ECON 351* -- NOTE 7

<u>Interval Estimation in the Classical Normal Linear Regression Model</u>

This note outlines the basic elements of **interval estimation** in the Classical Normal Linear Regression Model (the CNLRM). Interval estimation – i.e., the construction of confidence intervals for unknown population parameters – is one of the two alternative approaches to statistical inference; the other is hypothesis testing.

1. Introduction

- We have previously derived **point estimators** of all the unknown population parameters in the Classical Normal Linear Regression Model (CNLRM) for which the **population regression equation**, or **PRE**, is

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad \text{where } u_i \text{ is iid as } N(0, \sigma^2) \quad (i = 1, \ldots, N) \qquad \textbf{(1)}$$

- The **unknown parameters** of the PRE are

  (1) the **regression coefficients β_0 and β_1**

  and

  (2) the **error variance σ^2**.

- The **point estimators** of these unknown population parameters are

  (1) the ***unbiased*** **OLS regression coefficient estimators $\hat{\beta}_0$ and $\hat{\beta}_1$**

  and

  (2) the ***unbiased*** **error variance estimator $\hat{\sigma}^2$**.

- Assume that we have computed the point estimates $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\sigma}^2$ of the unknown parameters for a given set of sample data (Y_i, X_i), i = 1, ..., N.

- We therefore begin with the following **OLS sample regression equation** (or **OLS-SRE**):

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{u}_i = \hat{Y}_i + \hat{u}_i \qquad (i = 1, \ldots, N) \qquad \textbf{(2)}$$

where

$$\hat{\beta}_1 = \frac{\sum_i x_i y_i}{\sum_i x_i^2} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{N}(X_i - \overline{X})^2} = \text{OLS estimate of } \beta_1;$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} = \text{OLS estimate of } \beta_0;$$

$$\hat{\sigma}^2 = \frac{\sum_i \hat{u}_i^2}{(N-2)} = \frac{RSS}{(N-2)} = \text{unbiased OLS estimate of } \sigma^2;$$

$$V\hat{a}r(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum_i x_i^2} = \frac{\hat{\sigma}^2}{\sum_i (X_i - \overline{X})^2};$$

$$s\hat{e}(\hat{\beta}_1) = \sqrt{V\hat{a}r(\hat{\beta}_1)} = \left(\frac{\hat{\sigma}^2}{\sum_i x_i^2}\right)^{\frac{1}{2}} = \frac{\hat{\sigma}}{\sqrt{\sum_i x_i^2}};$$

$$V\hat{a}r(\hat{\beta}_0) = \frac{\hat{\sigma}^2 \sum_i X_i^2}{N \sum_i x_i^2} = \frac{\hat{\sigma}^2 \sum_i X_i^2}{N \sum_i (X_i - \overline{X})^2};$$

$$s\hat{e}(\hat{\beta}_0) = \sqrt{V\hat{a}r(\hat{\beta}_0)} = \left(\frac{\hat{\sigma}^2 \sum_i X_i^2}{N \sum_i x_i^2}\right)^{\frac{1}{2}}.$$

- Under the assumptions of the Classical Normal Linear Regression Model (CNLRM) – including in particular the ***normality assumption A9*** – the **sample *t-statistics* for $\hat{\beta}_1$ and $\hat{\beta}_0$** each have the **t-distribution with (N – 2) degrees of freedom**: i.e.,

$$t(\hat{\beta}_1) = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{V\hat{a}r(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - \beta_1}{s\hat{e}(\hat{\beta}_1)} \sim t[N-2];$$

$$t(\hat{\beta}_0) = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{V\hat{a}r(\hat{\beta}_0)}} = \frac{\hat{\beta}_0 - \beta_0}{s\hat{e}(\hat{\beta}_0)} \sim t[N-2].$$

2. Interval Estimation: Some Basic Ideas

2.1 General Form of a Confidence Interval

A confidence interval for the slope coefficient β_1 takes the general form

$$\Pr(\hat{\beta}_{1L} \le \beta_1 \le \hat{\beta}_{1U}) = \Pr(\hat{\beta}_1 - \hat{\delta} \le \beta_1 \le \hat{\beta}_1 + \hat{\delta}) = 1 - \alpha \qquad \textbf{(3)}$$

where

α = the ***significance*** **level** ($0 < \alpha < 1$),
$1 - \alpha$ = the ***confidence*** **level** (or confidence coefficient),
$\hat{\delta}$ = a positively-valued sample statistic,
$\hat{\beta}_{1L} = \hat{\beta}_1 - \hat{\delta}$ = the ***lower*** **confidence limit**,
$\hat{\beta}_{1U} = \hat{\beta}_1 + \hat{\delta}$ = the ***upper*** **confidence limit**.

The interval $[\hat{\beta}_{1L}, \hat{\beta}_{1U}] = [\hat{\beta}_1 - \hat{\delta}, \hat{\beta}_1 + \hat{\delta}]$ is called the **two-sided (1 – α)-level confidence interval**, or **two-sided 100(1 – α) percent confidence interval**, for the slope coefficient β_1.

2.2 Interpretation of Confidence Intervals

1. The confidence interval $[\hat{\beta}_{1L}, \hat{\beta}_{1U}]$ is a <u>random</u> interval.

- The **confidence limits** $\hat{\beta}_{1L} = \hat{\beta}_1 - \hat{\delta}$ and $\hat{\beta}_{1U} = \hat{\beta}_1 + \hat{\delta}$ are **random variables** (or **sample statistics**) that vary in value from one sample to another because the values of $\hat{\beta}_1$ and $\hat{\delta}$ vary from sample to sample.

- But for any one sample of data of size N and the corresponding estimates of $\hat{\beta}_1$ and $\hat{\delta}$, the confidence limits $\hat{\beta}_{1L} = \hat{\beta}_1 - \hat{\delta}$ and $\hat{\beta}_{1U} = \hat{\beta}_1 + \hat{\delta}$ are simply fixed numbers, i.e., they take fixed values. Therefore, any one confidence interval calculated for a particular sample of data is a fixed, meaning nonrandom, interval.

2. The **correct interpretation of the confidence interval [$\hat{\beta}_{1L}$, $\hat{\beta}_{1U}$]** is based on the **concept of *repeated sampling***.

- Suppose a very large number of random samples of the same size N (e.g., N = 50 observations) are independently selected from a given population.

- For each of these random samples of N observations, the values of the confidence limits $\hat{\beta}_{1L} = \hat{\beta}_1 - \hat{\delta}$ and $\hat{\beta}_{1U} = \hat{\beta}_1 + \hat{\delta}$ are calculated for some fixed value of the confidence level $1 - \alpha$ (such as $1 - \alpha = 0.99$ or 0.95 or 0.90).

- The probability statement in (3) means that ***100(1 − α) percent*** **of all the confidence intervals so constructed will contain the true (but unknown) population value of β_1**.

- But note that ***any one*** confidence interval **[$\hat{\beta}_{1L}$, $\hat{\beta}_{1U}$]** based on one sample of N observations may or may not contain the true value of β_1.

  ◊ For one sample of N observations, the confidence limits $\hat{\beta}_{1L} = \hat{\beta}_1 - \hat{\delta}$ and $\hat{\beta}_{1U} = \hat{\beta}_1 + \hat{\delta}$ take fixed values because the values of $\hat{\beta}_1$ and $\hat{\delta}$ calculated for a single sample of N observations are fixed numbers.

  ◊ Because β_1 is some fixed but unknown number, β_1 either lies inside or outside the fixed confidence interval calculated for any one sample of N observations. That is, a single confidence interval computed for one specific sample of N observations either does or does not contain the true population value of β_1.

  ◊ ***Result:*** A ***single*** **confidence interval [$\hat{\beta}_{1L}$, $\hat{\beta}_{1U}$]** based on ***one*** **sample** of N observations is a **fixed**, or **nonrandom**, **interval**.

3. Summary: Interpretation of Confidence Intervals

The **two-sided 100(1−α) percent confidence interval** for the slope coefficient β_1 is defined by probability statement (3):

$$\Pr(\hat{\beta}_{1L} \le \beta_1 \le \hat{\beta}_{1U}) = \Pr(\hat{\beta}_1 - \hat{\delta} \le \beta_1 \le \hat{\beta}_1 + \hat{\delta}) = 1 - \alpha \qquad \textbf{(3)}$$

(1) Any one confidence interval for β_1, based on one sample of data, may or may not contain the true value of β_1. Since the true value of β_1 is unknown, we do not know whether that value does or does not lie inside any one confidence interval.

(2) The probability statement (3) is therefore a statement about **the procedure used to construct the confidence interval**, not about any one confidence interval estimate calculated for a particular sample of data.

3. Confidence Intervals for the Regression Coefficients β_0 and β_1

3.1 Confidence Interval for β_1: Derivation

A **two-step** derivation:

Step 1: Start with a probability statement formulated in terms of $t(\hat{\beta}_1)$, the t-statistic for $\hat{\beta}_1$. This probability statement ***implicitly*** **defines** the two-sided $(1-\alpha)$-level confidence interval for β_1.

Step 2: Re-arrange this probability statement to obtain an equivalent probability statement formulated in terms of β_1 rather than $t(\hat{\beta}_1)$. The resultant probability statement ***explicitly*** **defines** the two-sided $(1-\alpha)$-level confidence interval for β_1.

STEP 1: The **two-sided $(1-\alpha)$-level confidence interval for β_1** is implicitly defined by the probability statement

$$\Pr\left(-t_{\alpha/2}[N-2] \le t(\hat{\beta}_1) \le t_{\alpha/2}[N-2]\right) = 1-\alpha \qquad \textbf{(4)}$$

where

$1-\alpha$ = the ***confidence*** **level** attached to the confidence interval;
α = the ***significance*** **level**, where $0 < \alpha < 1$;
$t_{\alpha/2}[N-2]$ = the ***critical value*** of the t-distribution with $(N-2)$ degrees of freedom at the $\alpha/2$ (or $100\alpha/2$ percent) significance level;

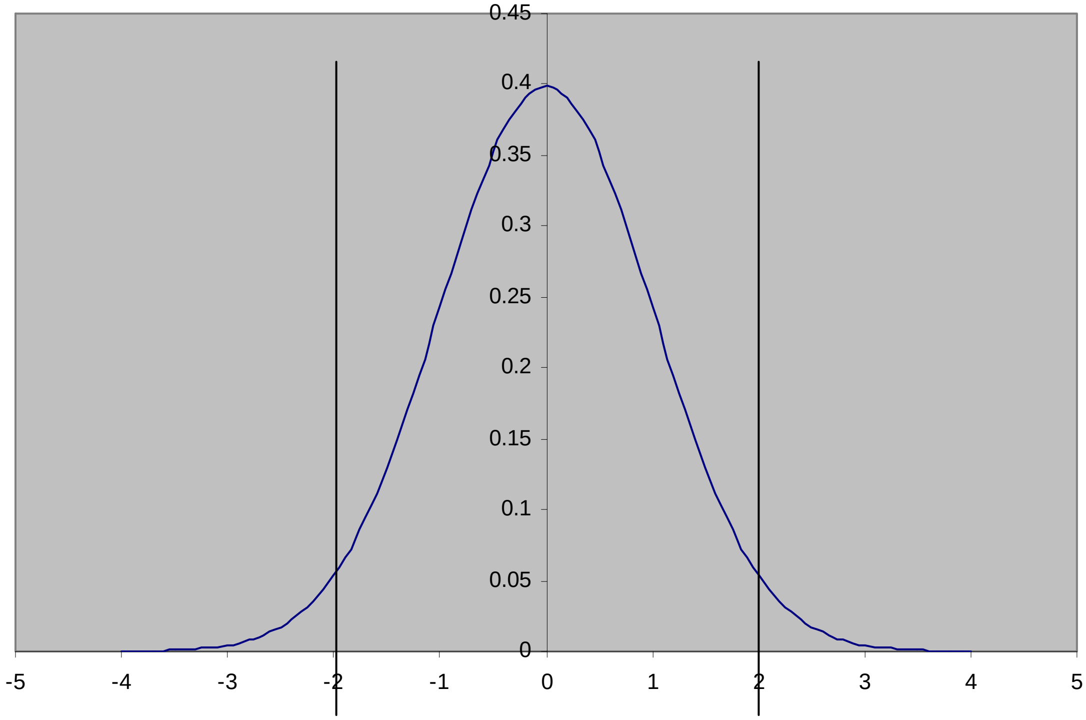
and $t(\hat{\beta}_1)$ is the t-statistic for $\hat{\beta}_1$ given by

$$t(\hat{\beta}_1) = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - \beta_1}{\widehat{\mathrm{se}}(\hat{\beta}_1)}. \qquad \textbf{(5)}$$

Upper and Lower $\alpha/2$ Critical Values of t[N-2] Distribution

left-tail area = $\alpha/2$ ↓ confidence region ↓ right-tail area = $\alpha/2$

area = $1 - \alpha$



$-t_{\alpha/2}$ ↑ $t_{\alpha/2}$ ↑

$\Pr(t < -t_{\alpha/2}) = \alpha/2$ ↑ $\Pr(-t_{\alpha/2} \le t \le t_{\alpha/2}) = 1-\alpha$ ↑ $\Pr(t > t_{\alpha/2}) = \alpha/2$

STEP 2: Express the double inequality inside the brackets in probability statement (4) in terms of β_1 rather than $t(\hat{\beta}_1)$.

$$\Pr\left(-t_{\alpha/2}[N-2] \le t(\hat{\beta}_1) \le t_{\alpha/2}[N-2]\right) = 1-\alpha \qquad \textbf{(4)}$$

(1) Substitute in the double inequality

$$-t_{\alpha/2}[N-2] \le t(\hat{\beta}_1) \le t_{\alpha/2}[N-2]$$

the expression for $t(\hat{\beta}_1)$ given in (5) above:

$$-t_{\alpha/2}[N-2] \le \frac{\hat{\beta}_1 - \beta_1}{\hat{se}(\hat{\beta}_1)} \le t_{\alpha/2}[N-2]. \qquad (6.1)$$

(2) Multiply the double inequality (6.1) by the positive number $\hat{se}(\hat{\beta}_1) > 0$:

$$-t_{\alpha/2}\hat{se}(\hat{\beta}_1) \le \hat{\beta}_1 - \beta_1 \le t_{\alpha/2}\hat{se}(\hat{\beta}_1). \qquad (6.2)$$

(3) Subtract $\hat{\beta}_1$ from both sides of inequality (6.2):

$$-\hat{\beta}_1 - t_{\alpha/2}\hat{se}(\hat{\beta}_1) \le -\beta_1 \le -\hat{\beta}_1 + t_{\alpha/2}\hat{se}(\hat{\beta}_1). \qquad (6.3)$$

(4) Multiply all terms in inequality (6.3) by −1, remembering to reverse the direction of the inequalities:

$$\hat{\beta}_1 - t_{\alpha/2}\hat{se}(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + t_{\alpha/2}\hat{se}(\hat{\beta}_1). \qquad (6.4)$$

RESULT: The probability statement (4) can be written as

$$\Pr\left(\hat{\beta}_1 - t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_1)\right) = 1-\alpha \,. \qquad \textbf{(7)}$$

The **two-sided (1 − α)-level confidence interval for β_1** can therefore be written as

$$\hat{\beta}_1 - t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_1)$$

or more compactly as

$$\hat{\beta}_1 \pm t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_1) \quad \textit{or} \quad [\hat{\beta}_1 - t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_1),\ \hat{\beta}_1 + t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_1)]$$

where at the (1 − α) confidence level, or 100(1 − α) percent confidence level,

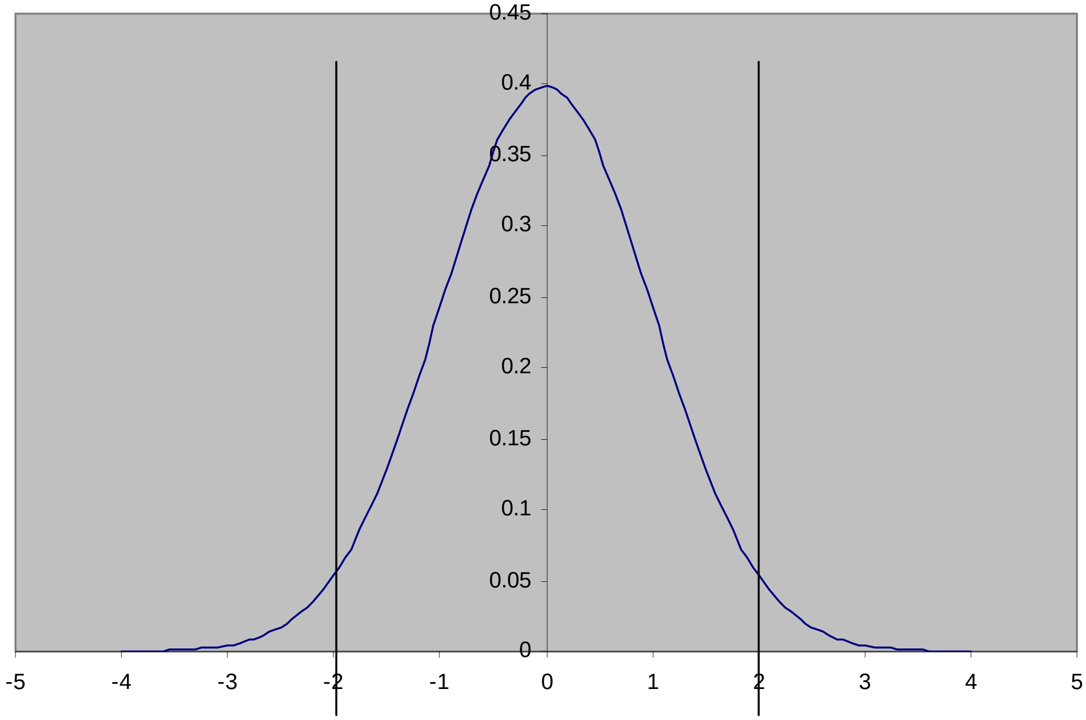
$\hat{\beta}_{1L} = \hat{\beta}_1 - t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_1)$ = the ***lower*** **100(1 − α) percent confidence limit for β_1**

and

$\hat{\beta}_{1U} = \hat{\beta}_1 + t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_1)$ = the ***upper*** **100(1 − α) percent confidence limit for β_1**

Two-Sided (1 – α)-level Confidence Interval for β_1

left-tail area = α/2 ↓ confidence area = 1 – α ↓ right-tail area = α/2



↑ ↑ ↑

$\hat{\beta}_{1L} = \hat{\beta}_1 - t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_1)$ $\hat{\beta}_1$ $\hat{\beta}_{1U} = \hat{\beta}_1 + t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_1)$

3.2 Confidence Interval for β_0: Derivation

The confidence interval (or interval estimator) for the intercept coefficient β_0 is derived, interpreted, and constructed in exactly the same way as the confidence interval for the slope coefficient β_1.

1. The **two-sided (1 – α)-level confidence interval for β_0** is implicitly defined by the probability statement

$$\Pr\left(-\,t_{\alpha/2}[N-2] \le t(\hat{\beta}_0) \le t_{\alpha/2}[N-2]\right) = 1-\alpha \qquad \textbf{(8)}$$

where

$1-\alpha$ = the **confidence level** attached to the confidence interval;
α = the **significance level**, where $0 < \alpha < 1$;
$t_{\alpha/2}[N-2]$ = the **critical value** of the t-distribution with (N–2) degrees of freedom at the α/2 (or 100(α/2) percent) significance level;

and $t(\hat{\beta}_0)$ is the t-statistic for $\hat{\beta}_0$ given by

$$t(\hat{\beta}_0) = \frac{\hat{\beta}_0-\beta_0}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_0)}} = \frac{\hat{\beta}_0-\beta_0}{\widehat{\mathrm{se}}(\hat{\beta}_0)}. \qquad (9)$$

2. The double inequality inside the brackets in probability statement (8) can be expressed in terms of β_0 rather than $t(\hat{\beta}_0)$, using a derivation analogous to that used in deriving the confidence interval for β_1.

<u>RESULT</u>: The probability statement (8) can be written as

$$\Pr\Big(\hat{\beta}_0 - t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_0) \le \beta_0 \le \hat{\beta}_0 + t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_0)\Big) = 1 - \alpha . \qquad \textbf{(10)}$$

The **two-sided (1 − α)-level confidence interval for β_0** can therefore be written as

$$\hat{\beta}_0 - t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_0) \le \beta_0 \le \hat{\beta}_0 + t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_0)$$

or more compactly as

$$\hat{\beta}_0 \pm t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_0) \quad \textit{or} \quad [\hat{\beta}_0 - t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_0),\ \hat{\beta}_0 + t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_0)]$$

where at the (1 − α) confidence level, or 100(1 − α) percent confidence level,

$\hat{\beta}_{0L} = \hat{\beta}_0 - t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_0) =$ the ***lower*** **100(1 − α) percent confidence limit for β_0**

and

$\hat{\beta}_{0U} = \hat{\beta}_0 + t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_0) =$ the ***upper*** **100(1 − α) percent confidence limit for β_0**

3.3 Procedure for Computing Confidence Intervals

Consider the problem of computing a confidence interval for the slope coefficient β_1. Recall that the **two-sided (1 – α)-level confidence interval for β_1** is given by the double inequality

$$\hat{\beta}_1 - t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_1).$$

<u>Step 1</u>: After estimating the PRE (1) by OLS, retrieve from the estimation results the OLS estimate $\hat{\beta}_1$ of β_1 and the estimated standard error $\hat{se}(\hat{\beta}_1)$.

<u>Step 2</u>: Select the value of the confidence level (1 – α), which amounts to selecting the value of α. Although the choice of confidence level is essentially arbitrary, the values most commonly used in practice are:

$\alpha = 0.01 \Rightarrow (1-\alpha) = 0.99$, i.e., the $100(1-\alpha) = 100(0.99) = 99$ percent confidence level;

$\alpha = 0.05 \Rightarrow (1-\alpha) = 0.95$, i.e., the $100(1-\alpha) = 100(0.95) = 95$ percent confidence level;

$\alpha = 0.10 \Rightarrow (1-\alpha) = 0.90$, i.e., the $100(1-\alpha) = 100(0.90) = 90$ percent confidence level.

<u>Step 3</u>: Obtain the value of $t_{\alpha/2}[N-2]$, the α/2 critical value of the t-distribution with N–2 degrees of freedom, either from statistical tables of the t-distribution or from a computer software program.

<u>Step 4</u>: Use the values of $\hat{\beta}_1$, $\hat{se}(\hat{\beta}_1)$, and $t_{\alpha/2}[N-2]$ to compute the upper and lower 100(1 – α) percent confidence limits for β_1:

$\hat{\beta}_{1U} = \hat{\beta}_1 + t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_1) =$ the ***upper* 100(1 – α)% confidence limit for β_1**;

$\hat{\beta}_{1L} = \hat{\beta}_1 - t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_1) =$ the ***lower* 100(1 – α)% confidence limit for β_1**.

4. Determinants of the Confidence Interval for β_j

Consider the **two-sided 100(1 – α)% confidence interval for β_j (j = 0, 1)**:

$$\hat{\beta}_j - t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_j) \le \beta_j \le \hat{\beta}_j + t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_j)$$

or

$$\left[\hat{\beta}_j - t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_j),\ \hat{\beta}_j + t_{\alpha/2}[N-2]\hat{se}(\hat{\beta}_j)\right]$$

By inspection, it is apparent that the **two-sided confidence interval for β_j** is ***wider***

(1) the ***greater*** **the value of $\hat{se}(\hat{\beta}_j)$**, the estimated standard error of $\hat{\beta}_j$, i.e., the ***less*** **precise** is the estimate of $\hat{\beta}_j$;

(2) the ***greater*** **the critical value $t_{\alpha/2}[N-2]$,** i.e., the ***greater*** **the chosen value of the confidence level (1 – α)** for the given sample size N.

Explanation: Given sample size N, **the value of $t_{\alpha/2}[N-2]$ is *negatively* related to the value of α**, and so is ***positively*** **related to the value of (1 – α)**.

Example: Suppose sample size N = 30, so that the degrees-of-freedom N-2 = 28. Then from a table of percentage points for the t-distribution, we obtain the following values of $t_{\alpha/2}[N-2] = t_{\alpha/2}[28]$ for different values of α:

$\alpha = 0.01 \Rightarrow (1-\alpha) = 0.99$: $\alpha/2 = 0.005$ and $t_{0.005}[28] = 2.763$;

$\alpha = 0.02 \Rightarrow (1-\alpha) = 0.98$: $\alpha/2 = 0.01$ and $t_{0.01}[28] = 2.467$;

$\alpha = 0.05 \Rightarrow (1-\alpha) = 0.95$: $\alpha/2 = 0.025$ and $t_{0.025}[28] = 2.048$;

$\alpha = 0.10 \Rightarrow (1-\alpha) = 0.90$: $\alpha/2 = 0.05$ and $t_{0.05}[28] = 1.701$.

Note that higher values of (1 – α) -- i.e., higher confidence levels -- correspond to higher critical values of $t_{\alpha/2}[28]$.

5. Numerical Example: Computing a Two-Sided 95 Percent Confidence Interval for β_1

- Estimate by OLS on the **auto1.dta** sample of N = 74 observations the simple linear regression model given by the population regression equation (*Stata 10* Tutorial 3)

$$price_i = \beta_0 + \beta_1 weight_i + u_i$$

```
. regress price weight

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  1,    72) =   29.42
       Model |   184233937     1   184233937           Prob > F      =  0.0000
    Residual |   450831459    72  6261548.04           R-squared     =  0.2901
-------------+------------------------------           Adj R-squared =  0.2802
       Total |   635065396    73  8699525.97           Root MSE      =  2502.3

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      weight |   2.044063   .3768341     5.42   0.000     1.292857    2.795268
       _cons |  -6.707353    1174.43    -0.01   0.995     -2347.89    2334.475
------------------------------------------------------------------------------
```

```
. display _b[weight]
2.0440626

. display _se[weight]
.37683413

. display invttail(72, 0.025)
1.9934636
```

- Selected results from OLS estimation of the above linear regression model:

$\hat{\beta}_1$ = **2.0440626**

$\hat{se}(\hat{\beta}_1)$ = **0.37683413**

$(1 - \alpha) = 0.95 \quad \Rightarrow \quad \alpha = 1 - 0.95 = 0.05 \quad \Rightarrow \quad \alpha/2 = 0.05/2 =$ **0.025**

$t_{\alpha/2}[N-2] = t_{0.025}[74-2] = t_{0.025}[72]$ = **1.9934636**

- Compute ***upper*** **95% confidence limit for β_1**

$$\begin{aligned}\hat{\beta}_{1U} = \hat{\beta}_1 + t_{0.025}[N-2]\hat{se}(\hat{\beta}_1) &= \mathbf{2.0440626 + 1.9934636(0.37683413)} \\ &= \mathbf{2.0440626 + 0.7512051} \\ &= \mathbf{2.7952677 = \underline{2.795}}\end{aligned}$$

- Compute ***lower*** **95% confidence limit for β_1**

$$\begin{aligned}\hat{\beta}_{1L} = \hat{\beta}_1 - t_{0.025}[N-2]\hat{se}(\hat{\beta}_1) &= \mathbf{2.0440626 - 1.9934636(0.37683413)} \\ &= \mathbf{2.0440626 - 0.7512051} \\ &= \mathbf{1.2928575 = \underline{1.293}}\end{aligned}$$

- ***Result:*** The **two-sided 95% confidence interval for β_1** is: **[1.293, 2.795]**

6. Simulations for Two-Sided 95 Percent Confidence Interval for $\hat{\beta}_1$

The True Model: is given by the **population regression equation (PRE)**

$$Y_i = \beta_0 + \beta_1 X_i + u_i = 70.0 + 0.90 X_i + u_i$$

where

$\boldsymbol{\beta_0 = 70.0}$ and $\boldsymbol{\beta_1 = 0.90}$;

Y_i = weekly consumption expenditures of the i-th household;

X_i = weekly disposable income of the i-th household;

u_i = an iid random error term that is assumed to be $N(0, \sigma^2)$.

Model 3: sets $\sigma^2 = Var(u_i \mid X_i) = 25{,}600$, $\sigma = \sqrt{Var(u_i \mid X_i)} = se(u_i \mid X_i) = 160$.

The Monte Carlo Simulations

- Two different sample sizes: N = 60, N = 120.
- Set population values of X, β_0 and β_1, and $\sigma^2 = \text{Var}(u_i \mid X_i)$.
- Generate 1,000 independent random samples of Y_i and u_i values.
- For each of these 1,000 independent random samples, compute the values of the OLS slope coefficient estimator

  $$\hat{\beta}_1 = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$$

  and its estimated standard error

  $$\hat{se}(\hat{\beta}_1) = \sqrt{\hat{Var}(\hat{\beta}_1)} = \left(\frac{\hat{\sigma}^2}{\sum_i x_i^2}\right)^{\frac{1}{2}} = \frac{\hat{\sigma}}{\sqrt{\sum_i x_i^2}}$$

  where $x_i \equiv X_i - \bar{X}$, $y_i \equiv Y_i - \bar{Y}$, $\bar{X} = \sum_i X_i / N$, and $\bar{Y} = \sum_i Y_i / N$.
- Save the 1,000 values of $\hat{\beta}_1$ and the 1,000 values of $\hat{se}(\hat{\beta}_1)$.
- Use each of the 1,000 values of $\hat{\beta}_1$ and $\hat{se}(\hat{\beta}_1)$ to compute the **two-sided 95 percent confidence interval for the slope coefficient** β_1, and then count **the number and percentage of these 1,000 confidence intervals that contain the true population value of β_1**, which is **0.90**.

Simulation Results for Model 3 for Sample Sizes N = 60 and N = 120 Observations (1,000 Replications)

For sample size N = 60:

- **Number** of two-sided 95% confidence intervals **that contained *true* value of β_1** = **940/1000**
- **Percentage** of two-sided 95% confidence intervals **that contained *true* value of β_1** = **94.0%**

For sample size N = 120:

- **Number** of two-sided 95% confidence intervals **that contained *true* value of β_1** = **952/1000**
- **Percentage** of two-sided 95% confidence intervals **that contained *true* value of β_1** = **95.2%**