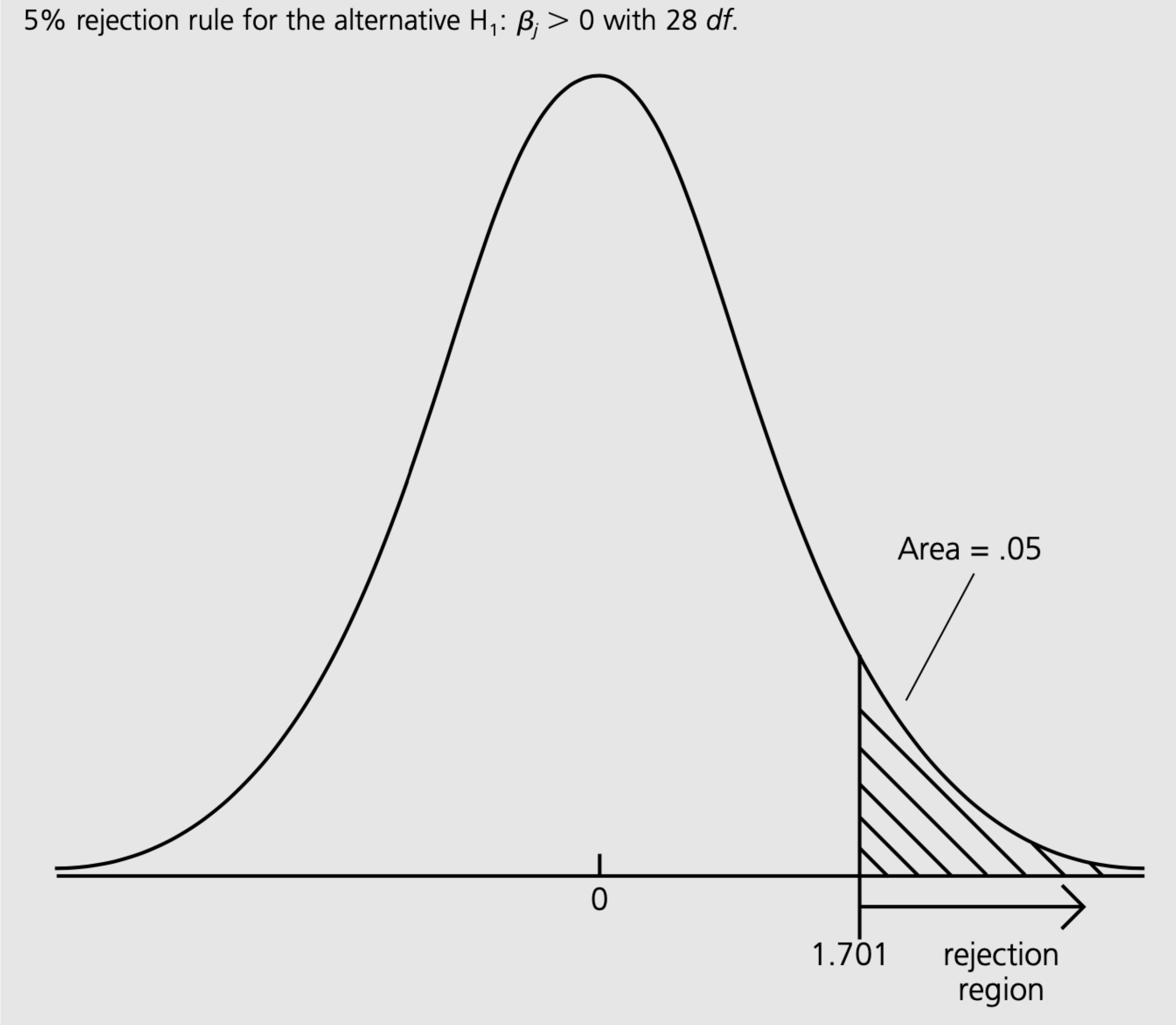
The significance level determines our decision rule for whether we reject or do not reject the null hypothesis. A general decision rule for a one-sided test is

$$\begin{cases} t > c, & \text{reject } \mathrm{H}_0 \\ t \leqslant c, & \text{do not reject } \mathrm{H}_0. \end{cases}$$

In other words, to reject the null hypothesis, the value of the t statistic has to be *high enough*. The *critical value* c is determined by our significance level α and the degrees of freedom $n - k - 1$. It is the $1 - \alpha$ percentile of the distribution of the test statistic.

For example, for a 5% significance level and 28 df, the critical value is the 95th percentile of the t distribution with 28 df (= 1.701). We will reject the null hypothesis wherever the t

statistic is above 1.701 and we will not reject the null wherever the t statistic is below 1.701.



> **Note**
>
> In statistical hypotheses testing, we either reject the null hypothesis or we do not reject the null hypothesis. We do not say that we **accept** the null hypothesis because there are infinitely many null hypotheses that we might not be able to reject.

Similarly, if our one-sided alternative is

$$\mathrm{H}_1 : \beta_j < 0,$$

our decision rule will look like

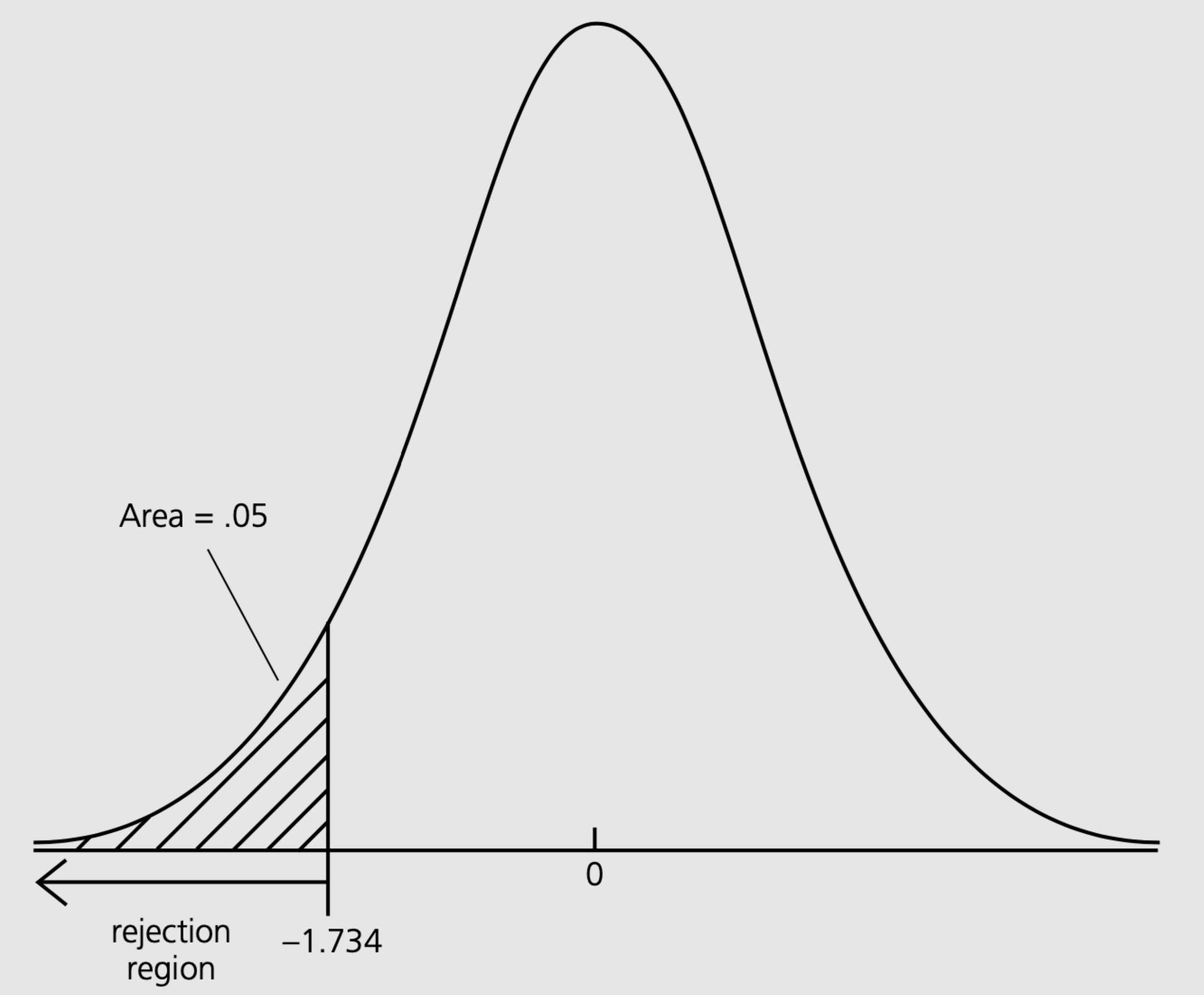
$$\begin{cases} t < -c, \text{ reject } \mathrm{H}_0 \\ t \geqslant -c, \text{ do not reject } \mathrm{H}_0, \end{cases}$$

where c (positive number) is defined as above.

For example, for a 5% significance level and 18 df, the critical value c is the 95th percentile of the t distribution with 18 df (= 1.734). We will reject the null hypothesis wherever the t

statistic is below 1.734 and we will not reject the null wherever the t statistic is above 1.734.

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



Often we do not have any expectation about the sign of a coefficient. In these cases, we use a two-sided alternative:

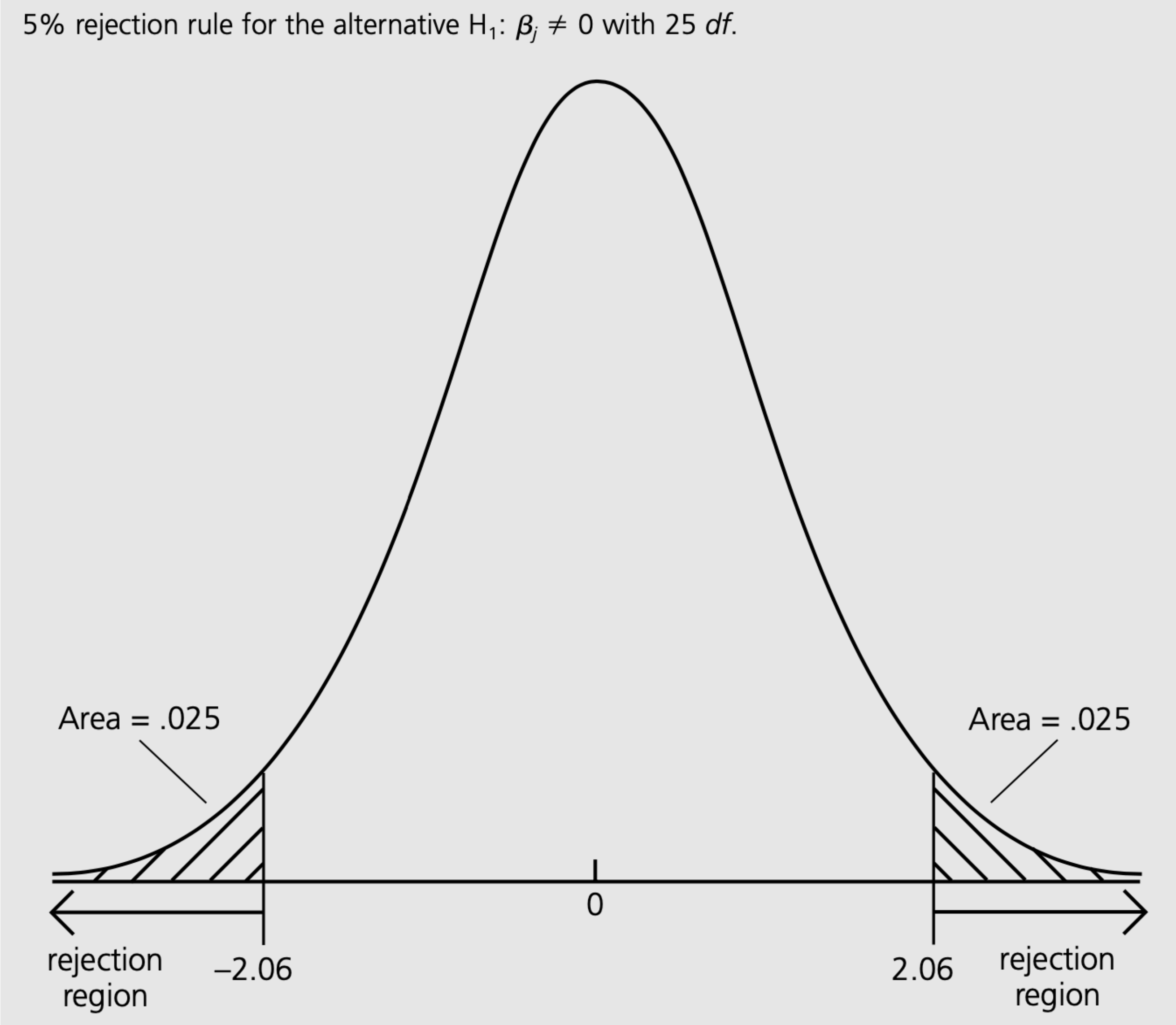
$$\mathrm{H}_1 : \beta_j \neq 0.$$

To form a decision rule, we now use the *absolute value* of the test statistic

$$\begin{cases} |t| > c, \text{ reject } \mathrm{H}_0 \\ |t| \leqslant c, \text{ do not reject } \mathrm{H}_0. \end{cases}$$

The critical value corresponding to the α% significance level will now be the $\frac{1-\alpha}{2}$ percentile of the distribution of the test statistic.

For example, for a 5% significance level and 18 df, the critical value c is the 97.5th percentile of the t distribution with 25 df (= 2.06). We will reject the null hypothesis wherever the absolute value of the t statistic is above 2.06 and we will not reject the null wherever the absolute value of the t statistic is below 2.06.



Hint

You can use a quick rule of thumb to determine whether a coefficient is statistically significant at a 5% level. If the absolute value of the estimate of the coefficient is at least twice as high as the standard error, the coefficient is statistically significant. This would correspond to a case when the absolute value of the t statistic is at least 2. Number 2 is an approximation of 1.96, which is the 97.5th percentile of the standard normal distribution.

In some cases we would like to test whether a coefficient is statistically different from a number, say b, other than zero. For example, if our coefficient has the interpretation of elasticity, we might want to test whether the elasticity is one. In these case our null hypothesis takes the form

$$\mathrm{H}_0 : \beta_j = b.$$

The test statistic corresponding to this null hypothesis will now be

$$t = \frac{\hat{\beta}_j - b}{\mathrm{se}(\hat{\beta}_j)}.$$

Notice that when $b = 0$ we are back to our previous test statistic.

> **Statistical vs. economic significance**
>
> One should not confuse statistical significance with economic significance. For example, in large samples even very small effects can be statistically significant. However, that would not make them economically significant

p-values

The procedure for testing a hypothesis we described so far follows a *classical* approach:

1. State the null and alternative hypotheses
2. Compute the test statistic
3. Choose a significance level
4. Compute the critical value based on the significance level and degrees of freedom
5. Determine the rejection region based on the critical value and the alternative hypothesis
6. Compare the value of the test statistic with the rejection region

An alternative approach would be to compute the p-value of a test. The p-value is the probability of observing a test statistic as extreme as we did if the null hypothesis is true. It also tells us what is the smallest significance level at which the null hypothesis will be rejected. And it also tells us the probability of making a Type I error (rejecting the null hypothesis when it is true) by rejecting the null.

Let T be a random variable following a standard t distribution with $n - k - 1$ degrees of freedom and $F_t(\cdot)$ be its cumulative distribution function. We compute the p-value of a test for a given value of the test statistic t and a two-sided alternative as

$$p \equiv \mathbb{P}(|T| > |t|) = \mathbb{P}(T > |t|) + \mathbb{P}(T < -|t|) = 2\mathbb{P}(T > |t|) = 2(1 - F_t(|t|)).$$

Note that we multiple by two because of the symmetry of the t distribution.

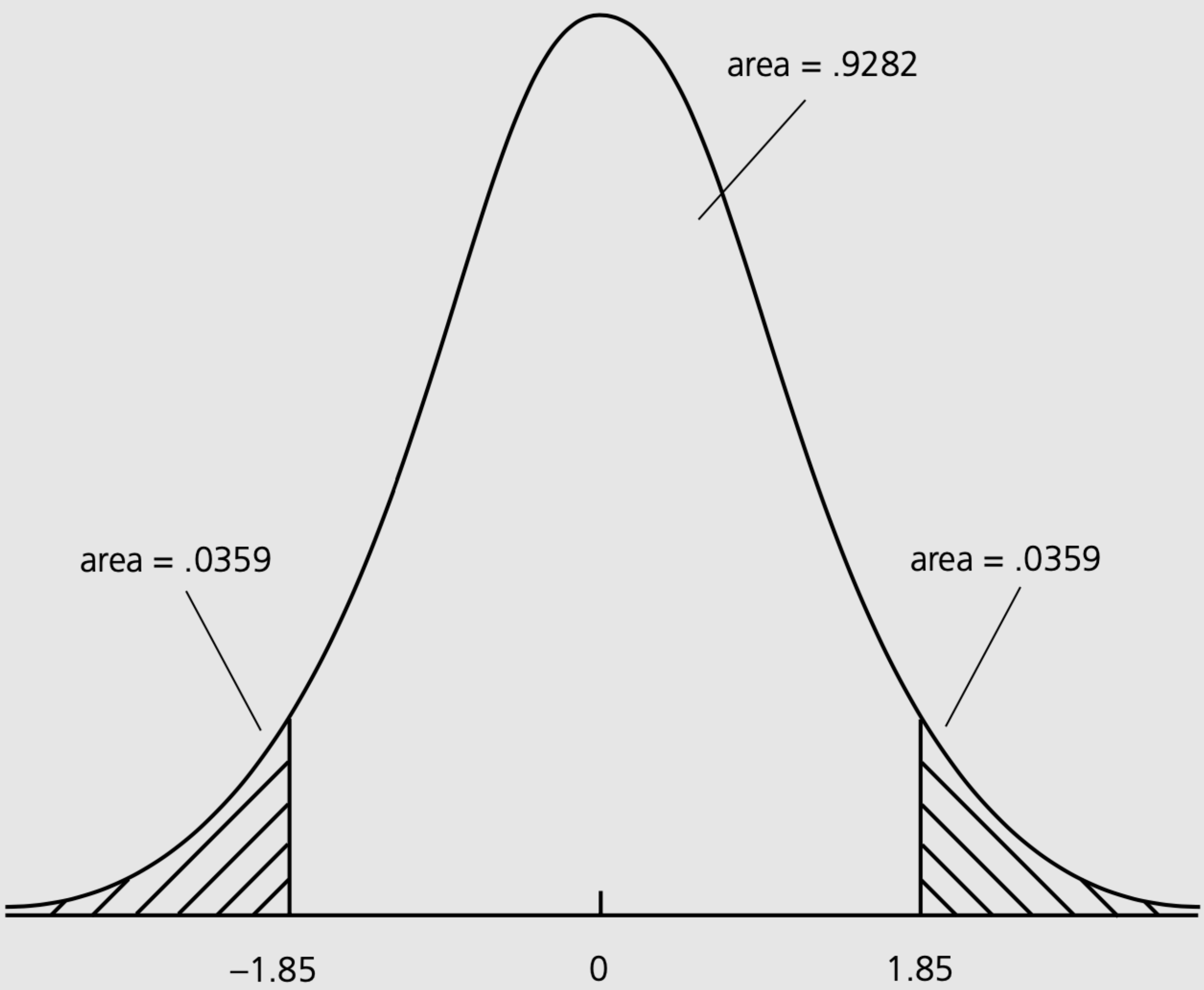
After computing the p-value, we can compare it with a chosen significance level α. Our decision rule will be

$$\begin{cases} p \leqslant \alpha, & \text{reject } H_0, \\ p > \alpha, & \text{do not reject } H_0. \end{cases}$$

In other words, to reject the null hypothesis, the p-value has to be *low enough*. For a common significance level of 5%, for example, this means that the p-value has to be lower than 0.05 to reject the null. If it is greater than 0.05, we do not reject the null.

For example, suppose our t statistic is 1.85 and we have 40 df. Then the p-value will be 2 $\times$ 0.036 = 0.072. We will reject the null at a 10% significance level, however, we will not reject the null at a 5% level.

Obtaining the p-value against a two-sided alternative, when t = 1.85 and df = 40.



Confidence intervals

Another useful quantity we can compute is a *confidence interval*. A $(1-\alpha)\times 100\%$ confidence interval (CI) for a coefficient estimate is computed as

$$\left[\hat{\beta}_j - c \times se(\hat{\beta}_j), \hat{\beta}_j + c \times se(\hat{\beta}_j)\right],$$

where c is the $\frac{1-\alpha}{2}$ percentile of the t distribution with $n-k-1\ df$.

For example, suppose our coefficient estimate is 2.04, the standard error is 0.56, and we have 40 df. For a 95% CI, the critical value c is 2.021, which is the 97.5th percentile of the t distribution with 40 df. Then the 95% CI is

$$[20.4 - 2.021 \times 0.56, 20.4 + 2.021 \times 0.56]$$
$$[0.908, 3.172]$$

The meaning of a 95% CI is the following. If we drew many random samples over and over again and computed CIs, then 95% of those intervals would contain the true value of β_j. For a given sample, we do not know whether the true value is contained in the CI or not, but there is a 95% chance that our CI does contain the true value.

Using a CI, we can also conduct hypothesis testing. Suppose our null hypothesis is $H_0 : \beta_j = b$. Then we would reject this null in favor of a two-sided alternative at a 5% significance level if the 95% CI does not contain b.

Example

Let's return to our trade data and estimate the following model with four predictors

$$\ln(imports_i) = \beta_0 + \beta_1 \ln(gdp_i) + \beta_2 \ln(distance_i) + \beta_3 liberal_i + \beta_4 \ln(area_i) + u_i.$$

The table below shows the coefficients estimates, standard errors, t statistics, p-values, and 95% confidence intervals.

	/ Est.	/ S.E.	/ t	/ p	/ 2.5 %	/ 97.5 %
(Intercept)	2.451	2.132	1.150	0.257	-1.848	6.750
log(gdp)	1.030	0.077	13.436	0.000	0.875	1.185
log(distance)	-0.888	0.156	-5.688	0.000	-1.203	-0.573
liberal	0.333	0.207	1.607	0.115	-0.085	0.751
log(area)	-0.159	0.086	-1.857	0.070	-0.331	0.014