Multiple Regression Analysis

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$

- 2. Inference

Assumptions of the Classical Linear Model (CLM)

- Knowing the expected value and variance of the OLS estimators is useful for describing the location and precision of the OLS estimators.
- However, in order to perform statistical inference, we need more than just the first two moments of $\hat{\beta}_j$'s. We need to know the full sampling distribution of the $\hat{\beta}_j$'s .

Assumptions of the Classical Linear Model (CLM)

- Even under the Gauss-Markov assumptions (MLR1-MLR5), the distribution of $\hat{\beta}_j$'s can have virtually any shape.
- In order to do classical hypothesis testing, we need to add another assumption, beyond the Gauss-Markov assumptions.
- Assume that u is independent of $x_1, x_2, \ldots, x_k$ and u is normally distributed with zero mean and variance σ^2: $u \sim \text{Normal}(0,\sigma^2)$.

Assumptions

- **MLR.6: NORMALITY**

  The error u is independent of x_1, x_2,…, x_k and u is normally distributed with zero mean and variance σ^2

  $$u \sim \text{Normal}(0,\sigma^2)$$

  If we make MLR.6, then we are necessarily assuming MLR.3 and MLR.5

CLM Assumptions

- Assumptions MLR.1 through MLR.6 are called the **classical linear model (CLM) assumptions**.
- It is best to think of the CLM assumptions as containing all the Gauss-Markov assumptions *plus* the assumption of a normally distributed error term.

CLM Assumptions

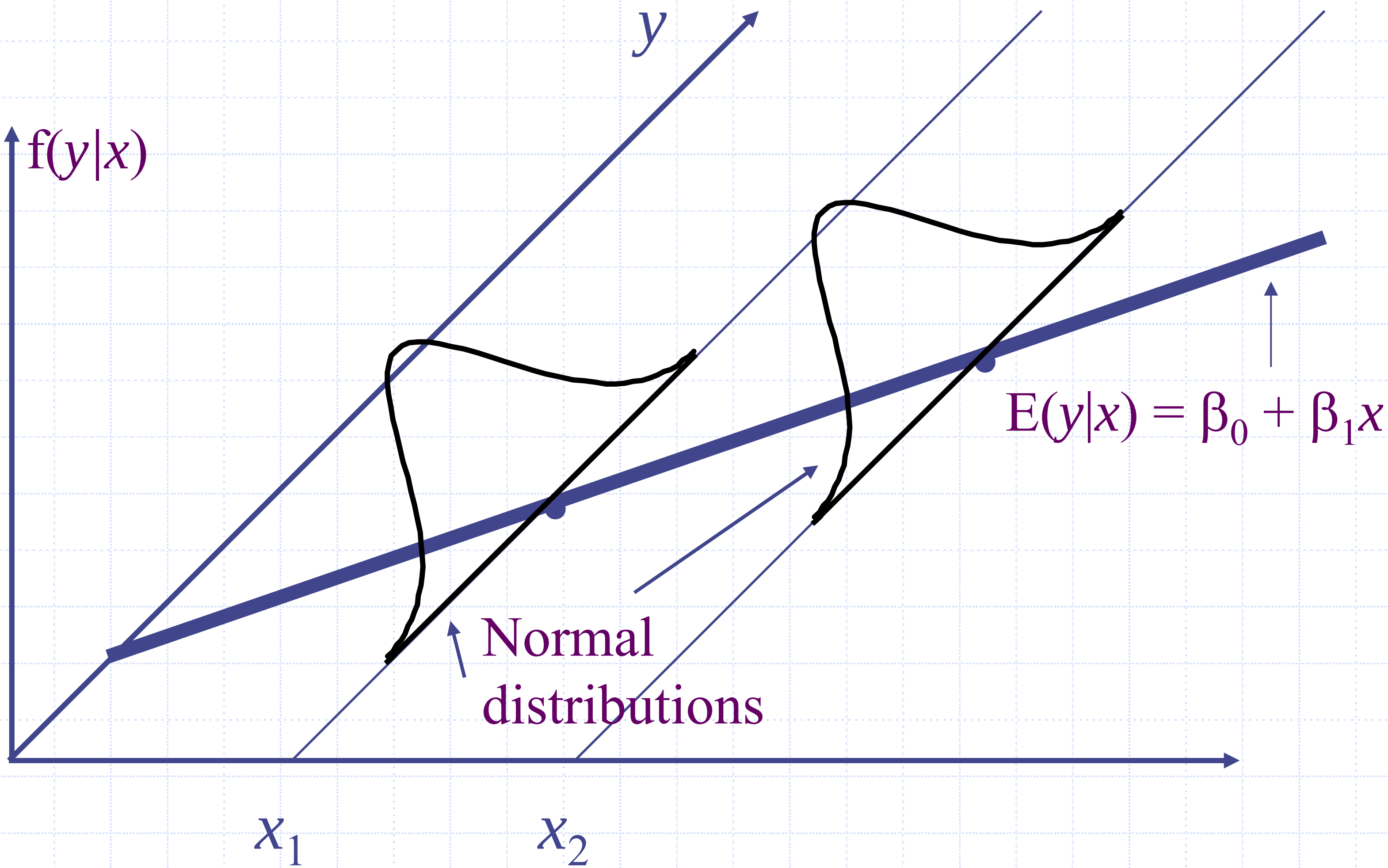
- Under CLM, **OLS** is not only BLUE, but is the **minimum variance unbiased estimator**.
- This means that OLS has the smallest variance among all unbiased estimators; we no longer have to restrict our comparison to estimators that are linear in the y_i.
- We can summarize the population assumptions of CLM as follows

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

Normality

- The claim for normality is usually done on the basis of a Central Limit Theorem (CLT), but this is restrictive in some cases.
- Normality cannot be assumed always.
- In any application, whether normality of u can be assumed is really an empirical matter.
- Often, using a transformation, i.e. taking log's, yields a distribution that is closer to normal.

The homoskedastic normal distribution with a single explanatory variable



Normality

- Normality is easy to handle from a mathematical point of view.
- Large samples will allow us to drop normality without affecting to much the results.
- Normality of the error term translates into normal sampling distributions of the OLS estimators.

Normality of OLS

- **THEOREM 4.1 NORMAL SAMPLING DISTRIBUTIONS**

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

$$\hat{\beta}_j \sim Normal\left[\beta_j ; Var(\hat{\beta}_j)\right]$$

  where $Var(\hat{\beta}_j)$ was given in topic 3. Therefore,

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

Normality of OLS

- The proof of Theorem 4.1 is not difficult, and is based on the fact that $\hat{\beta}_j$ is a linear combination of the errors, that are independent normal variables, and the important fact that a linear combination of normal variables has a normal distribution (Appendix B).
- Previously we showed that $E(\hat{\beta}_j) = \beta_j$ and derived $Var(\hat{\beta}_j)$

Normality of OLS

- The conclusion of Theorem 4.1 can be strengthened.
- In fact, any linear combination of the

$$\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$$

is also normally distributed, and any subset of the $\hat{\beta}_j$'s has a joint normal distribution.
- In the next topic we shall show that normality holds as an approximation even without MLR.6.

Hypothesis Testing: Fundamentals

- Hypothesis testing entails making a decision, on the basis of sample data, whether to accept that certain restrictions are satisfied by the basic assumed model.
- The restrictions we are going to test are known as the *null hypothesis*, denoted by H_0.
- We also define an *alternative hypothesis*, denoted by H_1, which represents our conclusion if the experimental test indicates that H_0 is false.

Hypothesis Testing: Example

- In the SLR, $y = \beta_0 + \beta_1 x + u$, we may be interested in the hypothesis

$$H_0\text{: } \beta_1 = 1$$

versus

$$H_1\text{: } \beta_1 = 0$$

- In order to conclude that H_0 is false and that H_1 is true, we must have evidence "beyond reasonable doubt" against H_0.

Hypothesis Testing: Example

- Usually we are not so specific about H_1, so this is generally stated as

  $$H_1 : \beta_1 < 1 \qquad \text{(one side alternative)}$$

  or

  $$H_1 : \beta_1 \neq 1 \qquad \text{(two side alternative)}$$

- **Important**: We are testing hypothesis about the *population* parameters. We are not testing hypothesis about the estimates from a particular sample.

Hypothesis Testing: Fundamentals

- In hypothesis testing, we can make two kinds of mistakes.

1. First, we can reject H_0 when it is in fact true. This is called **Type I error**.
2. Second, we can fail to reject H_0 when it is actually false. This is called **Type II error**.

Hypothesis Testing: Fundamentals

- After we have made the decision of whether or not to reject H_0, we have either decided correctly or we have committed an error. We shall never know with certainty whether an error was committed.
- However, we can compute the *probability* of making either a Type I error or a Type II error.

Hypothesis Testing: Fundamentals

- Hypothesis testing rules are constructed to make the probability of committing a Type I error fairly small.
- Generally, we define the **significance level** (α) of a test as the probability of a Type I error. Symbolically

  $$\alpha = \Pr(\text{Reject } H_0 \mid H_0)$$

  Read as: "The probability of rejecting H_0 given that H_0 is true."

Hypothesis Testing: Fundamentals

- Classical hypothesis testing requires that we initially specify a **significance level** for the test. When we specify a value for α, we are essentially quantifying our tolerance for Type I error.
- Common values for α are .10, .05 and .01.
- If $\alpha = .05$, then the researcher is willing to falsely reject H_0 5% of the time, in order to detect deviations from H_0.

Hypothesis Testing: Fundamentals

- Once we have chosen α, we would then like to minimize the probability of a Type II error, $Pr(\text{Fail to Reject } H_0 \mid H_1)$
- Alternatively, we would like to maximize the **power of a test** against all relevant alternatives. The power of a test, $\pi(\beta_1)$, is

$$\pi(\beta_1) = 1 - Pr(\text{Fail to Reject } H_0 \mid \beta_1) = Pr(\text{Reject } H_0 \mid \beta_1)$$

where β_1 denotes the actual value of the parameter.

Hypothesis Testing: Fundamentals

- Naturally, we would like the power function to be 1 under H_1 (a false null) and 0 under H_0 (a true null).
- But this is not possible!.
- It can be shown that there is a trade-off between both types of errors, so reducing Type I error increases Type II error.
- Instead, we choose our tests to maximize the power for a given α.

Hypothesis Testing: Fundamentals

- In order to test a null hypothesis against an alternative, we need to choose a test statistic and a critical value.
- The choices for the statistic and the critical value are based on convenience and on the desire to maximize power given a significance level for the test.
- A **test statistic**, *T*, is some function of the random sample, so is itself a random variable.

Hypothesis Testing: Fundamentals

- When we compute the statistic for a particular sample, we obtain an outcome of the test statistic, say *t*.
- In order to perform an statistical test we should know the distribution of *T* under the null hypothesis.
- Given the test statistic and its distribution, we can define a rejection rule that determines when H_0 is rejected in favor of H_1.

Hypothesis Testing: Fundamentals

- In this course, all rejection rules are based on comparing the value of the test statistic, t, to a **critical value**, c.
- The set of values of t that result in rejection of the null hypothesis are collectively known as the rejection region.
- In order to determine the critical value, we must first decide on a significance level of the test, α.

Hypothesis Testing: Fundamentals

- Then, given α, the critical value associated with α is determined by the distribution of T, *assuming* that H_0 is true.
- We shall write this critical value as c, but it should be understood that c depends on α.

Hypothesis Testing: CLM

- We are now interested testing hypothesis about a single population parameter in the context of the CLM,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

- Even if β_j is unknown, we can hypothesize about the value of β_j and then use statistical inference to test our hypothesis.
- The main result we need is the next one.

Hypothesis Testing: *t* test

- **THEOREM 4.2 *t* DISTRIBUTION FOR THE STANDARDIZED ESTIMATORS**

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

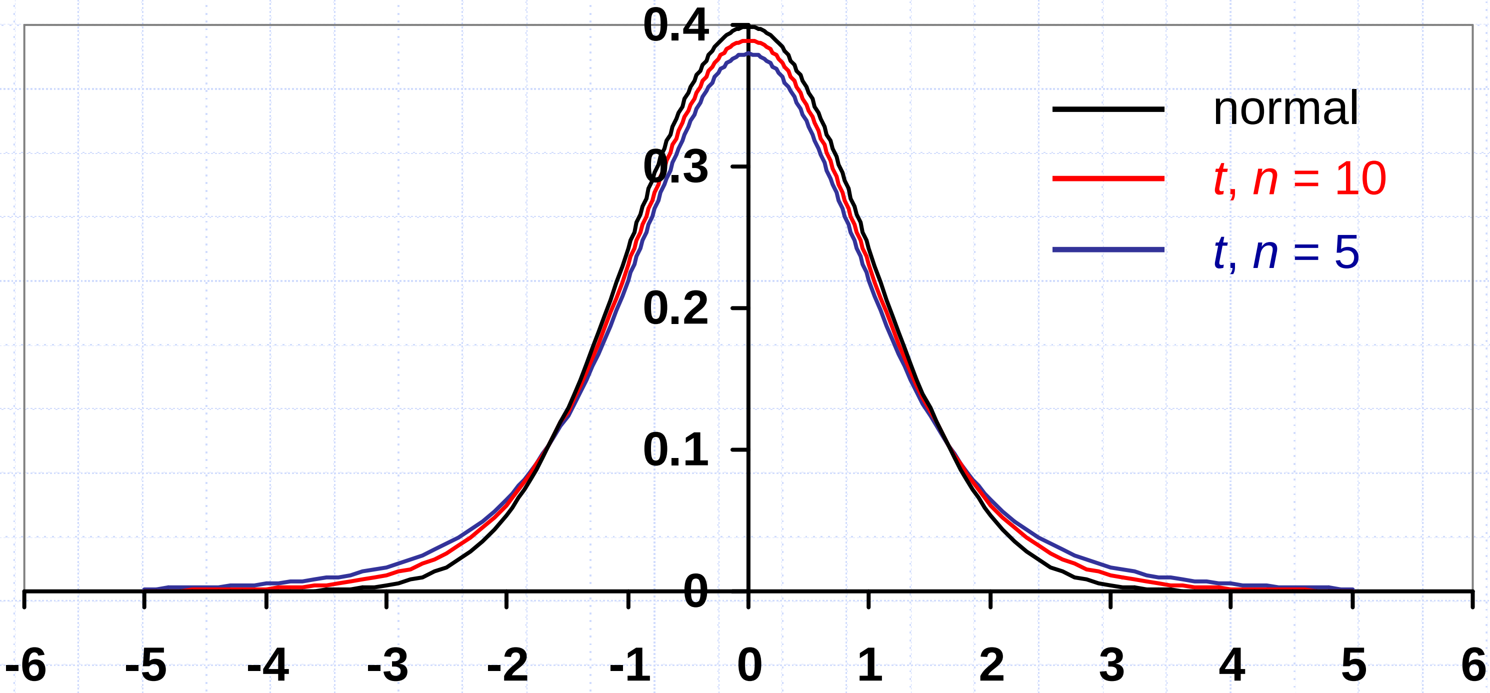
$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$$

  where $k + 1$ is the number of unknown parameters in the population model (k slope parameters and the intercept, β_0).

Hypothesis Testing: *t* test

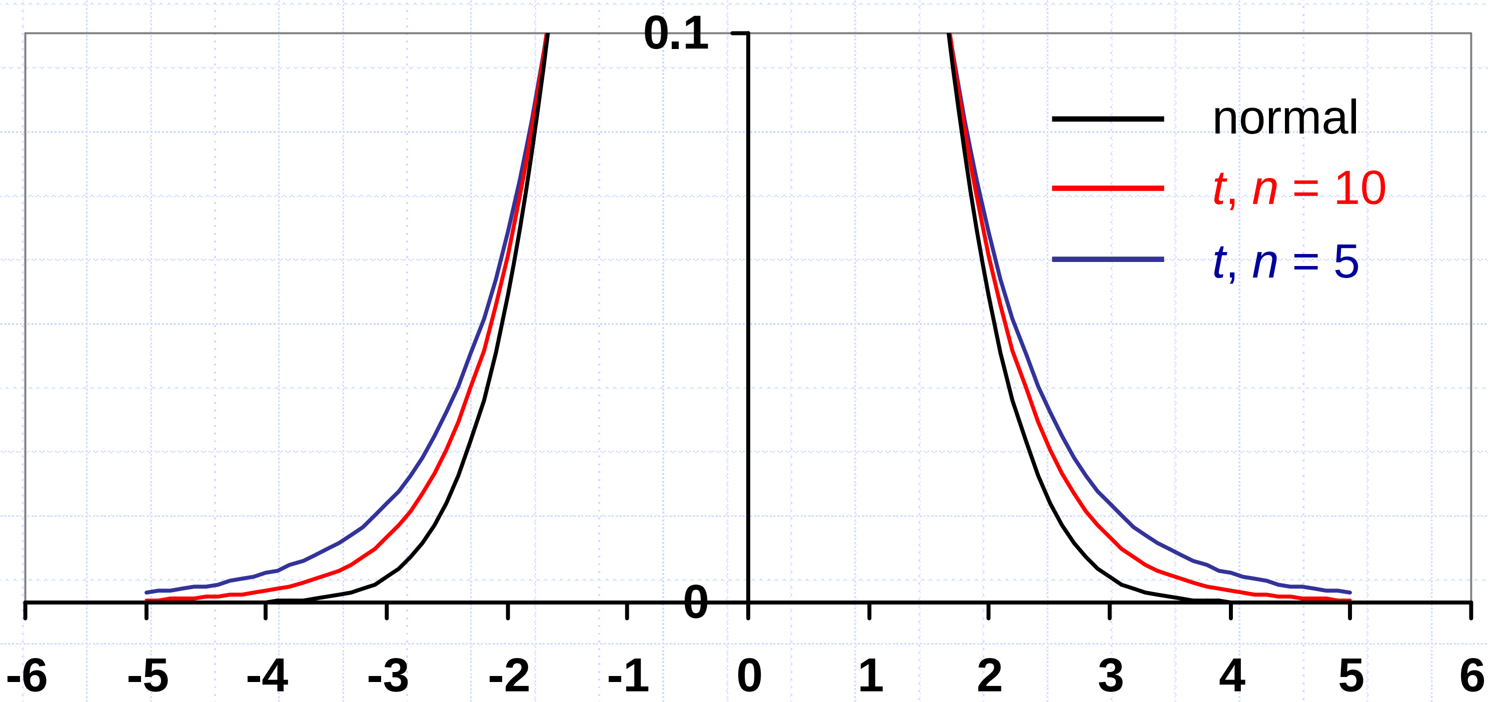
- Theorem 4.2 is important in that it allows as to test hypothesis involving the β_j.
- Compare this result with Theorem 4.1. The *t* distribution comes from the fact that the constant σ in $sd(\hat{\beta}_j)$ has been replaced with the random variable $\hat{\sigma}$.
- Are the normal and *t* distributions very different?.

Normal *versus* *t* distributions



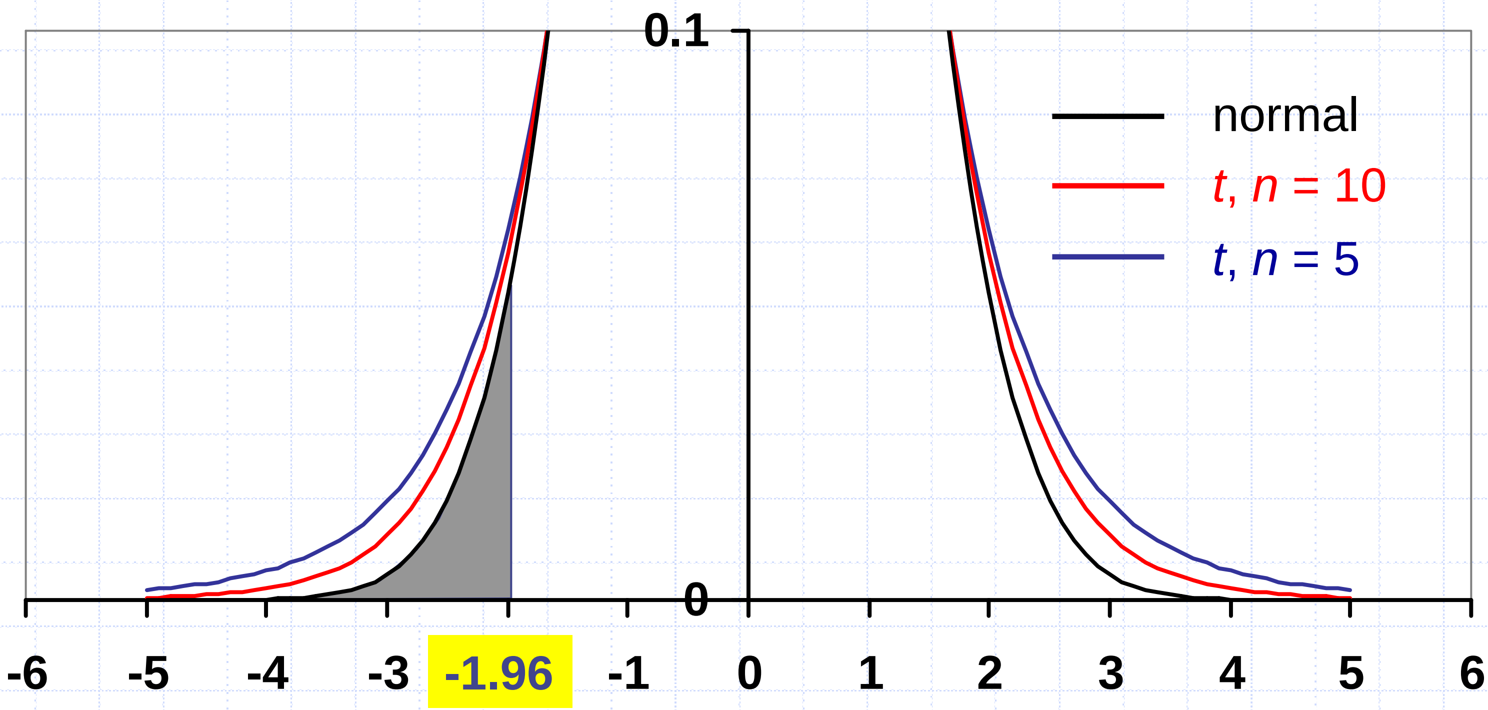
Generally not, they are very similar in shape. Both symmetric and centered around zero, but…

Normal *versus* *t* distributions



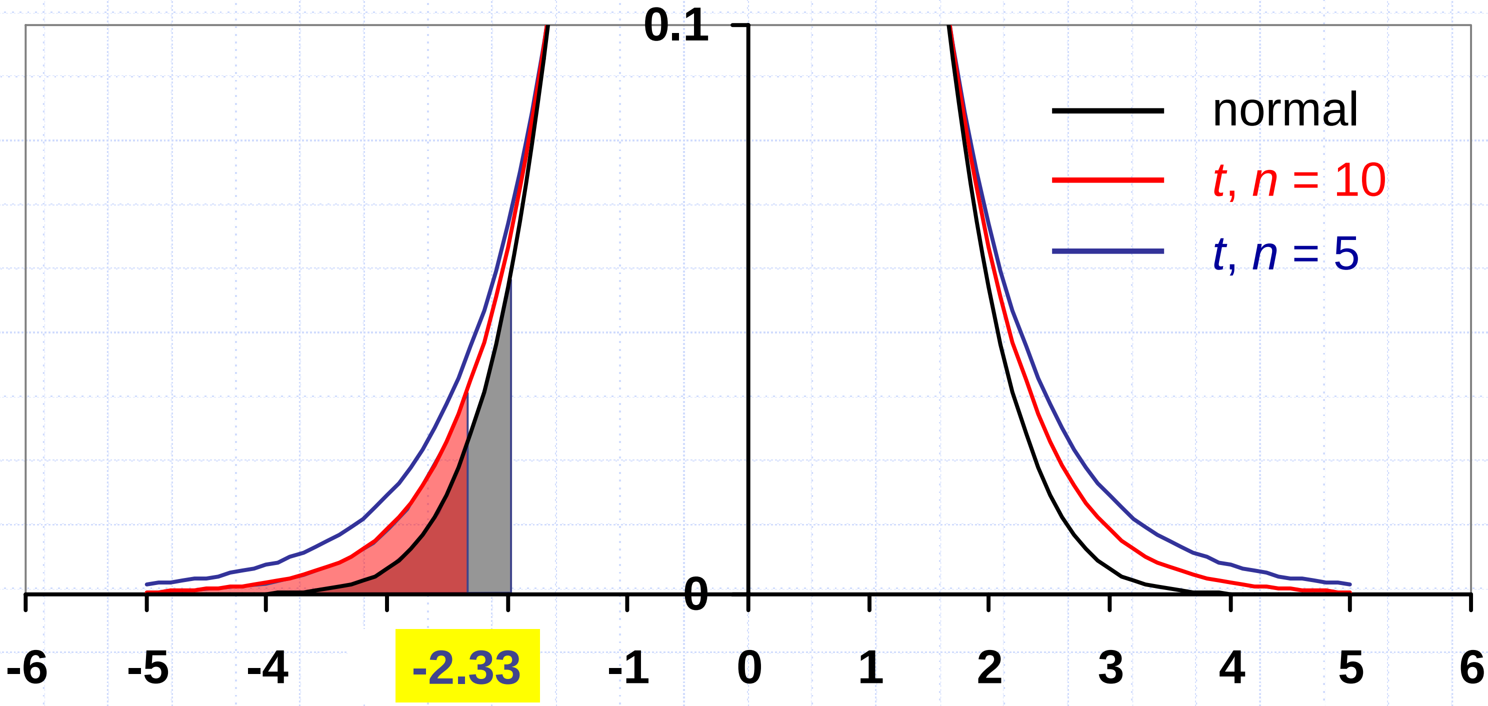
they are quite different at the tails. And these are the important parts of the distributions in hypothesis testing.

Normal *versus* *t* distributions



The 2.5% tail of a normal distribution starts 1.96 standard deviations from its mean.

Normal *versus t* distributions



The 2.5% tail of a *t* distribution with 10 degrees of freedom starts 2.33 standard deviations from its mean.

Hypothesis Testing: *t* test

- Consider the null hypothesis,

$$H_0\!: \beta_j = 0$$

- Since β_j measures the partial effect of x_j on y, after controlling for all other independent variables, $H_0\!: \beta_j = 0$ means that, once $x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k$ have been accounted for, x_j has *no effect* on y.
- This is called a **significance test**.

Hypothesis Testing: *t* test

- The statistic we use to test H_0: $\beta_j = 0$, against any alternative, is called "the" ***t* statistic** or "the" ***t* ratio** of $\hat{\beta}_j$ and is defined as

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

- Why is this a good statistic to test this hypothesis?

Hypothesis Testing: *t* test

- Since $se(\hat{\beta}_j) > 0$, $t_{\hat{\beta}_j}$ will have the same sign as $\hat{\beta}_j$.
- In order to test H_0: $\beta_j = 0$; first, it is natural to look at our unbiased estimator of β_j, $\hat{\beta}_j$.
- In a given sample $\hat{\beta}_j$ will never be zero exactly, but a small value will indicate a true null, whereas a big value will indicate a false null.
- The question is: How far is $\hat{\beta}_j$ from zero?

Hypothesis Testing: *t* test

- We must recognize that there is a sampling error in our estimate $\hat{\beta}_j$, so the size of $\hat{\beta}_j$ must be weighted against its sampling error.
- This is precisely what we do using $t_{\hat{\beta}_j}$, since this statistic measures how many estimated standard deviations $\hat{\beta}_j$ is away from zero.

One-Sided Alternatives

- In order to determine a rule for rejecting H_0, we need to decide on the relevant **alternative hypothesis**.
- First, consider a one-sided alternative of the form

$$H_1: \beta_j > 0$$

- How should we choose a rejection rule?

One-Sided Alternatives

- First, decide on a significance level, α, or the probability of rejecting H_0 when it is in fact true, i.e. $\alpha = .05$.
- While $t_{\hat{\beta}_j}$ has a *t* distribution under H_0, so it has zero mean, under the alternative $\beta_j > 0$, so the expected value of $t_{\hat{\beta}_j}$ is positive.
- Thus we are looking for a "sufficiently large" positive value of $t_{\hat{\beta}_j}$ in order to reject H_0: $\beta_j = 0$ in favor of H_1: $\beta_j > 0$.

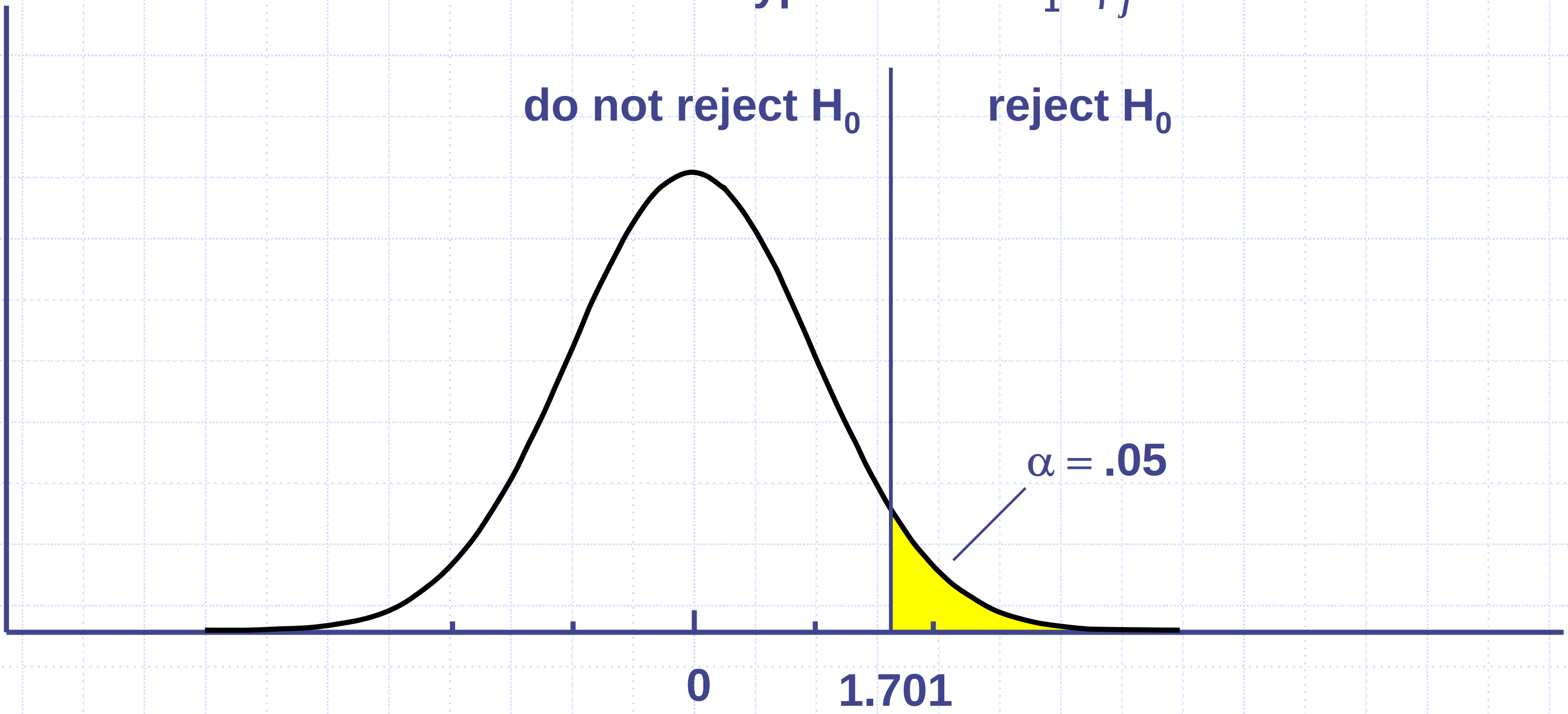
One-Sided Alternatives

- Negative values of $t_{\hat{\beta}_j}$ provide no evidence in favor of H_1: $\beta_j > 0$.
- The definition of a "sufficiently large", with $\alpha = .05$, is the 95^{th} percentile in a t distribution with $n - k - 1$ degrees of freedom, say c.
- **Rejection rule**: Reject H_0: $\beta_j = 0$ in favor of H_1: $\beta_j > 0$ at $\alpha = .05$ if $t_{\hat{\beta}_j} > c$

One-Sided Alternatives

null hypothesis: $H_0 : \beta_j = 0$

alternative hypothesis: $H_1 : \beta_j > 0$



Example of a rejection rule for 28 degrees of freedom.

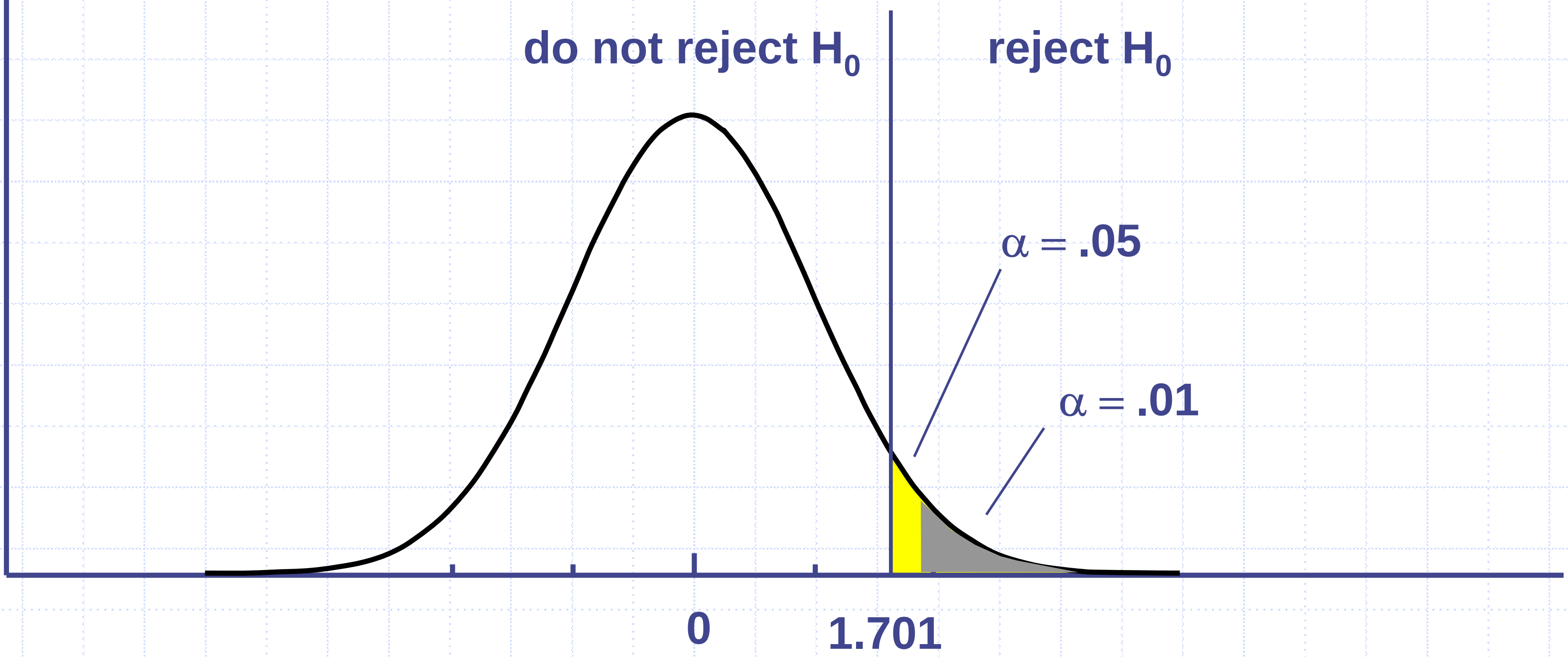
One-Sided Alternatives

- By our choice of the critical value *c*, rejection of H_0 will occur for 5% of all random samples when H_0 is true.
- This is an example of a **one-tailed test**.
- In order to obtain *c*, we only need the significance level and the degrees of freedom.
- You should note a pattern in the critical values of the *t* distribution: as α falls, *c* increases.

One-Sided Alternatives

null hypothesis: $H_0 : \beta_j = 0$

alternative hypothesis: $H_1 : \beta_j > 0$



Thus, if H_0 is rejected at, say, the 1% level, then it is automatically rejected at the 5% level.

t distribution

t* Distribution: Critical values of *t

Degrees of freedom	One-tailed test	5%	2.5%	1%	0.5%	0.1%	0.05%
1		6.314	12.706	31.821	63.657	318.31	636.62
2		2.920	4.303	6.965	9.925	22.327	31.598
3		2.353	3.182	4.541	5.841	10.214	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
…		…	…	…	…	…	…
…		…	…	…	…	…	…
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
…		…	…	…	…	…	…
…		…	…	…	…	…	…
120		1.658	1.980	2.358	2.617	3.160	3.373
∞		1.645	1.960	2.326	2.576	3.090	3.291

Normal *versus t* distributions

- As the degrees of freedom (*df*) in the *t* distribution gets large, the *t* distribution approaches the standard normal distribution.
- As a practical rule for *df* larger than 120 we can take the critical values from the normal.

One-Sided Alternatives

- Consider the one sided alternative,

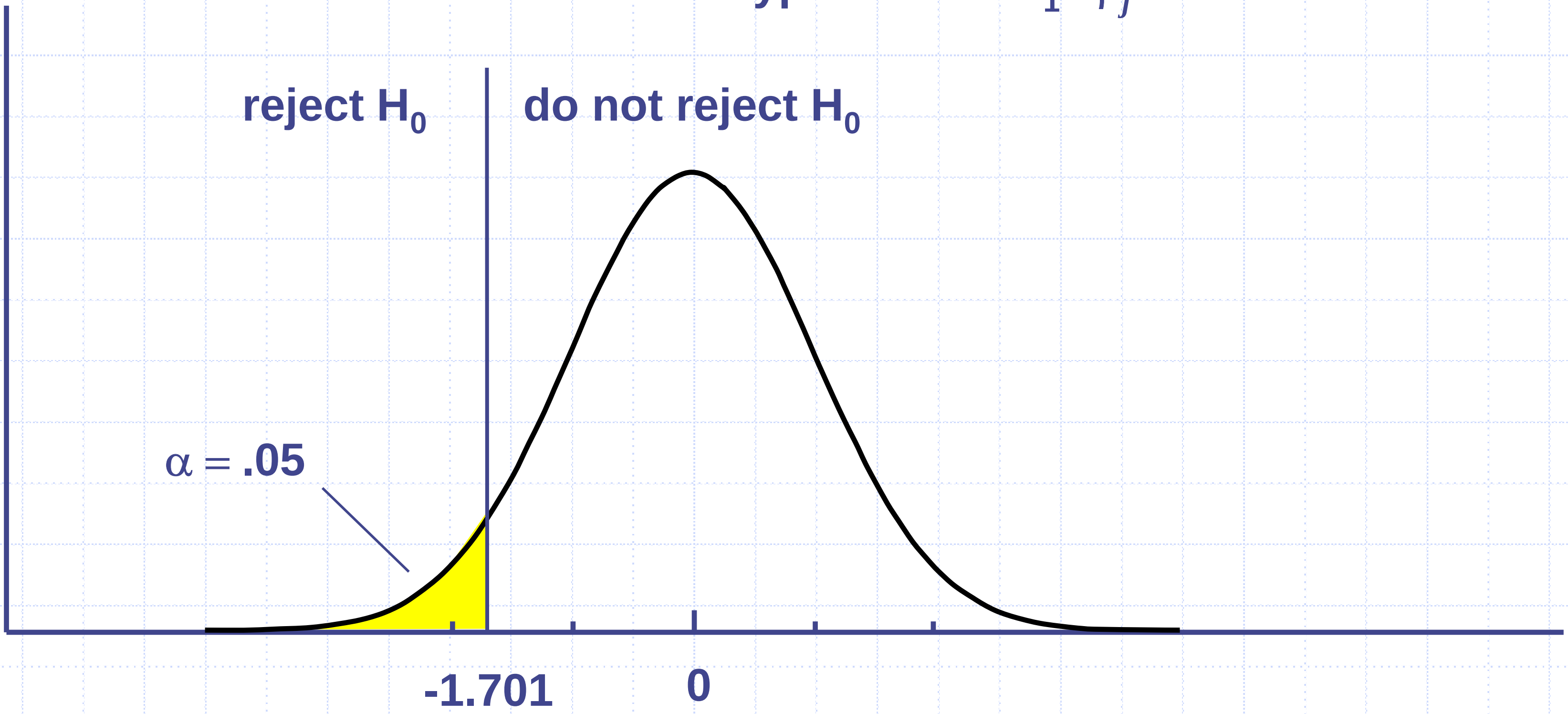
$$H_1\text{: } \beta_j < 0$$

- The rejection rule is just the mirror image of the previous case. Now, the critical value comes from the left tail of the *t* distribution.
- In practice, it is easiest to think of the **rejection rule** as: Reject H_0: $\beta_j = 0$ in favor of H_1: $\beta_j < 0$ at $\alpha = .05$ if $t_{\hat{\beta}_j} < -c$, where c is the critical value for H_1: $\beta_j > 0$.

One-Sided Alternatives

null hypothesis: $H_0 : \beta_j = 0$

alternative hypothesis: $H_1 : \beta_j < 0$



Example of a rejection rule for 28 degrees of freedom.

One-Sided Alternatives

- For simplicity, we always assume *c* is positive, since this is how critical values are reported in *t* tables, and so –*c* is a negative number.
- To reject H_0 against the alternative H_1: $\beta_j < 0$, we must get a negative *t* statistic. A positive *t* ratio, no matter how large, provides no evidence in favor of H_1: $\beta_j < 0$.

Two-Sided Alternatives

- Consider now the null hypothesis

$$H_0\colon \beta_j = 0$$

against a two sided-alternative

$$H_1\colon \beta_j \neq 0$$

- Under H_1, x_j has a ceteris paribus effect on y without specifying whether the effect is positive or negative. This is the relevant alternative when the sign of β_j is not well determined by theory or common sense.

Two-Sided Alternatives

- When the alternative is two-sided, we are interested in the *absolute value* of the *t* statistic.
- **Rejection rule**: Reject H_0: $\beta_j = 0$ in favor of H_1: $\beta_j \neq 0$ at $\alpha = .05$ if $\left|t_{\hat{\beta}_j}\right| > c$, where $|\bullet|$ denotes absolute value and *c* is an appropriately chosen critical value.

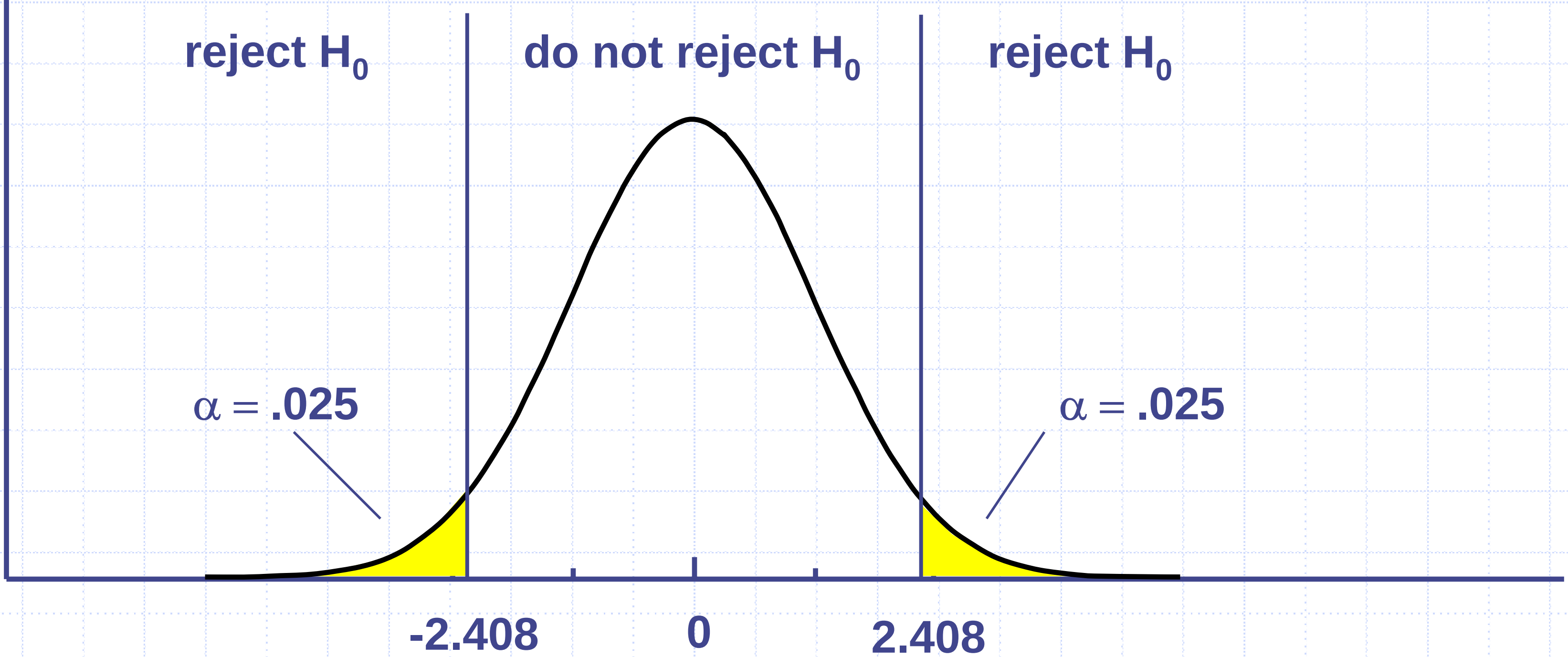
Two-Sided Alternatives

- Given α, for a **two-tailed test**, c is chosen to make the area in each tail of the t distribution equal to $\alpha/2$.
- Hence, for $\alpha = .05$, c is chosen to make the area in each tail of the t distribution equal to 0.025.
- In other words, c is the 97.5th percentile in the t distribution with $n - k - 1$ *df*.

Two-Sided Alternatives

null hypothesis: $H_0 : \beta_j = 0$

alternative hypothesis: $H_1 : \beta_j \neq 0$

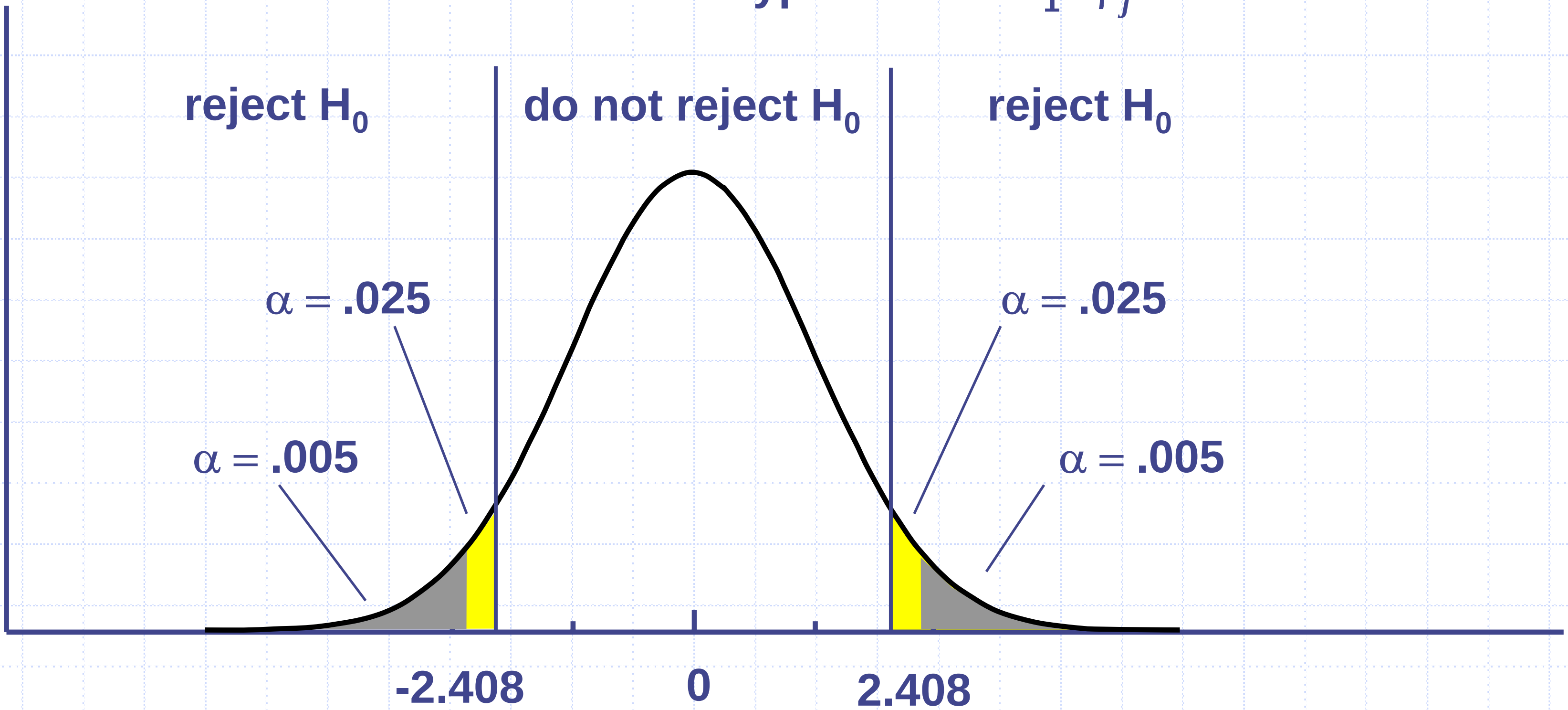


Example of a rejection rule for 28 degrees of freedom and a two-sided alternatives.

Two-Sided Alternatives

null hypothesis: $H_0: \beta_j = 0$

alternative hypothesis: $H_1: \beta_j \neq 0$



As before, if H_0 is rejected at, say, the 1% level, then it is automatically rejected at the 5% level.

t distribution

t* Distribution: Critical values of *t

Degrees of freedom	Two-tailed test / One-tailed test	10% / 5%	5% / 2.5%	2% / 1%	1% / 0.5%	0.2% / 0.1%	0.1% / 0.05%
1		6.314	12.706	31.821	63.657	318.31	636.62
2		2.920	4.303	6.965	9.925	22.327	31.598
3		2.353	3.182	4.541	5.841	10.214	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
…		…	…	…	…	…	…
…		…	…	…	…	…	…
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
…		…	…	…	…	…	…
…		…	…	…	…	…	…
120		1.658	1.980	2.358	2.617	3.160	3.373
∞		1.645	1.960	2.326	2.576	3.090	3.291

Two-Sided Alternatives

- 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 at $\alpha = 5\%$, we usually say that "x_j is **statistically significant** at the 5% level".
- If H_0 is not rejected, we say that "x_j is **statistically insignificant** at the 5% level".

Two-Sided Alternatives

- The two-sided significance test we have just seen is calculated routinely by regression software for each variable included in a model.
- These tests and associate probability values are reported together with estimates and standard errors.

Other Hypothesis about β_j

- Consider now the null hypothesis

$$H_0\colon \beta_j = a_j$$

- Then the appropriate t statistic is $t_{\hat{\beta}_j} = \dfrac{\hat{\beta}_j - a_j}{se(\hat{\beta}_j)}$

- As before, $t_{\hat{\beta}_j}$ measures how many estimated standard deviations $\hat{\beta}_j$ is away from the hypothesized value of β_j.

- In general: $t_{\hat{\beta}_j} = \dfrac{\text{estimate} - \text{value under } H_0}{\text{standar error}}$

Other Hypothesis about β_j

- The rejection rule, one or two tails, depend on the form of the alternative.

Computing *p*-values for *t* Tests

- We have explained the classical approach:

1. State H_0 and H_1, the last one either explicitly or implicitly.
2. Choose α, which determines c (i.e. the rejection region).
3. Compare the value of the t statistic with c.
4. Eventually H_0 is either rejected or nor rejected at the given α.

Computing *p*-values for *t* Tests

- To some extend the classical approach is in some sense arbitrary, since we have to choose α in advance, and eventually H_0 is either rejected or not.
- If H_0 is eventually rejected, we don't know if this rejection is strong o weak. And the same is true if H_0 is not rejected.

Computing *p*-values for *t* Tests

- Instead of testing at a given α, consider de following question: "Given the observed value of the *t* statistic, what is the smallest significance level at which H_0 would be rejected? This level is known as the ***p*-value** for the test".
- That is, the *p*-value is the significance level of the test when we use the value of the test statistic as the critical value for the test.

Computing *p*-values for *t* Tests

- **Example**: Assume $df = 40$ and $t_{\hat{\beta}_j} = 1.85$.

  What would be the *p*-value for a two-tailed tests?

$$p\text{-value} = \Pr(|T| > 1.85) = 2.\Pr(T > 1.85) = 2(.0359) = .0718$$

  This means that, if the H_0 is true, we would observe an absolute value of the *t* statistic as large as 1.85 about 7.2% of the time.

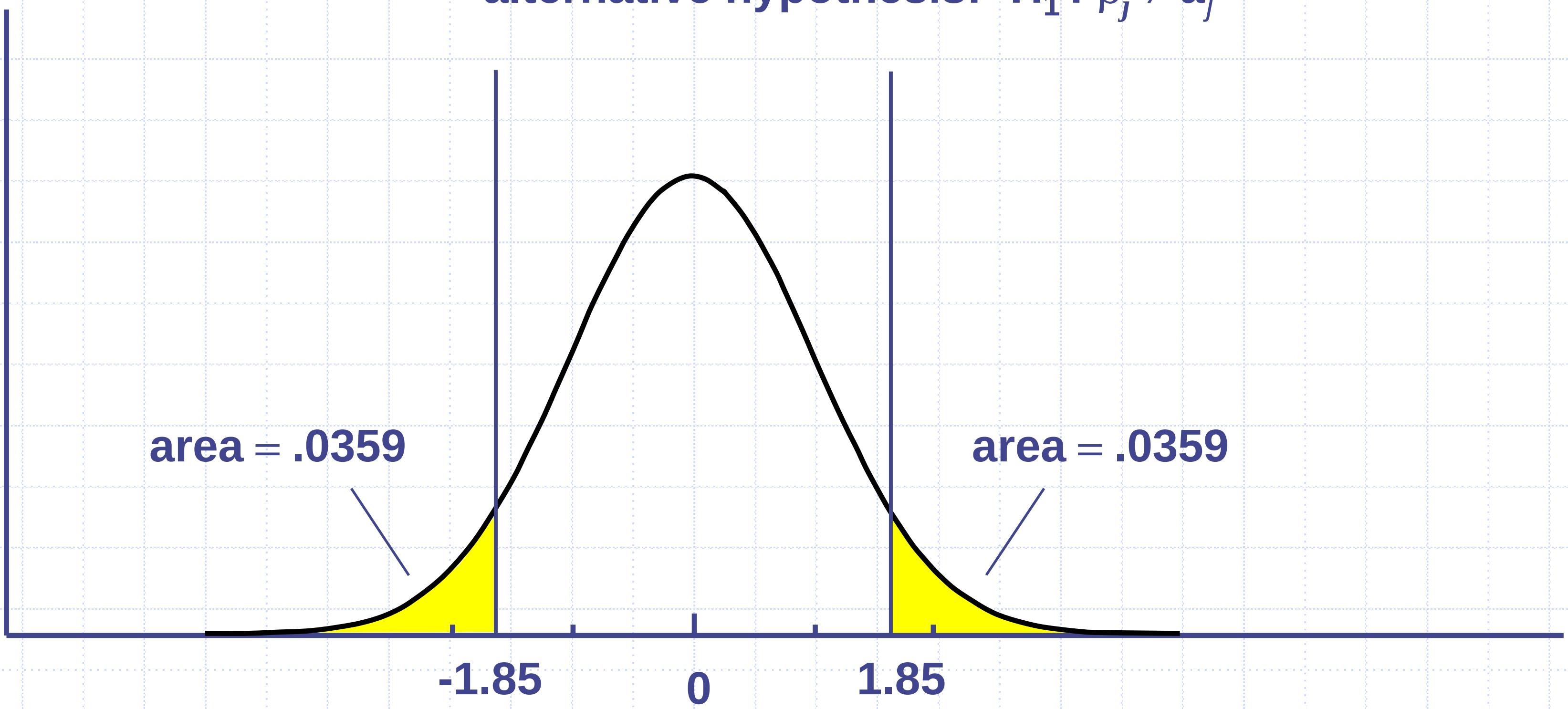
  This provides some evidence against H_0, since it would be rejected at $\alpha = 10\%$, but not at $\alpha = 5\%$.

  What would be the *p*-value for a positive one-tailed alternative? And for a negative one?

Computing *p*-values for *t* Tests

null hypothesis: $H_0: \beta_j = a_j$

alternative hypothesis: $H_1: \beta_j \neq a_j$



The *p*-value takes as the critical value the observed value of the test statistic, and from this computes the significance level of the test.

Computing *p*-values for *t* Tests

- This example shows that once the *p*-value has been computed, a classical test can be carried out at any desired level.
- If α denotes the significance level of the test, then H_0 is rejected if *p*-value $< \alpha$; otherwise, H_0 is not rejected at the $100.\alpha\%$ level.

Computing *p*-values for *t* Tests

- The *p*-value nicely summarizes the strength or weakness of the empirical evidence against H_0.
- A useful interpretation is the following: the *p*-value is the probability of observing a *t* statistic as extreme as we did *if the null hypothesis is true*.
- This means that small *p*-values are evidence against H_0; large *p*-values provide little evidence against H_0.

Computing *p*-values for *t* Tests

- Since a *p*-value is a probability, its value is always between 0 and 1.
- To compute *p*-values we need a very detailed statistical tables or a computer program that computes areas under probability density functions.
- You don't have to worry about because statistical software computes *p*-values for all statistical tests.

A Note on Terminology

- When H_0 is not rejected, we prefer to say: “we fail to reject H_0 at the α% level”, rather than “H_0 is accepted at the α% level”.
- The reason why the former is preferred is that, if we change the value of H_0 a little bit, we can also “accept” this new hypothesis, which is meaningless. We cannot “accept” both of these hypothesis.
- All we can say is that the data does not allow us to reject either of these hypothesis. So our sample is consistent with both of them.

Economic *versus* Statistical Significance

- So far we have emphasized statistical significance. However it is important to remember that we should pay attention to the magnitude of the coefficient estimates in addition to the *t* statistics.
- **Statistical significance** of a variable x_j is determined entirely by the size of $t_{\hat{\beta}_j}$, whereas the **economic significance** of a variable is related to the size (and sign) of $\hat{\beta}_j$.

Economic *versus* Statistical Significance

- Too much focus on statistical significance can lead to the false conclusion that a variable is "important" for explaining y even though its estimated effect is modest.
- So even if a variable is statistically significant, you need to discuss the magnitude of the estimated coefficient to get an idea of its practical or economic importance.
- This step requires some care, depending on how the variables appear in the equation.

Confidence Intervals

- Under the CLM, we can easily construct a **confidence interval** (CI) for the population parameter, β_j .
- CI are also called interval estimates, because they provide a range of likely values for β_j , and not just a point estimate.

Confidence Intervals

- Using the fact that

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

a simple manipulation leads to a CI for the unknown β_j.

- A 95% CI, is given by $\hat{\beta}_j \pm c.se(\hat{\beta}_j)$

where c is the 97.5th percentile in a t_{n-k-1} distribution.

Confidence Intervals

- In general a $(1-\alpha)\%$ CI is defined as

$$\hat{\beta}_j \pm c.se(\hat{\beta}_j)$$

where c is the $\left(1-\frac{\alpha}{2}\right)$ percentile in a t_{n-k-1} distribution.

Confidence Intervals

- More precisely, the lower and upper bounds of the confidence interval are given by

$$\underline{\beta}_j \equiv \hat{\beta}_j - c.se(\hat{\beta}_j)$$

and

$$\overline{\beta}_j \equiv \hat{\beta}_j + c.se(\hat{\beta}_j)$$

respectively.

Confidence Intervals: Meaning

- If random samples were obtained over and over again, with $\underline{\beta}_j$, and $\overline{\beta}_j$ computed each time, then the (unknown) population value β_j would lie in the interval $(\underline{\beta}_j, \overline{\beta}_j)$ for $(1 - \alpha)\%$ of the samples.
- 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.

Confidence Intervals

- Once a CI is constructed, it is easy to carry out two-tailed hypothesis tests.
- If the null hypothesis is H_0: $\beta_j = a_j$, then H_0 is rejected against H_1: $\beta_j \neq a_j$ at (say) the 5% significance level if, and only if, a_j is *not* in the 95% CI.
- Hence all values contained in the CI are consistent with our data, in the sense that wouldn't be rejected in a two-tailed test.

Testing a Linear Combination

- In many applications we are interested in testing hypothesis involving more than one of the population parameters.
- **Example**: Cobb-Douglas Production Function

$$\log(y) = \beta_0 + \beta_1 \log(l) + \beta_2 \log(k) + u$$

Hypothesis of interest: Constant Returns

$$H_0: \beta_1 + \beta_2 = 1$$

Against

$$H_1: \beta_1 + \beta_2 \neq 1$$

Testing a Linear Combination

- We cannot simply use the individual *t* statistics for $\hat{\beta}_1$ and $\hat{\beta}_2$ to test H_0.
- However, it is easy to see that the *t* statistic is now based on whether the estimated sum $\hat{\beta}_1 + \hat{\beta}_2$ is sufficiently different from one to reject H_0 in favor of H_1.

Testing a Linear Combination

- To account for the sampling error in our estimators, we standardize this sum by dividing by the standard error

$$t_{\hat{\beta}_1+\hat{\beta}_2} = \frac{\hat{\beta}_1 + \hat{\beta}_2 - 1}{se(\hat{\beta}_1 + \hat{\beta}_2)}$$

- Once we have the *t* statistic, testing proceeds as before. We choose a significant level for the test, α, and, based on the *df*, obtain a critical value, *c*.

Testing a Linear Combination

- Or, we compute the *t* statistic and then compute the *p*-value of the test.
- The procedure is the same if H_1 is a one sided alternative, H_1: $\beta_1 + \beta_2 < 1$ or H_1: $\beta_1 + \beta_2 > 1$.
- The only tedious part in obtaining $t_{\hat{\beta}_1+\hat{\beta}_2}$ is $se(\hat{\beta}_1 + \hat{\beta}_2)$, since you cannot compute it from the individual standard errors of the estimates, that is the information you get from the regression output.

Testing a Linear Combination

- In fact, to compute $se(\hat{\beta}_1 + \hat{\beta}_2)$ you need information on the estimated covariance, since

$$Var(\hat{\beta}_1 + \hat{\beta}_2) = Var(\hat{\beta}_1) + Var(\hat{\beta}_2) + 2Cov(\hat{\beta}_1, \hat{\beta}_2)$$

- Hence

$$se(\hat{\beta}_1 + \hat{\beta}_2) = \sqrt{\left[se(\hat{\beta}_1)\right]^2 + \left[se(\hat{\beta}_2)\right]^2 + 2.s_{12}}$$

where s_{12} denotes an estimate of $Cov(\hat{\beta}_1, \hat{\beta}_2)$

Testing a Linear Combination

- Many econometric software packages have an option to display estimates of the covariance terms like $Cov(\hat{\beta}_1, \hat{\beta}_2)$.
- More generally, we can always restate the problem to get the test we want.
- So in practice is usually much easier to estimate a different model that directly delivers the standard error of interest.
- Lets see our previous example.

Testing a Linear Combination

- Define $\theta = \beta_1 + \beta_2 - 1$, so H_0: $\theta = 0$.

  From this $\beta_1 = \theta - \beta_2 + 1$, so substitute β_1 in the original equation

  $$\log(y) = \beta_0 + (\theta - \beta_2 + 1)\log(l) + \beta_2 \log(k) + u$$

  Hence

  $$\log(y/l) = \beta_0 + \theta \log(l) + \beta_2 \log(k/l) + u$$

  Regress $\log(y/l)$ on a constant, $\log(l)$ and $\log(k/l)$;

  and get the t statistic $t_{\hat{\theta}} = \dfrac{\hat{\theta}}{se(\hat{\theta})}$ from the regression output.

Testing a Linear Combination

- The strategy of rewriting the model, so that it contains the parameter of interest, works in all cases and is usually easy to implement.
- Other examples of hypotheses about a single linear combination of parameters are:

  $\beta_1 = \beta_2$; $\beta_1 = -(1/2)\beta_2$; $\beta_1 = 1 + \beta_2$; $\beta_1 = 5\beta_2$;...

Multiple Linear Restrictions

- So far, we have only considered hypothesis involving a *single* restriction. But frequently, we wish to test *multiple* hypothesis about the underlying parameters $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$.
- We begin with the leading case of testing whether a set of independent variables has no partial effect on the dependent variable, y.
- These are called **exclusion restrictions**.

Testing Exclusion Restrictions

- **Example**: Consider the model

  $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + u$$

  A typical example of **exclusion restrictions** is

  $$H_0: \beta_3 = \beta_4 = \beta_5 = 0$$

  This is an example of a set of **multiple restrictions**, because we are putting more than one restriction on the parameters in the above equation.

Testing Exclusion Restrictions

- A test of multiple restrictions is called a **multiple hypothesis test** or a **joint hypothesis test**.
- As before we need:

1. H_1, either explicitly or implicitly.
2. A significance level, α.
3. A statistic whose distribution is known under H_0.
4. A critical value, *c*, which determines the rejection region.

Testing Exclusion Restrictions

- What should be the alternative?

$$H_1: H_0 \text{ is } \textbf{not} \text{ true}$$

- The test we study now is constructed to detect any violation of H_0.
- It is also valid when the alternative is something like

$$H_1: \beta_3 > 0, \beta_4 > 0, \beta_5 > 0$$

but it will not be the best possible test under such alternatives.

Testing Exclusion Restrictions

- We do not have the statistical background necessary to cover tests that have more power under multiple one-sided alternatives.
- To test H_0 it is temping to use the individual t statistics on x_3, x_4 and x_5.
- This option is not appropriate.
- A particular t statistic tests a hypothesis that puts no restriction on the other parameters.
- We need a way to test exclusion restrictions *jointly*.

Testing Exclusion Restrictions

- It turns out that the sum of squared residuals, SSR, provide us with a very convenient basis for testing multiple hypothesis.
- Before we go into the details of the statistic to use, we need two more concepts in relation to H_0.

1. **Unrestricted model**: The model we begin with.
2. **Restricted model**: The model under H_0.

Testing Exclusion Restrictions

- In the above example the **restricted model** is

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

- By definition the restricted model always have less parameters than the unrestricted one.
- The restricted model is obtained by imposing H_0 on the original model.

Testing Exclusion Restrictions

- Moreover it is always true that

$$SSR_r \geq SSR_{ur}$$

where SSR_r is the SSR of the restricted model,

and SSR_{ur} is the SSR of the unrestricted model.

Testing Exclusion Restrictions

- To see this, note that imposing restrictions on a model cannot lower the SSR.
- Remember that, because OLS estimates are chosen to minimize the sum of squared residuals, the SSR never decreases (and generally increases) when some restrictions (like dropping variables) are introduced into the model.
- This is an algebraic fact.

Testing Exclusion Restrictions

- Hence, even if the SSR itself tells us nothing about the truth of H_0. The increase in the SSR when the restrictions are imposed can tell us something about the likely truth of H_0.
- If we get a large increase, this is evidence against H_0, and this hypothesis will be rejected.
- If the increase is small, this is not evidence against H_0, and this hypothesis will not be rejected.

Testing Exclusion Restrictions

- The question is then whether the observed increased in the SSR when the restrictions are imposed is large enough, relative to the SSR in the unrestricted model, to warrant rejecting H_0.
- In others words, what we need to decide is whether the increased in the SSR in going from the unrestricted model to the restricted model is large enough to warrant rejection of H_0.

Testing Exclusion Restrictions

- As with all testing, the answer depends on α.
- But we cannot carry out the test at a chosen α until we have a statistic whose distribution is known, and can be tabulated, under H_0.
- Thus, we need a way to combine the information in SSR_r and SSR_{ur} to obtain a test statistic with a known distribution under H_0.
- Lets see the general case.

Testing Exclusion Restrictions

- **Unrestricted model**:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

- We have q **exclusion restrictions** to test, that is, H_0 states that q of the variables have zero coefficients.
- Assuming that they are the last q variables, H_0 is stated as

$$H_0\colon \beta_{k-q+1} = \beta_{k-q+2} = \ldots = \beta_k = 0$$

Testing Exclusion Restrictions

- **Restricted model**:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_{k-q} x_{k-q} + u$$

- Obtained by imposing H_0 on the unrestricted model.
- H_1 is stated as

$$H_1\text{: } H_0 \text{ is } \textbf{not} \text{ true}$$

Testing Exclusion Restrictions

- The ***F* statistic**, or *F ratio*, is defined by

$$F = \frac{(\mathrm{SSR}_r - \mathrm{SSR}_{ur})/q}{\mathrm{SSR}_{ur}/(n-k-1)}$$

where SSR_r is the SSR of the restricted model, and SSR_{ur} is the SSR of the unrestricted model.

- Note that

$$\mathrm{SSR}_r \geq \mathrm{SSR}_{ur} \quad \Rightarrow \quad F \geq 0$$

Testing Exclusion Restrictions

- The easiest way to remember where the SSR´s appear is to think of F as measuring the relative increase in SSR when moving from the unrestricted to the restricted model.
- The difference in SSR´s in the numerator of F is divided by q, which is the number of restrictions imposed in moving from the unrestricted to the restricted model.

Testing Exclusion Restrictions

- Note that we can write

q = **numerator degrees of freedom** = $df_r - df_{ur}$

so q is the difference in the *df* between the restricted and unrestricted model, $df_r > df_{ur}$.

- The SSR in the denominator of F is divided by df_{ur}.

$n-k-1$ = **denominator degrees of freedom** = df_{ur}

Testing Exclusion Restrictions

- In fact, the denominator of F is just the unbiased estimator of $\sigma^2 = \mathrm{Var}(u)$ in the unrestricted model.

$$\hat{\sigma}^2 = \frac{\mathrm{SSR}_{ur}}{n-k-1}$$

- In order to use the F statistic for hypothesis testing, we must know its sampling distribution under H_0 in order to choose c for a given α, and determine the rejection rule.

Testing Exclusion Restrictions

- It can be shown that, under H_0, and assuming the CLM assumptions hold, the F statistic is distributed as an F random variable with $(q\ ,\, n-k-1)$ *df*.
- We write this result as

$$F \sim F_{q,n-k-1} \quad \text{on } H_0$$

The $F_{q,n-k-1}$ distribution is readily tabulated and available in statistical tables.

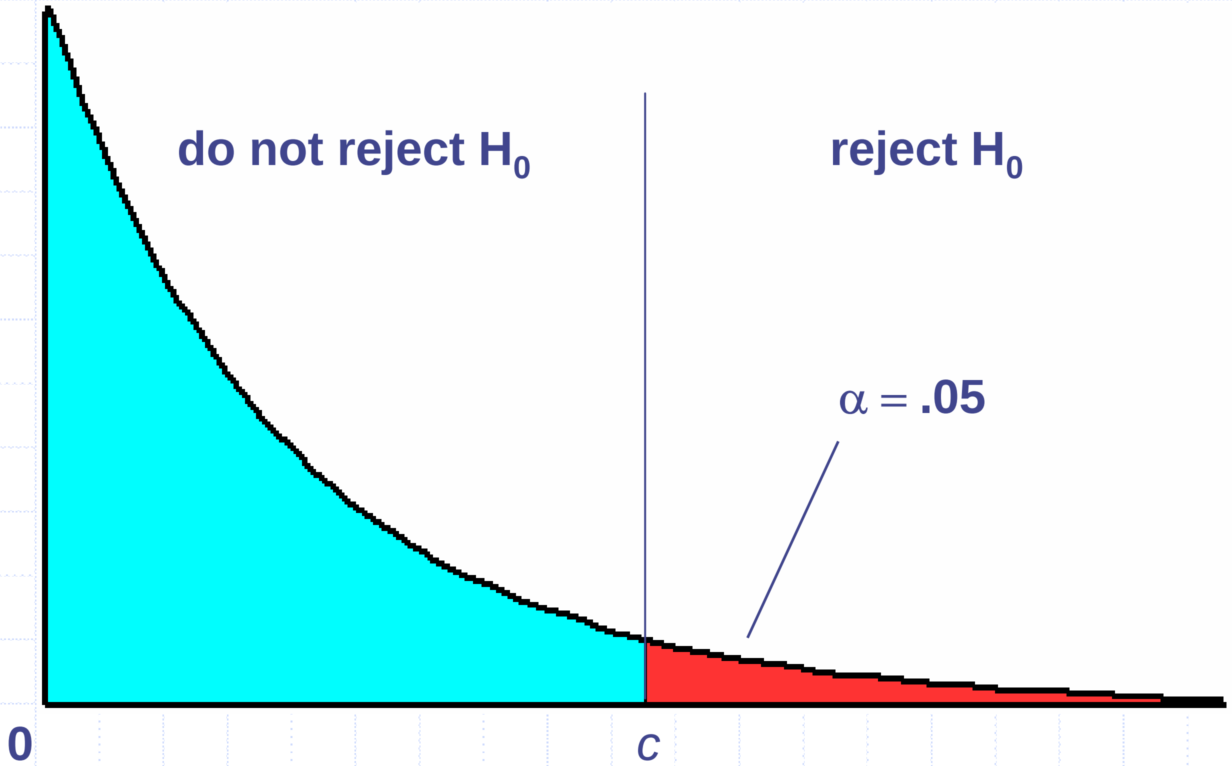
Testing Exclusion Restrictions

- It is pretty clear from the definition of F that we will reject H_0 in favor of H_1 when F is sufficiently "large".
- As usual, how large depends on α.
- For $\alpha = .05$, let c be the 95th percentile in the $F_{q,n-k-1}$ distribution. This critical value depends on q, the numerator *df*, and on $n - k - 1$, the denominator *df*.

Testing Exclusion Restrictions

- The rejection rule is quite simple.
- **Rejection rule**: Reject H_0 in favor of H_1 at the given α if $F > c$, where c is the corresponding percentile in the $F_{q,n-k-1}$ distribution.
- In general q will be notably smaller than $n - k - 1$.

The F statistic



Example of a rejection rule: Reject H_0 at $\alpha = .05$ if $F > c$.

Testing Exclusion Restrictions

- If H_0 is rejected, then we say that $x_{k-q+1}, x_{k-q+2}, \ldots, x_k$ are **jointly statistically significant**, or just *jointly significant*, at the appropriate significance level.
- This tests alone does not allow us to say which of the variables has a partial effect on *y*; they may all affect *y* or maybe only one affects *y*.

Testing Exclusion Restrictions

- If H_0 is not rejected, then we say that x_{k-q+1}, $x_{k-q+2}, \ldots, x_k$ are **jointly statistically insignificant**, or just *jointly insignificant*, which often justifies dropping them from the model.

- The F statistic is often useful for testing exclusion of a group of variables when the variables in the group are highly correlated.

Relation between *F* and *t* Statistics

- We have just seen how to use the *F* statistic to test whether a group of variables should be included in the model.
- What happens if we apply the *F* statistic to the case of testing significance of a *single* independent variable?
- This is, when $q = 1$ and H_0: $\beta_k = 0$?
- We know that the *t* statistic on β_k can be used to test this hypothesis.

Relation between *F* and *t* Statistics

- Do we have two different ways of testing the same H_0: $\beta_k = 0$?
- The answer is **no**.
- It can be shown that the *F* statistic for testing H_0: $\beta_k = 0$, is just equal to the square of the corresponding *t* statistic.
- Hence, the two approaches lead to exactly the same outcome, provided that the alternative is two-sided.

Relation between *F* and *t* Statistics

- It can be shown that $t^2_{n-k-1} \equiv F_{1,n-k-1}$.
- But, the *t* statistic is more flexible for testing a single hypothesis, because it can be used to test against one-sided alternatives.
- Moreover, since the *t* statistics are also easier to obtain than the *F* statistics, there is really no good reason to use an *F* statistic to test a single hypothesis. Use a *t* test instead.

Relation between *F* and *t* Statistics

- **Remember**: The *F* statistic is intended to detect whether any combination of a set of coefficients is jointly different from zero, but it is never the best test for determining whether a single coefficient is different from zero.
- Hence, if we fail to reject H_0 there is always the possibility that a single variable will be significant.
- The *t* test is best suited for testing a single hypothesis.

The *R*-Squared form of the *F* Statistic

- It often convenient to have a form of the *F* statistic that can be computed from the R^2 of the restricted and unrestricted models.
- Using the fact that $SSR_r = SST.(1 - R_r^2)$ and $SSR_{ur} = SST.(1 - R_{ur}^2)$, we can write the *F* as

$$F = \frac{\left(R_{ur}^2 - R_r^2\right)/q}{\left(1 - R_{ur}^2\right)/(n - k - 1)}$$

since the SST term cancels.

The *R*-Squared form of the *F* Statistic

- This is called the ***R*-squared form of the *F* statistic**.

- **Warning**: Whereas the *R*-squared form of the *F* statistic is very convenient for testing exclusion restrictions, it cannot be applied for testing all kind of linear restrictions.

  More on this later on.

Computing p-values for F Tests

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

$$p\text{-value} = \Pr(\mathfrak{F} > F)$$

where $\mathfrak{F}$ denotes an F random variable with $(q\ ,\, n - k - 1)$ *df*, and F is the actual value of the test statistic.

Computing *p*-values for *F* Tests

- The *p*-value still have the same interpretation as it did for *t* statistics: It is the probability of observing a value of *F* at least as large as we did, *given* that the null hypothesis is true.
- A small *p*-value is evidence against H_0.
- A large *p*-value is not evidence against H_0.
- As with *t* testing, once the *p*-value has been computed, the *F* test can be carried out at any significance level.

Overall significance

- A special case of exclusion restrictions is

$$H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0$$

- Which could be alternatively written as

$$H_0: x_1, x_2, \ldots, x_k \text{ do not help to explain } y$$

- This null states that *none* of the explanatory variables has an effect on y.
- Another useful way of stating the null is

$$H_0: E(y \mid x_1, x_2, \ldots, x_k) = E(y)$$

Overall significance

- And H_1 is that at least one of the β_j is different from zero,

$$H_1: \beta_j \neq 0 \quad \text{for some } j$$

- The **restricted model** is

$$y = \beta_0 + u$$

- For this model $\hat{\beta}_0 = \bar{y}$, $\hat{u}_i = y_i - \bar{y}$ and $R^2 = 0$

Overall significance

- Therefore, the F statistic for testing H_0 can be written as

$$F = \frac{R^2/k}{\left(1-R^2\right)/(n-k-1)}$$

where R^2 is just the usual R-squared from the regression of y on $x_1, x_2, \ldots, x_k$.

Overall significance

- **Warning**: This special form of the *F* statistic is valid only for testing joint exclusion restriction of *all* independent variables, excluded the intercept.

- This is called testing the **overall significance of the regression**, and it is usually computed by regression software after OLS estimation.

Testing General Linear Restrictions

- Sometimes we are interested in testing multiple joint restrictions, which not all are of the exclusion form.
- For example, in a model with $k = 5$,

$$H_0\text{: } \beta_1 = \beta_2 \text{ , } \beta_3 = 1, \beta_4 = 0, \beta_5 = 0$$

- The important thing we should remember is that the SSR form of the F test can always be applied in these situations.

Testing General Linear Restrictions

- All we need is the SSR of the restricted and unrestricted models.
- In order to get SSR_r we have to impose the restrictions on the model, to get the restricted model.
- Note that this can involve redefining some variables, before the restricted model can be estimated.

Testing General Linear Restrictions

- In the previous example the restricted model is

$$y = \beta_0 + \beta_1 x_1 + \beta_1 x_2 + x_3 + u$$

but before we can estimate it, we should write the model as

$$y - x_3 = \beta_0 + \beta_1 (x_1 + x_2) + u$$

So to get SSR_r, we regress $y - x_3$ on $x_1 + x_2$.

Testing General Linear Restrictions

- Once the restricted and unrestricted model have been estimated, the *F* statistic is computed in the usual way, using SSR from both models.

Testing General Linear Restrictions

- **Warning**: We cannot use the R-squared form of the F statistic in this example because the dependent variable in the restricted and the unrestricted model is different. This means that the SST are different in the two regressions and both formulas are no longer equivalent.

  As a general rule, the SSR form of the F statistic should be used if a different dependent variable is needed in running the restricted regression.

Reporting Regression Results

- Some guidelines on how to report multiple regression results:

1. Estimated OLS coefficients should always be reported.
2. For the key variables in an analysis, you should *interpret* the estimated coefficients.

   This often requires knowing the units of measurement of the variables.

Reporting Regression Results

3. The **economic** or practical importance of the estimates of the key variables should be discussed.
4. The standard errors should always be included along with the estimated coefficients.

   Better standard errors than *t* statistics.
5. The *R*-squared from the regression should always be included.

Reporting Regression Results

6. Reporting the overall significant F test, and its p-value, is good practice.
7. Reporting SSR and the standard error of the regression is good practice, but it is not crucial.
8. The number of observations, n, should be reported.
9. Reporting can be done in equation form, or in table form for many equations at a time.

Reporting Regression Results

10. This list will be updated as more material is covered.

    In particular when we study misspecification tests.