ECON 351* -- NOTE 6

The Fundamentals of Statistical Inference in the Simple Linear Regression Model

1. Introduction to Statistical Inference

1.1 Starting Point

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

$$\mathbf{Y_i = \beta_1 + \beta_2 X_i + u_i \qquad (i = 1, \ldots, N) \qquad (1)}$$

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

  (1) the regression coefficients β_1 and β_2

  and

  (2) the error variance σ^2.

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

  (1) the ***unbiased*** **OLS regression coefficient estimators** $\hat{\beta}_1$ **and** $\hat{\beta}_2$

  and

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

$$\hat{\sigma}^2 = \frac{\sum_i \hat{u}_i^2}{(N-2)} = \frac{RSS}{(N-2)} \quad \text{where } \hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i \quad (i = 1, ..., N).$$

Note: $\hat{\sigma}^2$ **is an *unbiased* estimator of the error variance** σ^2:

$$E(\hat{\sigma}^2) = \sigma^2 \quad \text{because} \quad E(RSS) = E(\Sigma_i \hat{u}_i^2) = (N-2)\sigma^2.$$

1.2 Nature of Statistical Inference

Statistical inference consists essentially of **using the point estimates $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\sigma}^2$** of the unknown population parameters β_1, β_2, and σ^2 **to make statements about the true values of β_1 and β_2 *within specified margins of statistical error***.

1.3 Two Related Approaches to Statistical Inference

1. **Interval estimation (or the confidence-interval approach)** involves using the point estimates $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\sigma}^2$ to construct confidence intervals for the regression coefficients that contain the true population parameters β_1 and β_2 ***with some specified probability***.

2. **Hypothesis testing (or the test-of-significance approach)** involves using the point estimates $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\sigma}^2$ to test hypotheses or assertions about the true population values of β_1 and β_2.

- These two approaches to statistical inference are mutually complementary and equivalent in the sense that **they yield identical inferences** about the true values of the population parameters.

- For both types of statistical inference, we need to know the ***form* of the sampling distributions** (or **probability distributions**) of the OLS coefficient estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ and the unbiased (degrees-of-freedom-adjusted) error variance estimator $\hat{\sigma}^2$.

  This is the purpose of the **error normality assumption A9**.

1.4 The Objective: Feasible Test Statistics for $\hat{\beta}_1$ and $\hat{\beta}_2$

For both forms of statistical inference -- interval estimation and hypothesis testing -- it is necessary to obtain ***feasible*** **test statistics for each of the OLS coefficient estimators $\hat{\beta}_1$ and $\hat{\beta}_2$** in the **OLS sample regression equation (OLS-SRE)**

$$\mathbf{Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \qquad (i = 1,\ldots,N) \qquad (2)}$$

1. The **OLS slope coefficient estimator $\hat{\beta}_2$** can be written in deviation-from-means form as:

$$\hat{\beta}_2 = \frac{\sum_i x_i y_i}{\sum_i x_i^2} \qquad \mathbf{(2.1)}$$

where $x_i \equiv X_i - \bar{X}$, $y_i \equiv Y_i - \bar{Y}$, $\bar{X} = \sum_i X_i / N$, and $\bar{Y} = \sum_i Y_i / N$.

2. The **OLS intercept coefficient estimator $\hat{\beta}_1$** can be written as:

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \qquad \mathbf{(2.2)}$$

What's a Feasible Test Statistic?

A ***feasible test statistic*** is **a sample statistic** that must satisfy ***two*** **properties**:

1. It **must have a known probability distribution -- a known sampling distribution**.

2. It **must be a function only of sample data** -- it **must contain no *unknown* parameters** other than the parameter(s) of interest, i.e., the parameters being tested.

What's Ahead in Note 6?

(1) a **statement of the *normality assumption*** and its implications for the sampling distributions of the estimators $\hat{\beta}_1$ and $\hat{\beta}_2$;

(2) a demonstration of how these implications can be used to **derive *feasible* test statistics for $\hat{\beta}_1$ and $\hat{\beta}_2$**; and

(3) **definition of three probability distributions** related to the normal that are used extensively in statistical inference.

What do we know so far about the sampling distributions of $\hat{\beta}_1$ and $\hat{\beta}_2$?

- Their ***means***

$$E(\hat{\beta}_2) = \beta_2 \qquad \text{and} \qquad E(\hat{\beta}_1) = \beta_1$$

- Their ***variances*** *and* ***standard errors***

$$Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum_i x_i^2} \qquad \text{and} \qquad se(\hat{\beta}_2) = \sqrt{Var(\hat{\beta}_2)}$$

$$Var(\hat{\beta}_1) = \frac{\sigma^2 \sum_i X_i^2}{N \sum_i x_i^2} \qquad \text{and} \qquad se(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)}$$

What don't we know about the sampling distributions of $\hat{\beta}_1$ and $\hat{\beta}_2$?

- Their ***mathematical form***

2. The Normality Assumption A9

Statement of the Error Normality Assumption

The normality assumption states that the unobservable random error terms u_i in the population regression equation (PRE)

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{3}$$

are ***independently*** **and** ***identically*** **distributed as the** ***normal*** **distribution** with

(1) ***zero*** **means:** $E(u_i | X_i) = 0 \quad \forall \; i;$

(2) ***constant*** **variances:** $\text{Var}(u_i | X_i) = E(u_i^2 | X_i) = \sigma^2 > 0 \quad \forall \; i;$ **... (A9)**

(3) ***zero*** **covariances:** $\text{Cov}(u_i, u_s | X_i, X_s) = E(u_i u_s | X_i, X_s) = 0 \quad \forall \; i \neq s.$

Compact Forms of the Error Normality Assumption A9

(1) $\mathbf{u_i \sim N(0, \sigma^2)} \quad \forall \; \mathbf{i = 1, ..., N}$ **... (A9.1)**

where $N(0, \sigma^2)$ denotes a normal distribution with mean or expectation equal to 0 and variance equal to σ^2 and the symbol "~" means "is distributed as".

NOTE: The probability density function, or pdf, of the normal distribution is defined as follows: if a random variable $X \sim N(\mu, \sigma^2)$, then the pdf of X is denoted as $f(X_i)$ and takes the form $f(X_i) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{(X_i - \mu)^2}{2\sigma^2}\right]$ where X_i denotes a *realized or observed value* of X.

(2) $\mathbf{u_i}$ **are iid as** $\mathbf{N(0, \sigma^2)}$ **... (A9.2)**

where "iid" means "independently and identically distributed".

(3) $\mathbf{u_i}$ **are NID** $\mathbf{(0, \sigma^2)}$ **... (A9.3)**

where "NID" means "normally and independently distributed".

Implications of the Error Normality Assumption A9

What does Assumption A9 imply for the sampling distribution of $\hat{\beta}_2$?

We use **two linearity properties** to obtain these implications.

1. ***Linearity Property of the Normal Distribution:*** Any linear function of a normally distributed random variable is itself normally distributed.

The **PRE** states that **Y_i is a *linear* function of the error terms u_i**:

$$\underset{\uparrow}{Y_i} = \beta_1 + \beta_2 X_i + \underset{\uparrow}{u_i} \qquad (i = 1,\ldots,N) \tag{4}$$

<u>Implication</u>: The **Y_i must be *normally and independently distributed* (NID)** because the u_i are normally and independently distributed.

$$\mathbf{Y_i \sim N(\beta_1 + \beta_2 X_i, \sigma^2) \quad \forall\, i \quad \textit{or} \quad Y_i \text{ are } NID(\beta_1 + \beta_2 X_i, \sigma^2)}$$

That is, the **Y_i are *normally and independently distributed* (NID)** with

(1) conditional means

$$E(Y_i \mid X_i) = \beta_1 + \beta_2 X_i \qquad \forall\, i = 1, ..., N;$$

(2) conditional variances

$$Var(Y_i \mid X_i) = E(u_i^2 \mid X_i) = \sigma^2 \qquad \forall\, i = 1, ..., N;$$

(3) conditional covariances

$$Cov(Y_i, Y_s \mid X_i, X_s) = E(u_i u_s \mid X_i, X_s) = 0 \qquad \forall\, i \neq s.$$

2. *Linearity Property of the OLS Coefficient Estimator* $\hat{\beta}_2$

$\hat{\beta}_2 = \sum_i k_i Y_i \quad \Rightarrow \quad \hat{\beta}_2$ is a linear function of the *observed* Y_i sample values
$\quad = \beta_2 + \sum_i k_i u_i \quad \Rightarrow \quad \hat{\beta}_2$ is a linear function of the *unobserved* u_i values

where the weights k_i are defined as $k_i = \dfrac{x_i}{\sum_i x_i^2}$: $i = 1, ..., N$.

Implication: $\hat{\beta}_2$ **must be normally distributed**.

The logic of this implication is as follows:

normality of u_i $\Rightarrow$ normality of Y_i $\Rightarrow$ normality of $\hat{\beta}_1$ and $\hat{\beta}_2$.

1. **Error normality assumption A9** states the unobserved u_i values are normally distributed:

   the u_i are NID(0, σ^2)

2. **Linearity property of the normal distribution** implies that the Y_i values are normally distributed:

   the Y_i are NID($\beta_1 + \beta_2 X_i$, σ^2).

3. **Linearity property of the OLS coefficient estimator $\hat{\beta}_2$ *plus* the linearity property of the normal distribution** imply that $\hat{\beta}_2$ is normally distributed:

$$\hat{\beta}_2 \sim N\big(\beta_2, \operatorname{Var}(\hat{\beta}_2)\big) = N\left(\beta_2, \frac{\sigma^2}{\sum_i x_i^2}\right).$$

   Similarly, $\hat{\beta}_1$ is normally distributed:

$$\hat{\beta}_1 \sim N\big(\beta_1, \operatorname{Var}(\hat{\beta}_1)\big) = N\left(\beta_1, \frac{\sigma^2 \sum_i X_i^2}{N \sum_i x_i^2}\right).$$

Complete Implications of the Error Normality Assumption A9

There are ***four*** **specific distributional implications of the error normality assumption A9**.

Implication 1: The **sampling distribution of the OLS *slope* coefficient estimator $\hat{\beta}_2$ is *normal*** with mean $E(\hat{\beta}_2) = \beta_2$ and variance $Var(\hat{\beta}_2) = \sigma^2 / \sum_i x_i^2$: i.e.,

$$\hat{\beta}_2 \sim N\left(\beta_2, Var(\hat{\beta}_2)\right) \qquad \text{where} \qquad Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum_i x_i^2}.$$

Implication 2: The **sampling distribution of the OLS *intercept* coefficient estimator $\hat{\beta}_1$ is *normal*** with mean $E(\hat{\beta}_1) = \beta_1$ and variance $Var(\hat{\beta}_1) = \sigma^2 \sum_i X_i^2 / N \sum_i x_i^2$: i.e.,

$$\hat{\beta}_1 \sim N\left(\beta_1, Var(\hat{\beta}_1)\right) \qquad \text{where} \qquad Var(\hat{\beta}_1) = \frac{\sigma^2 \sum_i X_i^2}{N \sum_i x_i^2}.$$

Implication 3: The statistic $(N-2)\hat{\sigma}^2 / \sigma^2$ has a ***chi-square*** **distribution with (N – 2) degrees of freedom**: i.e.,

$$\frac{(N-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2[N-2] \quad \Rightarrow \quad \frac{\hat{\sigma}^2}{\sigma^2} \sim \frac{\chi^2[N-2]}{(N-2)}.$$

where $\boldsymbol{\chi^2[N-2]}$ denotes the ***chi-square*** **distribution with (N – 2) degrees of freedom** and $\boldsymbol{\hat{\sigma}^2}$ **is the degrees-of-freedom-adjusted estimator of the error variance $\boldsymbol{\sigma^2}$** given by

$$\hat{\sigma}^2 = \frac{\sum_i \hat{u}_i^2}{(N-2)} \qquad \text{where } \hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i \quad (i = 1, ..., N).$$

Implication 4: The **OLS coefficient estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ are *distributed independently* of the error variance estimator $\hat{\sigma}^2$.**

4. Derivation of Feasible Test Statistics for $\hat{\beta}_1$ and $\hat{\beta}_2$

4.1 Definition of a Feasible Test Statistic

A **feasible test statistic** must possess **two critical properties**:

(1) It **must have a known probability distribution**;

(2) It **must be capable of being calculated using only the given sample data (Y_i, X_i), $i = 1, ..., N$ --** it **must contain no unknown parameters** other than the parameter(s) of interest.

4.2 A Standard Normal Z-Statistic for $\hat{\beta}_2$

- **Definition of a Standard Normal Variable**

  If some random variable X ~ N(μ, σ^2), then the standardized normal variable defined as Z = (X − μ)/σ has the standard normal distribution N(0,1).

$$X \sim N(\mu, \sigma^2) \quad \Rightarrow \quad Z = \frac{X - \mu}{\sigma} \sim N(0,1).$$

- **Write the Z-statistic for the coefficient estimator $\hat{\beta}_2$**

  The **error normality assumption A9** implies that

$$\hat{\beta}_2 \sim N\left(\beta_2, \operatorname{Var}(\hat{\beta}_2)\right) \quad \Rightarrow \quad Z(\hat{\beta}_2) = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\operatorname{Var}(\hat{\beta}_2)}} = \frac{\hat{\beta}_2 - \beta_2}{\operatorname{se}(\hat{\beta}_2)} \sim N(0,1)$$

  where $\operatorname{se}(\hat{\beta}_2) = \sqrt{\operatorname{Var}(\hat{\beta}_2)} = \sigma \big/ \sqrt{\sum_i x_i^2}$.

❑ **<u>Result:</u>** The **Z-statistic for $\hat{\beta}_2$** is

$$Z(\hat{\beta}_2) = \frac{\hat{\beta}_2 - \beta_2}{\operatorname{se}(\hat{\beta}_2)} = \frac{\hat{\beta}_2 - \beta_2}{\sigma \big/ \sqrt{\sum_i x_i^2}} = \frac{\left(\hat{\beta}_2 - \beta_2\right)\sqrt{\sum_i x_i^2}}{\sigma} \sim N(0,1) \qquad (8)$$

$\operatorname{se}(\hat{\beta}_2) = \sqrt{\operatorname{Var}(\hat{\beta}_2)}$ is the ***true* standard error of $\hat{\beta}_2$** and $\operatorname{Var}(\hat{\beta}_2)$ is the ***true* variance of $\hat{\beta}_2$**.

❑ **<u>Problem</u>:** The $Z(\hat{\beta}_2)$ statistic in equation (8) is ***not* a feasible test statistic** for $\hat{\beta}_2$ because it contains the ***unknown* parameter σ**, the square root of the unknown error variance σ^2.

4.3 Derivation of the t-Statistic for $\hat{\beta}_2$

- To obtain a feasible test statistic for $\hat{\beta}_2$, we use the **Student's t-distribution**.

- **General Definition of the t-Distribution**

A random variable with the t-distribution is constructed by dividing

(1) a **standard normal random variable Z**

by

(2) the ***square root* of an *independent* chi-square random variable V** that has been divided by its degrees of freedom *m*

The resulting statistic has the **t-distribution with *m* degrees of freedom**.

Formally:

If **(1)** $\mathbf{Z \sim N(0,1)}$

(2) $\mathbf{V \sim \chi^2[m]}$

and **(3)** **Z and V are *independent***,

then the random variable

$$t = \frac{Z}{\sqrt{V/m}} \sim t[m]$$

where **t[m]** denotes the **t-distribution** (or Student's t-distribution) with ***m* degrees of freedom**.

◊ The ***numerator* of a t-statistic** is simply an N(0,1) variable Z.

◊ The ***denominator* of a t-statistic** is the square root of a chi-square distributed random variable divided by its degrees of freedom.

- **Derivation of the t-Statistic for $\hat{\beta}_2$**

- ***Numerator* of the t-statistic for $\hat{\beta}_2$.** The numerator of the t-statistic for $\hat{\beta}_2$ is the $Z(\hat{\beta}_2)$ statistic (8).

$$Z(\hat{\beta}_2) = \frac{\hat{\beta}_2 - \beta_2}{\mathrm{se}(\hat{\beta}_2)} = \frac{\hat{\beta}_2 - \beta_2}{\sigma \big/ \sqrt{\sum_i x_i^2}} = \frac{\left(\hat{\beta}_2 - \beta_2\right)\sqrt{\sum_i x_i^2}}{\sigma} \sim N(0,1) \tag{8}$$

- ***Denominator* of the t-statistic for $\hat{\beta}_2$.** Implication (3) of the normality assumption implies that the statistic $\hat{\sigma}^2 / \sigma^2$ has a degrees-of-freedom-adjusted chi-square distribution with $(N - 2)$ degrees of freedom; that is

$$\frac{(N-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2[N-2] \quad \Rightarrow \quad \frac{\hat{\sigma}^2}{\sigma^2} \sim \frac{\chi^2[N-2]}{(N-2)}. \tag{9}$$

The square root of this statistic is therefore distributed as the square root of a degrees-of-freedom-adjusted chi-square variable with $(N - 2)$ degrees of freedom:

$$\frac{\hat{\sigma}}{\sigma} \sim \left[\frac{\chi^2[N-2]}{(N-2)}\right]^{\frac{1}{2}}. \tag{10}$$

- The **OLS coefficient estimator $\hat{\beta}_2$ is *distributed independently* of the error variance estimator $\hat{\sigma}^2$.**

So the statistics $Z(\hat{\beta}_2) = \dfrac{\hat{\beta}_2 - \beta_2}{\mathrm{se}(\hat{\beta}_2)}$ and $\dfrac{(N-2)\hat{\sigma}^2}{\sigma^2}$ are ***statistically independent***.

- **The *t-statistic* for $\hat{\beta}_2$.** The t-statistic for $\hat{\beta}_2$ is therefore the ratio of (8) to (10): i.e.,

$$t(\hat{\beta}_2) = \frac{Z(\hat{\beta}_2)}{\hat{\sigma}/\sigma} = \frac{\left(\hat{\beta}_2 - \beta_2\right)\sqrt{\sum_i x_i^2}\,/\sigma}{\hat{\sigma}/\sigma}. \tag{11}$$

The t-statistic for $\hat{\beta}_2$ given by (11) can be rewritten without the unknown parameter σ.

◊ Since the unknown parameter σ is the divisor of both the numerator and denominator of $t(\hat{\beta}_2)$, multiply both the numerator and denominator of (11) by σ:

$$t(\hat{\beta}_2) = \frac{\left(\hat{\beta}_2 - \beta_2\right)\sqrt{\sum_i x_i^2}\,/\sigma}{\hat{\sigma}/\sigma} = \frac{\left(\hat{\beta}_2 - \beta_2\right)\sqrt{\sum_i x_i^2}}{\hat{\sigma}}. \tag{12}$$

◊ Dividing the numerator and denominator of (12) by $\sqrt{\sum_i x_i^2}$ yields

$$t(\hat{\beta}_2) = \frac{\left(\hat{\beta}_2 - \beta_2\right)}{\hat{\sigma}\Big/\sqrt{\sum_i x_i^2}} = \frac{\left(\hat{\beta}_2 - \beta_2\right)}{\hat{se}(\hat{\beta}_2)}. \tag{13}$$

where the **denominator of (13)** is simply the ***estimated* standard error of $\hat{\beta}_2$**; i.e.,

$$\frac{\hat{\sigma}}{\sqrt{\sum_i x_i^2}} = \sqrt{\hat{Var}(\hat{\beta}_2)} = \hat{se}(\hat{\beta}_2).$$

- **Result:** The **t-statistic for $\hat{\beta}_2$** takes the form

$$t(\hat{\beta}_2) = \frac{\left(\hat{\beta}_2 - \beta_2\right)}{\hat{\sigma}\Big/\sqrt{\sum_i x_i^2}} = \frac{\left(\hat{\beta}_2 - \beta_2\right)}{\sqrt{\hat{Var}(\hat{\beta}_2)}} = \frac{\left(\hat{\beta}_2 - \beta_2\right)}{\hat{se}(\hat{\beta}_2)}. \qquad \textbf{(14)}$$

Note that, unlike the $Z(\hat{\beta}_2)$ statistic in (8), **the $t(\hat{\beta}_2)$ statistic in (14) is a *feasible* test statistic for $\hat{\beta}_2$** because its satisfies both the requirements for a feasible test statistic.

(1) Its sampling distribution is known; it has the t[N−2] distribution, the t-distribution with (N − 2) degrees of freedom:

$$t(\hat{\beta}_2) = \frac{\left(\hat{\beta}_2 - \beta_2\right)}{\hat{se}(\hat{\beta}_2)} \sim t[N-2].$$

(2) Its value can be calculated from sample data for any hypothesized value of β_2 -- i..e, **it contains no unknown parameters other than β_2.**

- **Result:** The **t-statistic for $\hat{\beta}_1$** is analogous to that for $\hat{\beta}_2$ and has the same distribution: i.e.,

$$t(\hat{\beta}_1) = \frac{\left(\hat{\beta}_1 - \beta_1\right)}{\sqrt{\hat{Var}(\hat{\beta}_1)}} = \frac{\left(\hat{\beta}_1 - \beta_1\right)}{\hat{se}(\hat{\beta}_1)} \sim t[N-2]$$

where the estimated standard error for $\hat{\beta}_1$ is

$$\hat{se}(\hat{\beta}_1) = \sqrt{\hat{Var}(\hat{\beta}_1)} = \sqrt{\frac{\hat{\sigma}^2 \sum_i X_i^2}{N \sum_i x_i^2}}.$$

4.4 Derivation of the F-Statistic for $\hat{\beta}_2$

- A second feasible test statistic for $\hat{\beta}_2$ can be derived from the normality assumption A9 using the **F-distribution**.

- **General Definition of the F-Distribution**

A random variable with the F-distribution is the ratio of two ratios:

(1) one **chi-square distributed random variable V_1 divided by its degrees of freedom m_1**

and

(2) a second ***independent*** **chi-square distributed random variable V_2** that also has been **divided by its degrees of freedom m_2**.

The resulting statistic has the **F-distribution with *m_1 numerator* degrees of freedom and *m_2 denominator* degrees of freedom**.

Formally:

If **(1)** $\mathbf{V_1 \sim \chi^2[m_1]}$
(2) $\mathbf{V_2 \sim \chi^2[m_2]}$
and **(3)** **V_1 and V_2 are *independent***,

then the random variable

$$F = \frac{V_1/m_1}{V_2/m_2} \sim F[m_1, m_2]$$

where $F[m_1, m_2]$ denotes the **F-distribution** (or Fisher's F-distribution) **with *m_1 numerator* degrees of freedom and *m_2 denominator* degrees of freedom**.

❑ **Derivation of the F-Statistic for $\hat{\beta}_2$**

♦ ***Numerator* of the F-statistic for $\hat{\beta}_2$.** The numerator of the F-statistic for $\hat{\beta}_2$ is the square of the $Z(\hat{\beta}_2)$ statistic (8). Recall that **the *square* of a standard normal N(0,1) random variable has a *chi-square distribution* with *one* degree of freedom**. Re-write the $Z(\hat{\beta}_2)$ statistic as in (8) above:

$$Z(\hat{\beta}_2) = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = \frac{\hat{\beta}_2 - \beta_2}{\sigma \Big/ \sqrt{\sum_i x_i^2}} = \frac{\left(\hat{\beta}_2 - \beta_2\right)\sqrt{\sum_i x_i^2}}{\sigma} \sim N(0,1). \qquad (8)$$

The *square* of the $Z(\hat{\beta}_2)$ statistic is therefore:

$$\left(Z(\hat{\beta}_2)\right)^2 = \frac{\left(\hat{\beta}_2 - \beta_2\right)^2}{\left(se(\hat{\beta}_2)\right)^2} = \frac{\left(\hat{\beta}_2 - \beta_2\right)^2}{\sigma^2 \Big/ \left(\sum_i x_i^2\right)} = \frac{\left(\hat{\beta}_2 - \beta_2\right)^2 \left(\sum_i x_i^2\right)}{\sigma^2} \sim \chi^2[1]. \qquad (15)$$

♦ ***Denominator* of the F-statistic for $\hat{\beta}_2$.** Implication (3) of the normality assumption implies that the statistic $\hat{\sigma}^2/\sigma^2$ has a degrees-of-freedom-adjusted chi-square distribution with $(N-2)$ degrees of freedom; that is

$$\frac{(N-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2[N-2] \quad \Rightarrow \quad \frac{\hat{\sigma}^2}{\sigma^2} \sim \frac{\chi^2[N-2]}{(N-2)}. \qquad (9)$$

♦ It is possible to show that the $\chi^2[1]$-distributed statistic $\left(Z(\hat{\beta}_2)\right)^2$ in (15) and the $\chi^2[N-2]$-distributed statistic $(N-2)\hat{\sigma}^2/\sigma^2$ in (9) are **statistically *independent***.

♦ **The *F-statistic* for $\hat{\beta}_2$.** The F-statistic for $\hat{\beta}_2$ is therefore the ratio of (15) to (9):

$$
\begin{aligned}
F(\hat{\beta}_2) &= \frac{\left(Z(\hat{\beta}_2)\right)^2}{\hat{\sigma}^2 / \sigma^2} \\
&= \frac{\left(\hat{\beta}_2 - \beta_2\right)^2 \left(\sum_i x_i^2\right) / \sigma^2}{\hat{\sigma}^2 / \sigma^2} \\
&= \frac{\left(\hat{\beta}_2 - \beta_2\right)^2 \left(\sum_i x_i^2\right)}{\hat{\sigma}^2} \qquad (16) \\
&= \frac{\left(\hat{\beta}_2 - \beta_2\right)^2}{\hat{\sigma}^2 / \sum_i x_i^2} \\
&= \frac{\left(\hat{\beta}_2 - \beta_2\right)^2}{\hat{\mathrm{V}}\mathrm{ar}(\hat{\beta}_2)} \qquad \text{since } \hat{\sigma}^2 / \sum_i x_i^2 = \hat{\mathrm{V}}\mathrm{ar}(\hat{\beta}_2).
\end{aligned}
$$

❑ **<u>Result</u>:** The **F-statistic for $\hat{\beta}_2$** takes the form

$$F(\hat{\beta}_2) = \frac{\left(\hat{\beta}_2 - \beta_2\right)^2}{\hat{\sigma}^2 / \left(\sum_i x_i^2\right)} = \frac{\left(\hat{\beta}_2 - \beta_2\right)^2}{\hat{\mathrm{V}}\mathrm{ar}(\hat{\beta}_2)} \sim F[1, N-2]. \qquad \textbf{(17)}$$

Like the $t(\hat{\beta}_2)$ statistic in (14), **the $F(\hat{\beta}_2)$ statistic in (17) is a *feasible* test statistic for $\hat{\beta}_2$**; it satisfies both the requirements for a feasible test statistic.

(1) First, **its sampling distribution is known**; it has the **F[1, N–2] distribution**, the F-distribution with 1 numerator degree of freedom and (N − 2) denominator degrees of freedom:

$$F(\hat{\beta}_2) = \frac{\left(\hat{\beta}_2 - \beta_2\right)^2}{\hat{\mathrm{V}}\mathrm{ar}(\hat{\beta}_2)} \sim F[1, N-2].$$

(2) Second, **its value can be calculated entirely from sample data** for any hypothesized value of β_2.

❑ **<u>Result</u>:** The **F-statistic for $\hat{\beta}_1$** is analogous to that for $\hat{\beta}_2$ and has the same distribution: i.e.,

$$F(\hat{\beta}_1) = \frac{\left(\hat{\beta}_1 - \beta_1\right)^2}{\hat{Var}(\hat{\beta}_1)} \sim F[1, N-2].$$

where the estimated variance for $\hat{\beta}_1$ is

$$\hat{Var}(\hat{\beta}_1) = \frac{\hat{\sigma}^2 \sum_i X_i^2}{N \sum_i x_i^2}.$$

❑ **Relationship Between the t-statistic and the F-statistic for $\hat{\beta}_j$, j = 1, 2:**

- The ***F-statistic*** **for $\hat{\beta}_2$ is the *square* of the *t-statistic* for $\hat{\beta}_2$**:

$$F(\hat{\beta}_2) = \frac{\left(\hat{\beta}_2 - \beta_2\right)^2}{\hat{Var}(\hat{\beta}_2)} = \frac{\left(\hat{\beta}_2 - \beta_2\right)^2}{\left(\hat{se}(\hat{\beta}_2)\right)^2} = \left(\frac{\hat{\beta}_2 - \beta_2}{\hat{se}(\hat{\beta}_2)}\right)^2 = \left(t(\hat{\beta}_2)\right)^2.$$

- Similarly, the ***F-statistic*** **for $\hat{\beta}_1$ is the *square* of the *t-statistic* for $\hat{\beta}_1$**:

$$F(\hat{\beta}_1) = \frac{\left(\hat{\beta}_1 - \beta_1\right)^2}{\hat{Var}(\hat{\beta}_1)} = \frac{\left(\hat{\beta}_1 - \beta_1\right)^2}{\left(\hat{se}(\hat{\beta}_1)\right)^2} = \left(\frac{\hat{\beta}_1 - \beta_1}{\hat{se}(\hat{\beta}_1)}\right)^2 = \left(t(\hat{\beta}_1)\right)^2.$$

- The t-distribution and the F-distribution are also related as follows:

$$F[1, N-2] = \left(t[N-2]\right)^2 \qquad \textit{or} \qquad t[N-2] = \sqrt{F[1, N-2]}.$$

That is, the F-distribution with 1 numerator degree of freedom and N−2 denominator degrees of freedom *equals* the square of the t-distribution with N−2 degrees of freedom. Conversely, the t-distribution with N−2 degrees of freedom *equals* the square root of the F-distribution with 1 numerator degree of freedom and N−2 denominator degrees of freedom.

4.5 Important Results: Summary

1. Under the **error normality assumption A9**, the **sample statistics $t(\hat{\beta}_2)$ and $t(\hat{\beta}_1)$ have the t-distribution with N−2 degrees of freedom**:

$$t(\hat{\beta}_2) = \frac{\left(\hat{\beta}_2 - \beta_2\right)}{\sqrt{\hat{\text{Var}}(\hat{\beta}_2)}} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\text{se}}(\hat{\beta}_2)} \sim t[N-2];$$

$$t(\hat{\beta}_1) = \frac{\left(\hat{\beta}_1 - \beta_1\right)}{\sqrt{\hat{\text{Var}}(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\text{se}}(\hat{\beta}_1)} \sim t[N-2].$$

2. Under the **error normality assumption A9**, the **sample statistics $F(\hat{\beta}_2)$ and $F(\hat{\beta}_1)$ have the F-distribution with 1 numerator degree of freedom and N−2 denominator degrees of freedom**:

$$F(\hat{\beta}_2) = \frac{\left(\hat{\beta}_2 - \beta_2\right)^2}{\hat{\text{Var}}(\hat{\beta}_2)} \sim F[1, N-2];$$

$$F(\hat{\beta}_1) = \frac{\left(\hat{\beta}_1 - \beta_1\right)^2}{\hat{\text{Var}}(\hat{\beta}_1)} \sim F[1, N-2].$$

Note that $\hat{\text{se}}(\hat{\beta}_2) = \sqrt{\hat{\text{Var}}(\hat{\beta}_2)}$ and $\hat{\text{se}}(\hat{\beta}_1) = \sqrt{\hat{\text{Var}}(\hat{\beta}_1)}$ are the ***estimated* standard errors**, and $\hat{\text{Var}}(\hat{\beta}_2)$ and $\hat{\text{Var}}(\hat{\beta}_1)$ are the ***estimated* variances**, of the OLS coefficient estimators $\hat{\beta}_2$ and $\hat{\beta}_1$, respectively.

3. The **Z-statistics for $\hat{\beta}_1$ and $\hat{\beta}_2$ <u>are</u> <u>not</u> *feasible test statistics***.

$$Z(\hat{\beta}_1) = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\mathrm{Var}(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - \beta_1}{\mathrm{se}(\hat{\beta}_1)} \quad \textit{and} \quad Z(\hat{\beta}_2) = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\mathrm{Var}(\hat{\beta}_2)}} = \frac{\hat{\beta}_2 - \beta_2}{\mathrm{se}(\hat{\beta}_2)}.$$

They require for their computation the ***true* but *unknown* variances and standard errors** of the OLS coefficient estimators, and these require that the value of the error variance σ^2 be known.

But since the value of σ^2 is almost always unknown in practice, the values of $\mathrm{Var}(\hat{\beta}_1)$ and $\mathrm{Var}(\hat{\beta}_2)$, and of $\mathrm{se}(\hat{\beta}_1)$ and $\mathrm{se}(\hat{\beta}_2)$, are also unknown.

4. The **t-statistics for $\hat{\beta}_1$ and $\hat{\beta}_2$ <u>are</u> *feasible test statistics***.

$$t(\hat{\beta}_1) = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\mathrm{Var}}(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\mathrm{se}}(\hat{\beta}_1)} \quad \textit{and} \quad t(\hat{\beta}_2) = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{\hat{\mathrm{Var}}(\hat{\beta}_2)}} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\mathrm{se}}(\hat{\beta}_2)}.$$

They are obtained by replacing the ***unknown* variances and standard errors** of the OLS coefficient estimators in the Z-statistics $Z(\hat{\beta}_1)$ and $Z(\hat{\beta}_2)$ with their corresponding ***estimated* variances** $\hat{\mathrm{Var}}(\hat{\beta}_1)$ and $\hat{\mathrm{Var}}(\hat{\beta}_2)$ and ***estimated* standard errors** $\hat{\mathrm{se}}(\hat{\beta}_1) = \sqrt{\hat{\mathrm{Var}}(\hat{\beta}_1)}$ and $\hat{\mathrm{se}}(\hat{\beta}_2) = \sqrt{\hat{\mathrm{Var}}(\hat{\beta}_2)}$.

5. The **F-statistics for $\hat{\beta}_1$ and $\hat{\beta}_2$** also **<u>are</u> *feasible test statistics***.

$$F(\hat{\beta}_1) = \frac{\left(\hat{\beta}_1 - \beta_1\right)^2}{\hat{\mathrm{Var}}(\hat{\beta}_1)} \quad \textit{and} \quad F(\hat{\beta}_2) = \frac{\left(\hat{\beta}_2 - \beta_2\right)^2}{\hat{\mathrm{Var}}(\hat{\beta}_2)}.$$

The denominators of $F(\hat{\beta}_1)$ and $F(\hat{\beta}_2)$ are the ***estimated* variances** $\hat{\mathrm{Var}}(\hat{\beta}_1)$ and $\hat{\mathrm{Var}}(\hat{\beta}_2)$, ***not* the *true* variances** $\mathrm{Var}(\hat{\beta}_1)$ and $\mathrm{Var}(\hat{\beta}_2)$.

❑ **Why is the Error Normality Assumption A9 Important?**

- The **normality assumption A9** permits us **to derive the *functional form* of the sampling distributions of $\hat{\beta}_1$ (normal), $\hat{\beta}_2$ (normal), and $\hat{\sigma}^2$ (chi-square)**.

- Knowing the form of the sampling distributions of $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\sigma}^2$ enables us **to derive *feasible* test statistics for the OLS coefficient estimators $\hat{\beta}_1$ and $\hat{\beta}_2$**.

- These **feasible test statistics for $\hat{\beta}_1$ and $\hat{\beta}_2$** enable us **to conduct *statistical inference*** -- i.e.,

  (1) to *construct confidence intervals* for β_1 and β_2

  and

  (2) to *perform* statistical *hypothesis tests* about the values of β_1 and β_2 .