Lecture 3: Interval Estimation and Hypothesis Testing

Shuo Liu

UCLA Summer School Econ 103

July 3, 2017

Outline

Interval Estimation

Assumed Model Form:

$$y = \beta_1 + \beta_2 x + e \tag{1}$$

Fitted Model:

$$\hat{y} = b_1 + b_2 x \tag{2}$$

- b_1 and b_2 are **point estimates** of β_1 and β_2 by OLS method;
- But we need estimate with "more information", for example, what is the probability that the estimate(s) can "cover" the true population parameter(s)?
- So we need **interval estimate**: proposes a range (interval) of values in which the true population parameter is likely to fall **with certain probability**. We usually consider 95% and 90% confidence intervals.

Interval Estimation

The confidence interval must come from some distribution:

(1) Distribution of **OLS Estimators**$\Longrightarrow$(2) Distribution of **Transformed Variable**$\Longrightarrow$(3) Interval of **Transformed Variable**$\Longrightarrow$(4) Confidence Interval of **Population Parameter**.

1. When σ^2 is known:

- By properties of OLS estimator b_2:

$$b_2 \sim N\left(\beta_2, \sigma^2 \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right) \tag{3}$$

- **Transformed Variable** is obtained by transforming general normal to standard normal:

$$Z = \frac{b_2 - \beta_2}{\sqrt{\sigma^2 \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}} \sim N(0,1) \tag{4}$$

Interval Estimation

- If we want 95% confidence interval of β_2, we firstly need to find a "95% interval" (symmetric w.t. $E(Z)$) of Z:

$$P(-1.96 < Z < 1.96) = 0.95 \tag{5}$$

$$\Longleftrightarrow$$

$$P\left(-1.96 < \frac{b_2 - \beta_2}{\sqrt{\sigma^2 \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}} < 1.96\right) = 0.95 \tag{6}$$

$$\Longleftrightarrow$$

$$P\left(b_2 - 1.96\sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} < \beta_2 < b_2 + 1.96\sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}\right) = 0.95 \tag{7}$$

- **The 95% confidence interval of β_2 is:**

$$\left[b_2 - 1.96\sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}, b_2 + 1.96\sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}\right] \tag{8}$$

Interval Estimation

- Exercise: get 90%, 85% confidence intervals of β_1 and β_2;
- For β_1, we should start from the distribution of OLS estimator of β_1:

$$b_1 \sim N\left(\beta_1, \sigma^2 \frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}\right) \tag{9}$$

How to interpret a 95% confidence interval of β_2?

- Given a specific set of sample data, there is 95% probability that the true parameter β_2 falls in this confidence interval;
- If we do repeated sampling (obtain "a large number" of sets of sample data, thus "a large number" of such confidence intervals), 95% of these intervals (obtained by the same method) will "cover" the true parameter β_2.

2. When σ^2 is unknown:
We need to use $\hat{\sigma}^2$ to replace σ^2:

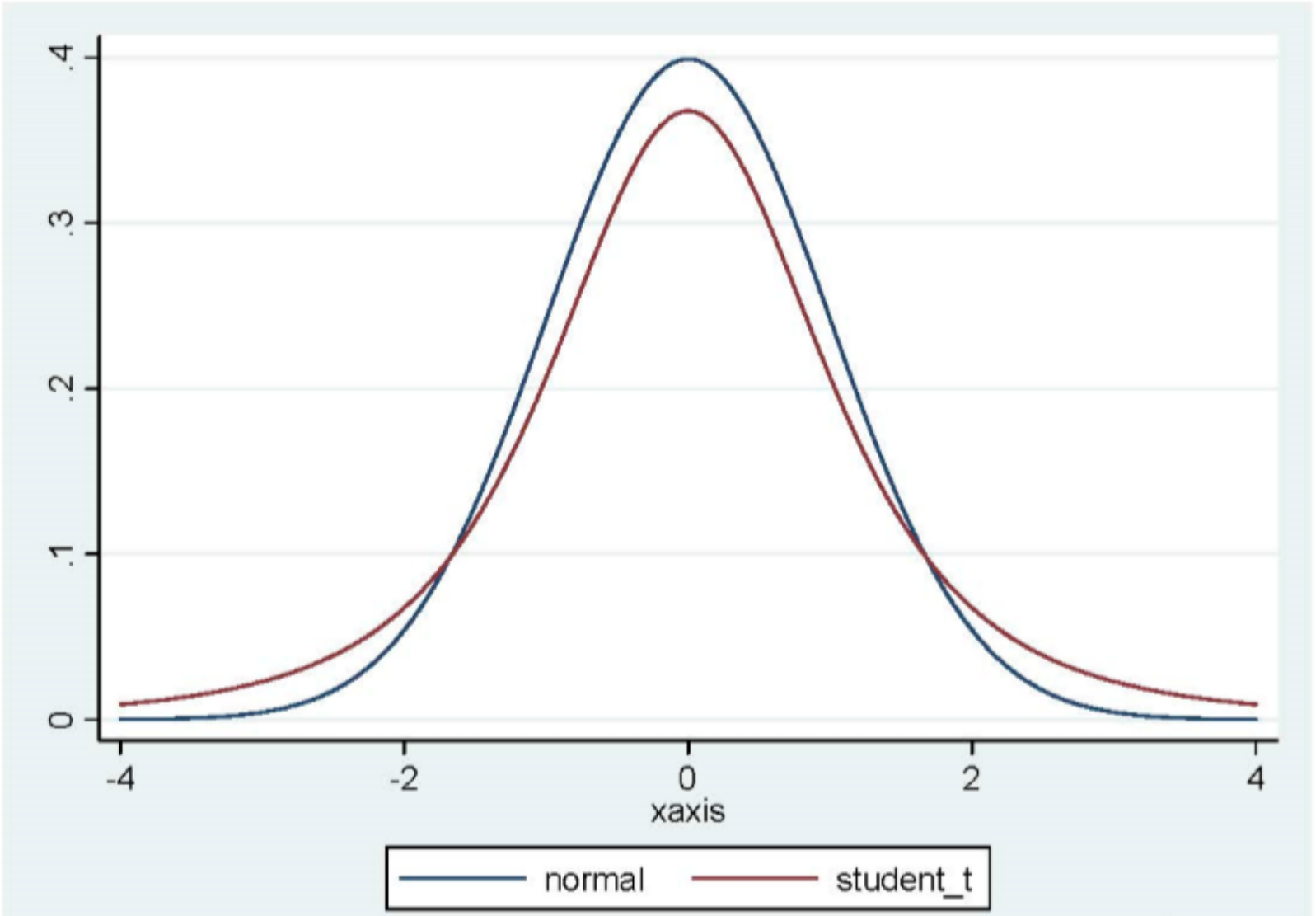
$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{e}_i^2 \tag{10}$$

- Directly start from distribution of **transformed variable** of b_2 using $\hat{\sigma}^2$ (remember the distribution of b_2 does not change):

$$t = \frac{b_2 - \beta_2}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}} = \frac{b_2 - \beta_2}{\sqrt{\widehat{Var}(b_2)}} \sim t_{n-2} \tag{11}$$

- If we want 95% confidence interval of β_2, we firstly need to find a "95% interval" (symmetric w.t. $E(t)$) of t:

$$P(-t_{(0.975,n-2)} < t < t_{(0.975,n-2)}) = 0.95 \tag{12}$$


.4
.3
.2
.1
0
-4
-2
0
2
4
xaxis
normal
student_t

Interval Estimation

$$P(-t_{(0.975,n-2)} < t < t_{(0.975,n-2)}) = 0.95 \tag{13}$$

$$\Longleftrightarrow$$

$$P\left(-t_{(0.975,n-2)} < \frac{b_2 - \beta_2}{\sqrt{\widehat{Var}(b_2)}} < t_{(0.975,n-2)}\right) = 0.95 \tag{14}$$

$$\Longleftrightarrow$$

$$P\left(b_2 - t_{(0.975,n-2)}\sqrt{\widehat{Var}(b_2)} < \beta_2 < b_2 + t_{(0.975,n-2)}\sqrt{\widehat{Var}(b_2)}\right) = 0.95 \tag{15}$$

- **The 95% confidence interval of β_2 is:**

$$\left[b_2 - t_{(0.975,n-2)}\sqrt{\widehat{Var}(b_2)}, b_2 + t_{(0.975,n-2)}\sqrt{\widehat{Var}(b_2)}\right] \tag{16}$$

- $t_{(0.975,n-2)}$ is an example of critical value of t_{n-2} to construct 95% confidence interval.

Interval Estimation

For $t_c = t_{(0.975,n-2)}$: $m = n-2$, $\frac{\alpha}{2} = 1 - 0.975 = 0.025$, so $\alpha = 0.05 = 1 - 0.95$.

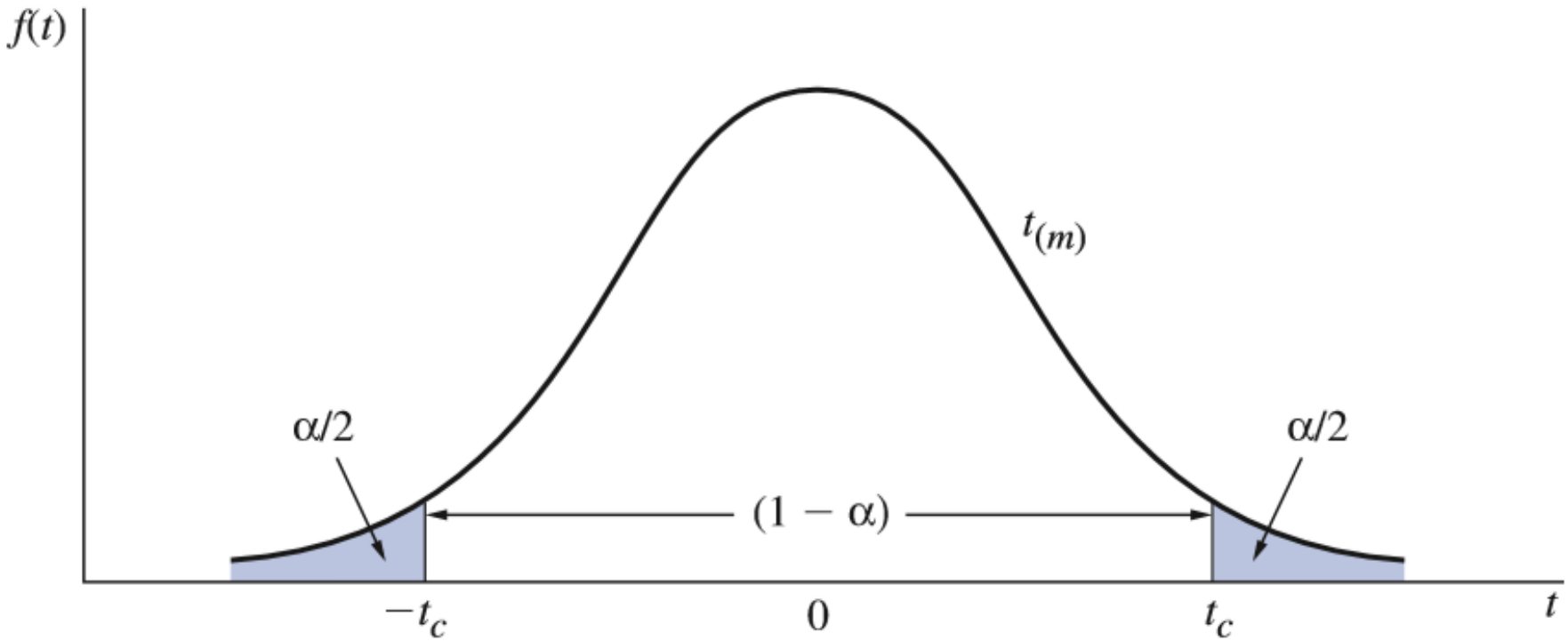


FIGURE 3.1 Critical values from a t-distribution.

Interval Estimation

More generally, if we want to construct X ($X = 0.99, 0.95, 0.9, 0.85 \cdots$) confidence interval of b_2: (do not know σ^2 but $\hat{\sigma}^2$)

- Firstly determine the shape (degree of freedom) of t distribution: $m = n-2$;
- Secondly determine α: $1 - \alpha = X \rightarrow \alpha = 1 - X$, then determine $\frac{\alpha}{2}$;
- Then determine the critical value $t_c = t_{(1-\frac{\alpha}{2}, n-2)}$ from the t table;
- Then construct corresponding X interval of transformed variable t:

$$P(-t_c < t < t_c) = X \tag{17}$$

- Through transformation, get final X confidence interval:

$$\left[b_2 - t_c \sqrt{\widehat{Var}(b_2)}, b_2 + t_c \sqrt{\widehat{Var}(b_2)} \right] \tag{18}$$

Question: if we fix the model form to be simple linear regression model and we fix the sample data we use, then which part in the confidence interval is fixed? which part depends on X through which parameter? How about the confidence interval of β_1?

Interval Estimation

Equivalently, if given α (usually $\alpha = 0.01, 0.05$), we can construct $100(1-\alpha)\%$ confidence intervals for β_k $(k = 1, 2)$:

$$\left[b_k - t_c\sqrt{\widehat{Var}(b_k)}, b_k + t_c\sqrt{\widehat{Var}(b_k)}\right] \tag{19}$$

Comment:

- We will never know whether the true parameter β_k $(k = 1, 2)$ is "definitely/for sure" inside the confidence interval or not;
- We can only say, if we use the above way to construct confidence interval, it will "work" $100(1-\alpha)\%$ times;
- Intuitively, interval estimate contains both "point estimate and standard error of point estimate", so interval estimate provides more information thus being more reliable than point estimate.

Hypothesis Testing

Test the **statements about model parameters** to evaluate whether our assumed model form is good or not:

- **Statements about model parameters**:
  1. each parameter separately: e.g. $\beta_k = 0$, $\beta_k = c$ and etc;
  2. linear combination of part or all of the parameters: e.g. $\lambda_1\beta_1 + \lambda_2\beta_2 = 0$, $\lambda_1\beta_1 + \lambda_2\beta_2 = c$ and etc;
  3. all parameters together: e.g. $\beta_1 = \beta_2 = 0$, in other words, whether model is good in the sense that "not all population parameters are equal to zero at the same time".
- **Evidence used to test the statements**:

  Sample data $\longrightarrow$ point estimates and standard errors $\longrightarrow$ test statistics

Hypothesis Testing

Components of Hypothesis Testing:

- A null hypothesis H_0, an alternative hypothesis H_1;
- A test statistic (know the distribution the test statistic follows and its realized value under H_0);
- A rejection region (depends on the distribution the test statistic follows and significance level α).

Comments:

- H_0 is the statement that is "not easy" to be rejected: we believe it is true unless we have strong enough evidence to reject it;
- H_0 and H_1 are complements;
- Rejection region represents "small probability event (defined by α)" happens on test statistic conditional on "H_0 is true".

Hypothesis Testing

Basic idea:

Value of test statistic (assuming "H_0 is true") is inside rejection region

$\Longleftrightarrow$

Small probability event happens if we assume "H_0 is true"

$\Longleftrightarrow$

H_0 is not true (we reject H_0 is true)

Hypothesis Testing

Assumed Model Form:

$$y = \beta_1 + \beta_2 x + e \tag{20}$$

Fitted Model:

$$\hat{y} = b_1 + b_2 x \tag{21}$$

Usually we want to test whether β_k (especially β_2) is significantly different from zero, why?

Null Hypothesis (k is either 1 or 2):

$$H_0 : \beta_k = 0;$$

Why do we usually put $\beta_k = 0$ on H_0? If we reject the null hypothesis when it is true, then we commit a Type 1 error, and $P(\text{Type 1 error}) = \alpha$.

Possible forms of alternative hypothesis:

$$H_1 : \beta_k \neq 0$$

$$H_1 : \beta_k > 0$$

$$H_1 : \beta_k < 0$$

Hypothesis Testing

Generally for $H_0 : \beta_k = c$
Step 1: calculate value of test statistic conditional on "H_0 is true";

- Construct test statistic as:

$$t = \frac{b_k - \beta_k}{\sqrt{\widehat{Var}(b_k)}} \sim t_{n-2} \tag{22}$$

- Evaluate the value of test statistic assuming "H_0 is true":

$$t = \frac{b_k - c}{\sqrt{\widehat{Var}(b_k)}} \tag{23}$$

Note that, if our null hypothesis is $H_0 : \beta_k = 0$, then the value of test statistic is $t = \frac{b_k}{\sqrt{\widehat{Var}(b_k)}}$

Hypothesis Testing

Step 2: construct the rejection region

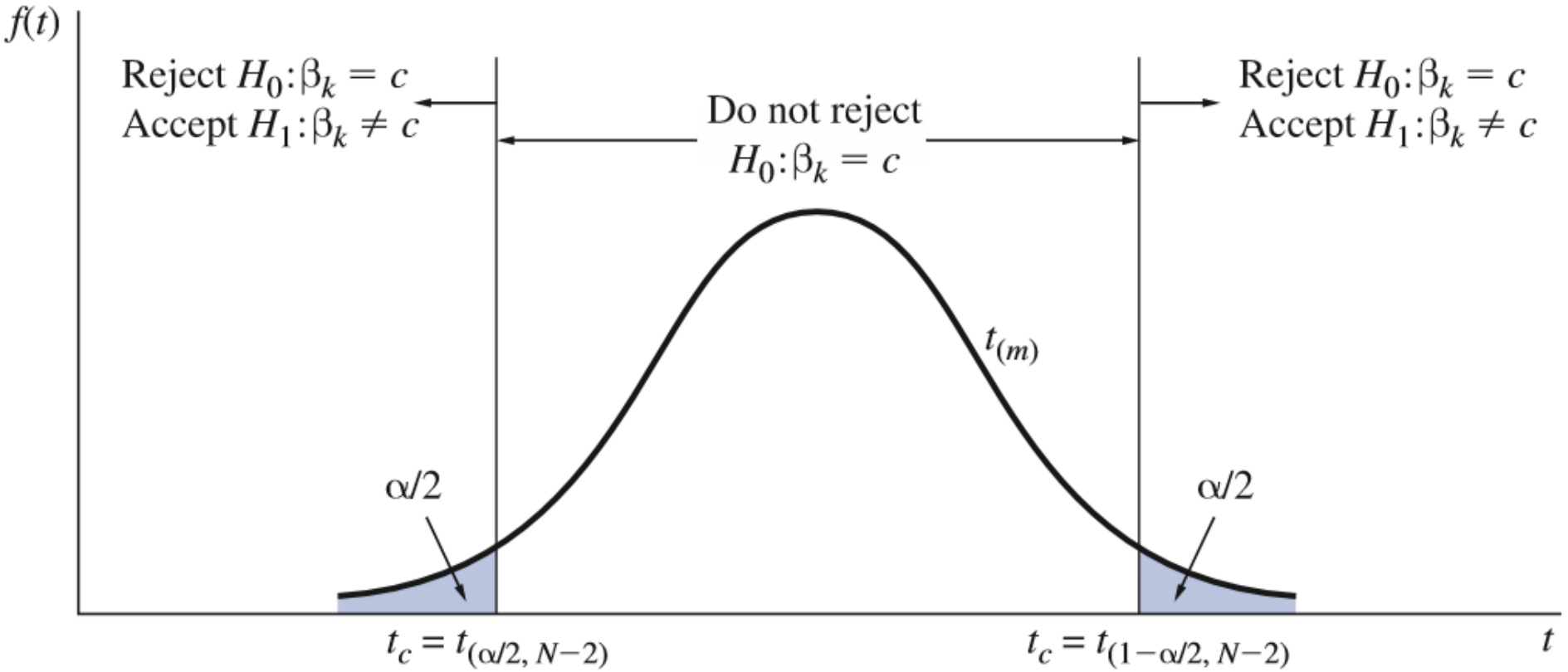
- Distribution of test statistic: t_{n-2};
- A level of significance α: probability that "the value of a random variable following t_{n-2} is inside the rejection region", i.e. the area of rejection region;
- The form of alternative hypothesis H_1:

$\Longrightarrow$ critical value t_c (which is different for different form of H_1)

H_1	**Test form**	**reject H_0 when value of test statistic is**
$\beta_k \neq c$	two-tail test	either too large or too small, $t > t_c$ or $t < -t_c$
$\beta_k > c$	one-tail(right) test	too large, $t > t_c$
$\beta_k < c$	one-tail(left) test	too small, $t < t_c$

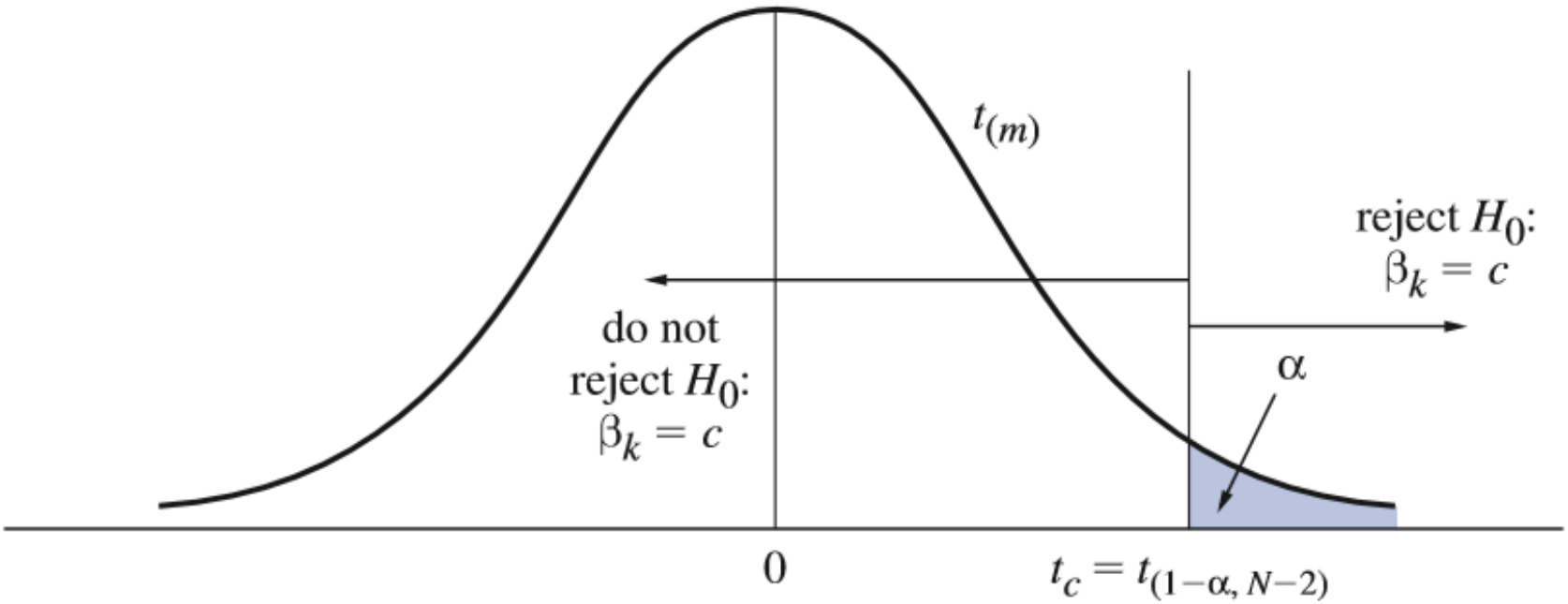
Hypothesis Testing

Rule for testing $H_0 : \beta_k = c$ against $\beta_k \neq c$, $m = n - 2$, significance level is α: reject H_0 (accept H_1) if $t > t_c = t_{(1-\frac{\alpha}{2},n-2)}$ or $t < -t_c = -t_{(1-\frac{\alpha}{2},n-2)} = t_{(\frac{\alpha}{2},n-2)}$



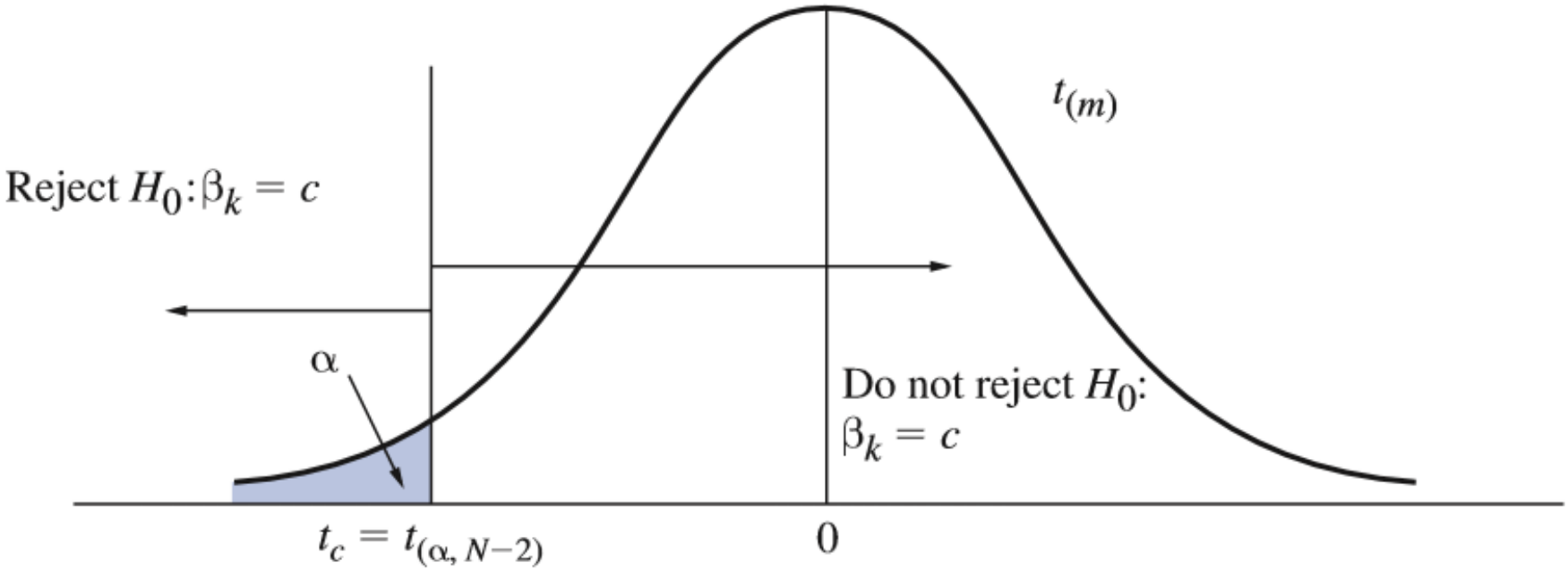
Hypothesis Testing

Rule for testing $H_0 : \beta_k = c$ against $\beta_k > c$, $m = n - 2$, significance level is α: reject H_0 (accept H_1) if $t > t_c = t_{(1-\alpha,n-2)}$



Hypothesis Testing

Rule for testing $H_0 : \beta_k = c$ against $\beta_k < c$, $m = n - 2$, significance level is α: reject H_0 (accept H_1) if $t < t_c = t_{(\alpha, n-2)}$



Hypothesis Testing

Example 1:

$$FOODEXP = \beta_1 + \beta_2 INCOME + e \tag{24}$$

Using sample data $\{FOODEXP_i, INCOME_i\}_{i=1}^{40}$, we estimate $b_2 = 10.21$ with standard error $\hat{Se}(b_2) = \sqrt{\hat{Var}(b_2)} = 2.09$, and do the hypothesis testing with selected $\alpha = 0.05$:

$$H_0 : \beta_2 = 0;$$

$$H_1 : \beta_2 > 0.$$

$$t = \frac{b_2 - 0}{\hat{Se}(b_2)} = \frac{10.21}{2.09} = 4.88 \tag{25}$$

and the critical value for right-tail test is

$$t_c = t_{(1-\alpha, n-2)} = t_{(0.95, 38)} = 1.686 \tag{26}$$

Since $t > t_c$, we reject H_0 and accept H_1. That is, we reject the hypothesis that there is no relationship between income and food expenditure, and conclude that there is **statistically significant** positive relationship between household income and food expenditure.

Hypothesis Testing

Example 2: Everything is same as Example 1 except that we do hypothesis testing as

$$H_0 : \beta_2 \leq 0;$$

$$H_1 : \beta_2 > 0.$$

What is the conclusion? Is it same as that of Example 1?

Hypothesis Testing

p-Value: more convenient way to determine whether we reject H_0 (against specific H_1), and totally **equal to** "checking whether the value of test statistic under 'H_0 is true' is inside the rejection region".

1. Definition: the minimal value of α for which we could reject a test.

2. For a certain significance level α and assume the value of test statistic is t under "H_0 is true". We use T to denote a random variable that follows the same distribution with t, i.e. $T \sim t_{n-2}$.

H_1	**p-Value**	**reject H_0 when**
$\beta_k \neq c$	$p = P(T \geq \lvert t \rvert) = P(T \geq t) + P(T \leq -t)$	$p \leq \alpha$
$\beta_k > c$	$p = P(T \geq t)$	$p \leq \alpha$
$\beta_k < c$	$p = P(T \leq t)$	$p \leq \alpha$

3. Simple rule: we just need to compare p-value with significance level α.

Hypothesis Testing

Go back to **Example 1**:

$$p = P(T > t) = P(T > 4.88) \approx 0 \tag{27}$$

where $T \sim t_{38}$

$\Longrightarrow$

$$p < \alpha = 0.05$$

Then we reject H_0.

Hypothesis Testing

Example 3:
Everything is same as Example 1 except that our hypothesis is,

$$H_0 : \beta_2 \leq 5.5;$$

$$H_1 : \beta_2 > 5.5.$$

and

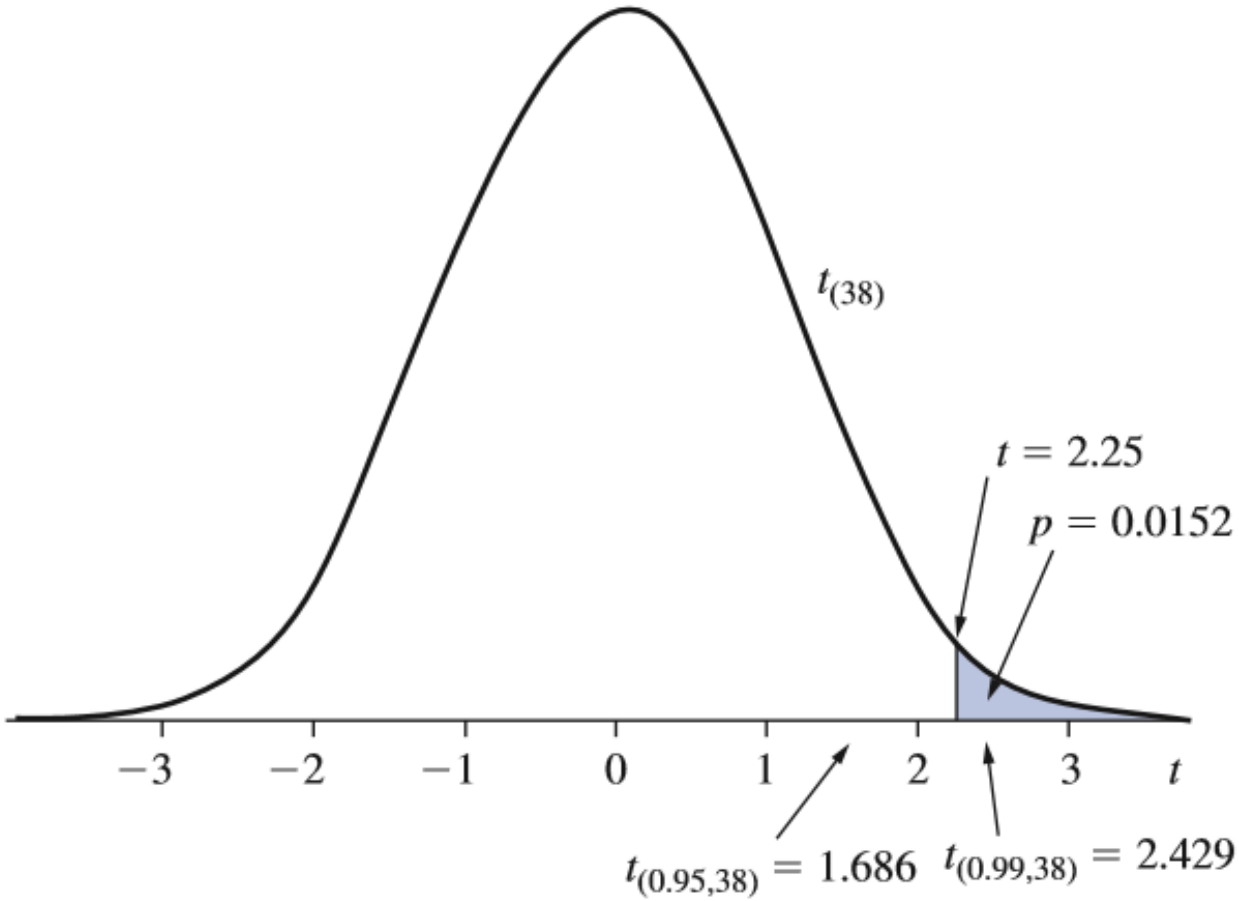
$$t = \frac{b_2 - 5.5}{\hat{S}e(b_2)} = 2.25$$

Then the p-value is,

$$p = P(T > t) = P(T > 2.25) = 1 - P(T \leq 2.25) = 1 - 0.9848 = 0.0152 < \alpha = 0.05 \tag{28}$$

where $T \sim t_{38}$.

Hypothesis Testing



Hypothesis Testing

Example 4: Everything is same as Example 1 except that our hypothesis is,

$$H_0 : \beta_2 = 7.5;$$

$$H_1 : \beta_2 \neq 7.5.$$

and

$$t = \frac{b_2 - 7.5}{\hat{Se}(b_2)} = 1.29$$

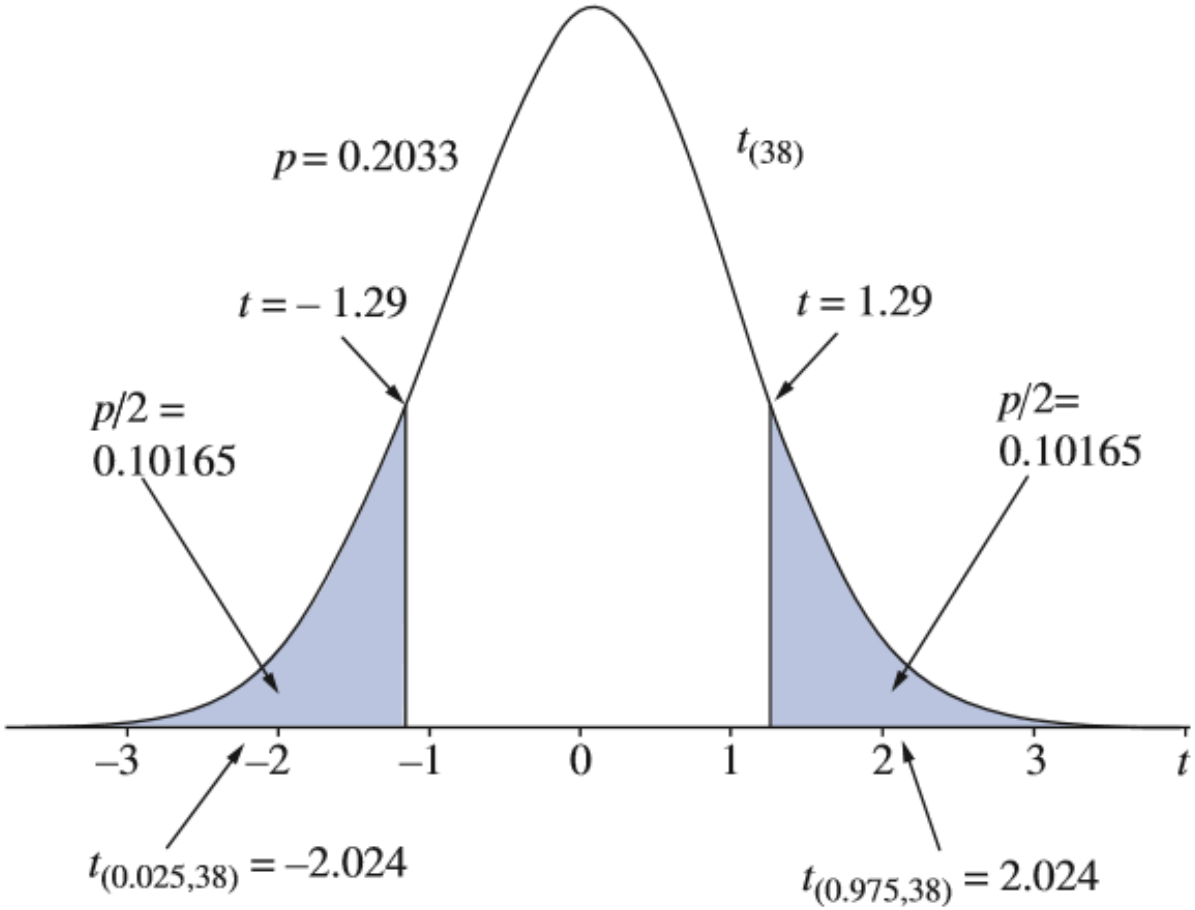
Then the p-value is,

$$p = P(T \geq |t|) = P(T \geq 1.29) + P(T \leq -1.29) = 0.2033 > \alpha = 0.05 \qquad (29)$$

where $T \sim t_{38}$.

Then we **cannot** reject H_0.

Hypothesis Testing



Hypothesis Testing

Linear Combination of Parameters:
We may want to **estimate and test** hypothesis about a linear combination of parameters:

$$\lambda = c_1\beta_1 + c_2\beta_2 \tag{30}$$

where c_1 and c_2 are constants.

- Under assumptions $SR1 - SR5$ and by Gauss Markov Theorem, b_1 and b_2 are best linear unbiased estimators of β_1 and β_2. So have $\hat{\lambda} = c_1 b_1 + c_2 b_2$ is the best linear unbiased estimator of $\lambda = c_1\beta_1 + c_2\beta_2$.
- To do hypothesis test for λ, we not only need to obtain $\hat{\lambda}$ as above, but also need to obtain estimator of $Var(\hat{\lambda})$

$$Var(\hat{\lambda}) = Var(c_1 b_1 + c_2 b_2) = c_1^2 Var(b_1) + c_2^2 Var(b_2) + 2c_1 c_2 Cov(b_1, b_2) \tag{31}$$

$$\Longrightarrow$$

$$\hat{Var}(\hat{\lambda}) = c_1^2 \hat{Var}(b_1) + c_2^2 \hat{Var}(b_2) + 2c_1 c_2 \hat{Cov}(b_1, b_2) \quad \text{by replacing } \sigma^2 \text{ with } \hat{\sigma}^2. \tag{32}$$

Hypothesis Testing

- By $SR6$ (or large enough sample data by Central Limit Theorem), get distribution of $\hat{\lambda}$:

$$\hat{\lambda} = c_1 b_1 + c_2 b_2 \sim N(\lambda, Var(\hat{\lambda})) \tag{33}$$

and if we use $\hat{\sigma}^2$,

$$\frac{\hat{\lambda} - \lambda}{\sqrt{\hat{Var}(\hat{\lambda})}} = \frac{(c_1 b_1 + c_2 b_2) - (c_1 \beta_1 + c_2 \beta_2)}{\sqrt{\hat{Var}(c_1 b_1 + c_2 b_2)}} \sim t_{n-2} \tag{34}$$

- We can construct confidence interval and test hypothesis about the linear combination of true parameters $\lambda = c_1 \beta_1 + c_2 \beta_2$ same as before.

Hypothesis Testing

Example:

Assume estimated model is,

$$\widehat{FOODEXP} = 83.42 + 10.21 INCOME \tag{35}$$

Unit of FOODEXP is \$1 and unit of INCOME is \$100, then $INCOME = 20$ means the household income is \$2000. **Question**: (1) construct 95% confidence interval of "expected food expenditure of household with income as \$2000"; (2) we want to test whether "expected food expenditure of household with income as \$2000" is significantly larger than \$250

$$E(FOODEXP|INCOME = 20) = \beta_1 + 20\beta_2 = \lambda \tag{36}$$

$\Longrightarrow$ We want to construct 95% confidence interval of $\beta_1 + 20\beta_2$ and test:

$$H_0 : \beta_1 + 20\beta_2 \leq 250$$

$$H_1 : \beta_1 + 20\beta_2 > 250$$

Hypothesis Testing

Assume $n = 40$ and estimated variance-covariance matrix of OLS estimators $b_1 = 83.42$ and $b_2 = 10.21$ is:

$$\begin{bmatrix} \widehat{Var}(b_1) & \widehat{Cov}(b_1, b_2) \\ \widehat{Cov}(b_1, b_2) & \widehat{Var}(b_2) \end{bmatrix} = \begin{bmatrix} 1884.442 & -85.9032 \\ -85.9032 & 4.3818 \end{bmatrix}$$

and we have $t_{0.975,38} = 2.024$ and $t_{0.95,38} = 1.686$.

- Then do exercise to construct confidence interval and do hypothesis testing on last page.
- Suggested textbook exercise for this lecture: **3.2, 3.4, 3.8, 3.14**