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

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad (i = 1, ..., N) \qquad \textbf{(1)}$$

- 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** $\hat{\sigma}^2$ given by the formula

$$\hat{\sigma}^2 = \frac{\sum_i \hat{u}_i^2}{(N-2)} = \frac{RSS}{(N-2)} \quad \text{where} \quad \hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i \quad (i = 1, ..., N).$$

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

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

1.2 Nature of Statistical Inference

Statistical inference consists essentially of **using the point estimates** $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\sigma}^2$ of the unknown population parameters β_0, β_1 and σ^2 **to make statements about the true values of β_0 and β_1 *within specified margins of statistical error***.

1.3 Two Related Approaches to Statistical Inference

1. **<u>Interval estimation</u> (or the confidence-interval approach)** involves using the point estimates $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\sigma}^2$ to construct confidence intervals for the regression coefficients that contain the true population parameters β_0 and β_1 ***with some specified probability***.

2. **<u>Hypothesis testing</u> (or the test-of-significance approach)** involves using the point estimates $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\sigma}^2$ to test hypotheses or assertions about the true population values of β_0 and β_1.

- 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 $\hat{\beta}_0$ and $\hat{\beta}_1$ 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 $\hat{\beta}_0$ and $\hat{\beta}_1$

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 $\hat{\beta}_0$ and $\hat{\beta}_1$** in the **OLS sample regression equation (OLS-SRE)**

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

1. The **OLS slope coefficient estimator $\hat{\beta}_1$** can be written in deviation-from-means form as:

$$\hat{\beta}_1 = \frac{\sum_i x_i y_i}{\sum_i x_i^2} = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2} \qquad \textbf{(2.1)}$$

   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 $\hat{\beta}_0$** can be written as:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \qquad \textbf{(2.2)}$$

<u>What's a Feasible Test Statistic?</u>

A ***feasible test statistic*** is **a sample statistic** that satisfies **<u>two</u> properties**:

1. It **has a known probability distribution** for the true population value of some parameter(s).

2. It **is a function only of sample data --** i.e., its value can be computed for some hypothesized value of the unknown parameter(s) using only sample data.

   The **formula** for a feasible test statistic **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 $\hat{\beta}_0$ and $\hat{\beta}_1$;

(2) a demonstration of how these implications can be used to **derive *feasible* test statistics for $\hat{\beta}_0$ and $\hat{\beta}_1$**; and

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

2. The Normality Assumption A9

2.1 Statement of Normality Assumption

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

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad (i = 1, ..., N) \qquad (3)$$

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

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

<u>Compact Forms of the Normality Assumption A9</u>

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

<u>NOTE</u>: The normal distribution has only two parameters: its mean or expectation, and its variance. 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) u_i are iid as N(0, σ^2) ... (A9.2)

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

- Parts (1) and (2) of A9 -- the assumptions of zero means and constant variances -- are the "identically" part of the iid specification.

- Part (3) of A9 -- the assumption of zero error covariances -- is the "independently" part of the iid specification.

(3) u_i are NID(0, σ^2) ... (A9.3)

where NID(0, σ^2) means "normally and independently distributed with zero mean and constant variance σ^2".

2.2 Implications of Normality Assumption A9 for the Distribution of Y_i

The error normality assumption A9 implies that the **sample values Y_i** of the regressand (or dependent variable) Y **are also *normally distributed***.

- **<u>Linearity Property of Normal Distribution</u>:** Any linear function of a normally distributed random variable is itself normally distributed.

- The **PRE**, or **population regression equation**, states that **Y_i is a *linear* function of the error terms u_i**:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad (4)$$
$$\uparrow\_\_\_\_\_\_\_\_\_\_\_\_\_\_\uparrow$$

- Since **any *linear* function of a *normally distributed random variable* is itself normally distributed**, the error normality assumption A9 implies that

$$\mathbf{Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2) \qquad i = 1, ..., N \qquad (5.1)}$$

or

$$\mathbf{Y_i \text{ are NID}(\beta_0 + \beta_1 X_i, \sigma^2) \qquad i = 1, ..., N \qquad (5.2)}$$

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

(1) conditional means

$$E(Y_i \mid X_i) = \beta_0 + \beta_1 X_i \qquad \forall\ i = 1, ..., N;$$

(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

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

3. The Sampling Distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$

To obtain the sampling distributions of the OLS coefficient estimators $\hat{\beta}_0$ and $\hat{\beta}_1$, we make use of

1. the **linearity property** of the **OLS coefficient estimators $\hat{\beta}_0$ and $\hat{\beta}_1$**;

and

2. the **normality assumption A9**.

3.1 Linearity Property of the OLS Coefficient Estimators $\hat{\beta}_0$ and $\hat{\beta}_1$

The linearity property of the OLS coefficient estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ means that the normality of u_i and Y_i carries over to the sampling distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$. For example, recall that the OLS slope coefficient estimator $\hat{\beta}_1$ can be written as the following linear functions of the Y_i and u_i:

$$\begin{aligned} \hat{\beta}_1 &= \textstyle\sum_i k_i Y_i \\ &= \beta_1 + \textstyle\sum_i k_i u_i \end{aligned} \qquad (6)$$

where the observation weights, the k_i, are defined as

$$k_i = \frac{x_i}{\sum_i x_i^2} \qquad (i = 1, ..., N).$$

- ***Result:*** Equation (6) indicates that the **OLS slope coefficient estimator $\hat{\beta}_1$** is a ***linear* function** of both the ***observed* Y_i values** and the ***unobserved* u_i values**.

3.2 Implications of Normality Assumption and Linearity Property

Because any linear function of normally distributed random variables is itself normally distributed, the sampling distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$ are also normally distributed. The reasoning is as follows:

normality of u_i $\Rightarrow$ normality of Y_i $\Rightarrow$ normality of $\hat{\beta}_0$ and $\hat{\beta}_1$.

There are **<u>four</u> specific distributional implications of the error normality assumption A9**.

<u>Implication 1</u>: The **sampling distribution of the OLS slope coefficient estimator $\hat{\beta}_1$ is *normal*** with mean $E(\hat{\beta}_1) = \beta_1$ and variance $Var(\hat{\beta}_1) = \sigma^2 / \sum_i x_i^2$: i.e.,

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

The normality of the sampling distribution of $\hat{\beta}_1$ implies that the **Z-statistic** $Z(\hat{\beta}_1)$ = $(\hat{\beta}_1 - \beta_1)/se(\hat{\beta}_1)$ has the ***standard* normal distribution** with mean 0 and variance 1: i.e.,

$$Z(\hat{\beta}_1) = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{Var(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1}{\sigma / \left(\sum_i x_i^2\right)^{1/2}} \sim N(0,1).$$

<u>Implication 2</u>: The **sampling distribution of the OLS intercept coefficient estimator $\hat{\beta}_0$ is *normal*** with mean $E(\hat{\beta}_0) = \beta_0$ and variance $Var(\hat{\beta}_0) = \sigma^2 \sum_i X_i^2 / N \sum_i x_i^2$: i.e.,

$$\hat{\beta}_0 \sim N\left(\beta_0, Var(\hat{\beta}_0)\right) = N\left(\beta_0, \frac{\sigma^2 \sum_i X_i^2}{N \sum_i x_i^2}\right).$$

The normality of the sampling distribution of $\hat{\beta}_0$ implies that the **Z-statistic** $Z(\hat{\beta}_0) = (\hat{\beta}_0 - \beta_0)/se(\hat{\beta}_0)$ has the ***standard* normal distribution** with mean 0 and variance 1: i.e.,

$$Z(\hat{\beta}_0) = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{Var(\hat{\beta}_0)}} = \frac{\hat{\beta}_0 - \beta_0}{se(\hat{\beta}_0)} = \frac{\hat{\beta}_0 - \beta_0}{\sigma\left(\sum_i X_i^2\right)^{1/2} / N^{1/2}\left(\sum_i x_i^2\right)^{1/2}} \sim N(0,1).$$

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

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

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

$$\hat{\sigma}^2 = \frac{\sum_i \hat{u}_i^2}{(N-2)} \qquad \text{where} \quad \hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i \quad (i = 1, ..., N).$$

<u>Implication 4</u>: The **OLS coefficient estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are *distributed independently* of the error variance estimator $\hat{\sigma}^2$.**

4. Derivation of Test Statistics for $\hat{\beta}_0$ and $\hat{\beta}_1$

4.1 Definition of a Feasible Test Statistic

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

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

(2) It must be capable of being **calculated using only the given sample data** (Y_i, X_i), i = 1, ..., N.

We illustrate the derivation of feasible test statistics for $\hat{\beta}_0$ and $\hat{\beta}_1$ by considering the slope coefficient estimator $\hat{\beta}_1$; an analogous argument can be used to obtain a feasible test statistic for $\hat{\beta}_0$.

4.2 A Standard Normal Z-Statistic for $\hat{\beta}_1$

The normality of the sampling distribution of $\hat{\beta}_1$ implies that $\hat{\beta}_1$ can be written in the form of a **standard normal variable**, or **Z-statistic**, with a mean of zero and a variance of one and is denoted as **N(0,1)**.

- **Definition of a Standard Normal Variable**

If some random variable $X \sim N(\mu, \sigma^2)$, then the standardized normal variable defined as $Z = (X - \mu)/\sigma$ 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 standard normal variable, or Z-statistic, for the coefficient estimator $\hat{\beta}_1$**

♦ The error normality assumption A9 implies that

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

♦ The Z-statistic for $\hat{\beta}_1$ is therefore defined as

$$Z(\hat{\beta}_1) = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{Var(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1}{\sigma \Big/ \left(\sum_i x_i^2\right)^{1/2}}$$

where $Var(\hat{\beta}_1)$ is the ***true* variance of** $\hat{\beta}_1$ and $se(\hat{\beta}_1) \equiv [Var(\hat{\beta}_1)]^{1/2}$ is the ***true* standard error of $\hat{\beta}_1$**.

♦ Since $\hat{\beta}_1$ has the normal distribution, the statistic $Z(\hat{\beta}_1)$ has the standard normal distribution N(0,1): that is,

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_i x_i^2}\right) \quad \Rightarrow \quad Z(\hat{\beta}_1) = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1}{\sigma \Big/ \left(\sum_i x_i^2\right)^{1/2}} \sim N(0,1). \qquad (7)$$

❑ **<u>Problem</u>:** The $Z(\hat{\beta}_1)$ statistic in equation (7) is ***not* a feasible test statistic** for $\hat{\beta}_1$ because it involves the ***unknown* parameter σ**, the square root of the unknown error variance σ^2.

4.3 Derivation of the t-Statistic for $\hat{\beta}_1$

- To obtain a feasible test statistic for $\hat{\beta}_1$, 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) $Z \sim N(0,1)$

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

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.

Result:

A **t-statistic** is simply the ratio of a standard normal variable to the square root of an *independent* degrees-of-freedom-adjusted chi-square variable with *m* degrees of freedom.

- **Derivation of the t-Statistic for $\hat{\beta}_1$**

- ***Numerator* of the t-statistic for $\hat{\beta}_1$.** The numerator of the t-statistic for $\hat{\beta}_1$ is the $Z(\hat{\beta}_1)$ statistic (7). It will be convenient to re-write the $Z(\hat{\beta}_1)$ statistic in the form

$$Z(\hat{\beta}_1) = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1}{\sigma\big/\left(\sum_i x_i^2\right)^{1/2}} = \frac{\left(\hat{\beta}_1 - \beta_1\right)\left(\sum_i x_i^2\right)^{1/2}}{\sigma} \sim N(0,1) \qquad (8)$$

- ***Denominator* of the t-statistic for $\hat{\beta}_1$.** 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

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

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:

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

- **The *t-statistic* for $\hat{\beta}_1$.** The t-statistic for $\hat{\beta}_1$ is therefore the ratio of (8) to (10): i.e.,

$$t(\hat{\beta}_1) = \frac{Z(\hat{\beta}_1)}{\hat{\sigma}/\sigma} = \frac{\left(\hat{\beta}_1 - \beta_1\right)\left(\sum_i x_i^2\right)^{1/2}/\sigma}{\hat{\sigma}/\sigma}. \qquad (11)$$

The t-statistic for $\hat{\beta}_1$ 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 $t(\hat{\beta}_1)$, multiplication of the numerator and denominator of (11) by σ permits the t-statistic for $\hat{\beta}_1$ to be written as

$$t(\hat{\beta}_1) = \frac{\left(\hat{\beta}_1 - \beta_1\right)\left(\sum_i x_i^2\right)^{1/2}/\sigma}{\hat{\sigma}/\sigma} = \frac{\left(\hat{\beta}_1 - \beta_1\right)\left(\sum_i x_i^2\right)^{1/2}}{\hat{\sigma}}. \qquad (12)$$

  - Dividing the numerator and denominator of (12) by $\left(\sum_i x_i^2\right)^{1/2}$ yields

$$t(\hat{\beta}_1) = \frac{\left(\hat{\beta}_1 - \beta_1\right)}{\hat{\sigma}/\left(\sum_i x_i^2\right)^{1/2}}. \qquad (13)$$

  - But the **denominator of (13)** is simply the ***estimated* standard error of** $\hat{\beta}_1$; i.e.,

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

- ***Result:*** The **t-statistic for $\hat{\beta}_1$** takes the form

$$t(\hat{\beta}_1) = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}\big/\left(\sum_i x_i^2\right)^{1/2}} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{Var}(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - \beta_1}{\hat{se}(\hat{\beta}_1)}. \qquad \textbf{(14)}$$

Note that, unlike the $Z(\hat{\beta}_1)$ statistic in (8), the $t(\hat{\beta}_1)$ statistic in (14) is a ***feasible* test statistic** for $\hat{\beta}_1$ because its satisfies both the requirements for a feasible test statistic.

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

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

(2) Second, **its value can be calculated from sample data** for any hypothesized value of β_1.

- ***Result:*** The **t-statistic for $\hat{\beta}_0$** is analogous to that for $\hat{\beta}_1$ and has the same distribution: i.e.,

$$t(\hat{\beta}_0) = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{Var}(\hat{\beta}_0)}} = \frac{\hat{\beta}_0 - \beta_0}{\hat{se}(\hat{\beta}_0)} \sim t[N-2]$$

where the estimated standard error for $\hat{\beta}_0$ is

$$\hat{se}(\hat{\beta}_0) = \sqrt{\hat{Var}(\hat{\beta}_0)} = \left[\frac{\hat{\sigma}^2 \sum_i X_i^2}{N \sum_i x_i^2}\right]^{\frac{1}{2}}.$$

4.4 Derivation of the F-Statistic for $\hat{\beta}_1$

- A second feasible test statistic for $\hat{\beta}_1$ 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 independent random variables**:

(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) $V_1 \sim \chi^2[m_1]$
(2) $V_2 \sim \chi^2[m_2]$
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 $\hat{\beta}_1$

- ***Numerator* of the F-statistic for $\hat{\beta}_1$.** The numerator of the F-statistic for $\hat{\beta}_1$ is the square of the $Z(\hat{\beta}_1)$ statistic (7). 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 $Z(\hat{\beta}_1)$ statistic as in (8) above:

$$Z(\hat{\beta}_1) = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1}{\sigma \Big/ \left(\sum_i x_i^2\right)^{1/2}} = \frac{\left(\hat{\beta}_1 - \beta_1\right)\left(\sum_i x_i^2\right)^{1/2}}{\sigma} \sim N(0,1). \qquad (8)$$

The *square* of the $Z(\hat{\beta}_1)$ statistic is therefore:

$$\left(Z(\hat{\beta}_1)\right)^2 = \frac{\left(\hat{\beta}_1 - \beta_1\right)^2}{\left(se(\hat{\beta}_1)\right)^2} = \frac{\left(\hat{\beta}_1 - \beta_1\right)^2}{\sigma^2 \Big/ \left(\sum_i x_i^2\right)} = \frac{\left(\hat{\beta}_1 - \beta_1\right)^2 \left(\sum_i x_i^2\right)}{\sigma^2} \sim \chi^2[1]. \qquad (15)$$

- ***Denominator* of the F-statistic for $\hat{\beta}_1$.** 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 $\left(Z(\hat{\beta}_1)\right)^2$ 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 $\hat{\beta}_1$.** The F-statistic for $\hat{\beta}_1$ is therefore the ratio of (15) to (9):

$$\begin{aligned} F(\hat{\beta}_1) &= \frac{\left(Z(\hat{\beta}_1)\right)^2}{\hat{\sigma}^2 / \sigma^2} \\ &= \frac{\left(\hat{\beta}_1 - \beta_1\right)^2 \left(\sum_i x_i^2\right) / \sigma^2}{\hat{\sigma}^2 / \sigma^2} \\ &= \frac{\left(\hat{\beta}_1 - \beta_1\right)^2 \left(\sum_i x_i^2\right)}{\hat{\sigma}^2} \qquad (16) \\ &= \frac{\left(\hat{\beta}_1 - \beta_1\right)^2}{\hat{\sigma}^2 / \sum_i x_i^2} \\ &= \frac{\left(\hat{\beta}_1 - \beta_1\right)^2}{\hat{Var}(\hat{\beta}_1)} \qquad \text{since } \hat{\sigma}^2 / \sum_i x_i^2 = \hat{Var}(\hat{\beta}_1). \end{aligned}$$

❑ ***Result:*** The **F-statistic for $\hat{\beta}_1$** takes the form

$$F(\hat{\beta}_1) = \frac{\left(\hat{\beta}_1 - \beta_1\right)^2}{\hat{\sigma}^2 / \left(\sum_i x_i^2\right)} = \frac{\left(\hat{\beta}_1 - \beta_1\right)^2}{\hat{Var}(\hat{\beta}_1)} \sim F[1, N-2]. \qquad \textbf{(17)}$$

Note that, like the $t(\hat{\beta}_1)$ statistic in (14), the $F(\hat{\beta}_1)$ statistic in (17) is a ***feasible* test statistic** for $\hat{\beta}_1$ because its 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:

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

(2) Second, **its value can be calculated entirely from sample data** for any hypothesized value of β_1.

- ***Result:*** The **F-statistic for $\hat{\beta}_0$** is analogous to that for $\hat{\beta}_1$ and has the same distribution: i.e.,

$$F(\hat{\beta}_0) = \frac{\left(\hat{\beta}_0 - \beta_0\right)^2}{\hat{Var}(\hat{\beta}_0)} \sim F[1, N-2]$$

where the estimated variance for $\hat{\beta}_0$ is

$$\hat{Var}(\hat{\beta}_0) = \frac{\hat{\sigma}^2 \sum_i X_i^2}{N \sum_i x_i^2}.$$

- **Relationship Between the t-statistic and the F-statistic for $\hat{\beta}_j$, j = 0, 1:**

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

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

- Similarly, the ***F-statistic*** **for $\hat{\beta}_0$ is the *square* of the *t-statistic* for $\hat{\beta}_0$**:

$$F(\hat{\beta}_0) = \frac{\left(\hat{\beta}_0 - \beta_0\right)^2}{\hat{Var}(\hat{\beta}_0)} = \frac{\left(\hat{\beta}_0 - \beta_0\right)^2}{\left(\hat{se}(\hat{\beta}_0)\right)^2} = \left(\frac{\hat{\beta}_0 - \beta_0}{\hat{se}(\hat{\beta}_0)}\right)^2 = \left(t(\hat{\beta}_0)\right)^2.$$

- The ***t-distribution*** and the ***F-distribution*** **are also related**.

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

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 $t(\hat{\beta}_1)$ and $t(\hat{\beta}_0)$ have the t-distribution with N−2 degrees of freedom**:

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

$$t(\hat{\beta}_0) = \frac{(\hat{\beta}_0 - \beta_0)}{\sqrt{\hat{Var}(\hat{\beta}_0)}} = \frac{\hat{\beta}_0 - \beta_0}{\hat{se}(\hat{\beta}_0)} \sim t[N-2].$$

2. Under the **error normality assumption A9**, the **sample statistics $F(\hat{\beta}_1)$ and $F(\hat{\beta}_0)$ have the F-distribution with 1 numerator degree of freedom and N – 2 denominator degrees of freedom**:

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

$$F(\hat{\beta}_0) = \frac{(\hat{\beta}_0 - \beta_0)^2}{\hat{Var}(\hat{\beta}_0)} \sim F[1, N-2].$$

Note that $\hat{se}(\hat{\beta}_1) = \sqrt{\hat{Var}(\hat{\beta}_1)}$ and $\hat{se}(\hat{\beta}_0) = \sqrt{\hat{Var}(\hat{\beta}_0)}$ are the ***estimated*** **standard errors**, and $\hat{Var}(\hat{\beta}_1)$ and $\hat{Var}(\hat{\beta}_0)$ are the ***estimated*** **variances**, of the OLS coefficient estimators $\hat{\beta}_1$ and $\hat{\beta}_0$, respectively.

3. The **Z-statistics for $\hat{\beta}_0$ and $\hat{\beta}_1$ <u>are</u> <u>not</u> *feasible test statistics***.

$$Z(\hat{\beta}_1) = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\text{Var}(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - \beta_1}{\text{se}(\hat{\beta}_1)} \quad \textit{and} \quad Z(\hat{\beta}_0) = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\text{Var}(\hat{\beta}_0)}} = \frac{\hat{\beta}_0 - \beta_0}{\text{se}(\hat{\beta}_0)}.$$

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 $\text{Var}(\hat{\beta}_0)$ and $\text{Var}(\hat{\beta}_1)$, and of $\text{se}(\hat{\beta}_0)$ and $\text{se}(\hat{\beta}_1)$, are also unknown.

4. The **t-statistics for $\hat{\beta}_0$ and $\hat{\beta}_1$ <u>are</u> *feasible test statistics***.

$$t(\hat{\beta}_0) = \frac{\left(\hat{\beta}_0 - \beta_0\right)}{\sqrt{\hat{\text{Var}}(\hat{\beta}_0)}} = \frac{\left(\hat{\beta}_0 - \beta_0\right)}{\hat{\text{se}}(\hat{\beta}_0)} \quad \textit{and} \quad t(\hat{\beta}_1) = \frac{\left(\hat{\beta}_1 - \beta_1\right)}{\sqrt{\hat{\text{Var}}(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\text{se}}(\hat{\beta}_1)}.$$

They are obtained by replacing the ***unknown* variances** and **standard errors** of the OLS coefficient estimators in the Z-statistics $Z(\hat{\beta}_0)$ and $Z(\hat{\beta}_1)$ with their corresponding ***estimated* variances** $\hat{\text{Var}}(\hat{\beta}_0)$ and $\hat{\text{Var}}(\hat{\beta}_1)$ and ***estimated* standard errors** $\hat{\text{se}}(\hat{\beta}_0) = \sqrt{\hat{\text{Var}}(\hat{\beta}_0)}$ and $\hat{\text{se}}(\hat{\beta}_1) = \sqrt{\hat{\text{Var}}(\hat{\beta}_1)}$.

5. The **F-statistics for $\hat{\beta}_0$ and $\hat{\beta}_1$** also **<u>are</u> *feasible test statistics***.

$$F(\hat{\beta}_0) = \frac{\left(\hat{\beta}_0 - \beta_0\right)^2}{\hat{\text{Var}}(\hat{\beta}_0)} \quad \textit{and} \quad F(\hat{\beta}_1) = \frac{\left(\hat{\beta}_1 - \beta_1\right)^2}{\hat{\text{Var}}(\hat{\beta}_1)}.$$

The denominators of $F(\hat{\beta}_0)$ and $F(\hat{\beta}_1)$ are the ***estimated* variances** $\hat{\text{Var}}(\hat{\beta}_0)$ and $\hat{\text{Var}}(\hat{\beta}_1)$, ***not* the *true* variances** $\text{Var}(\hat{\beta}_0)$ and $\text{Var}(\hat{\beta}_1)$.

❑ Important General Implication of the Normality Assumption A9

The normality assumption A9 permits us **to derive the *functional form* of the sampling distributions of $\hat{\beta}_0$ (normal), $\hat{\beta}_1$ (normal), and $\hat{\sigma}^2$ (chi-square)**.

Knowing the form of the sampling distributions of $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\sigma}^2$ enables us **to derive *feasible* test statistics for the OLS coefficient estimators $\hat{\beta}_0$ and $\hat{\beta}_1$**.

These feasible test statistics for $\hat{\beta}_0$ and $\hat{\beta}_1$ enable us **to conduct *statistical inference*** -- i.e., to derive procedures

(1) for constructing ***confidence intervals* for β_0 and β_1**

and

(2) for performing statistical ***hypothesis tests* about the values of β_0 and β_1**.

5. Distributions Related to the Normal Distribution: A Review

Three important probability distributions are related to the normal distribution:

(1) the **chi-square distribution**;

(2) the **t-distribution**;

(3) the **F-distribution**.

These three distributions are used extensively in constructing confidence intervals and performing hypothesis tests for the regression coefficients β_0 and β_1.

5.1 The Chi-Square Distribution

Definition: A random variable constructed as **the sum of squares of *m* independent standard normal N(0,1) random variables** has the **chi-square distribution with *m* degrees of freedom**.

***Formally*:**

If $Z_1, Z_2, ..., Z_m$ are m **independent N(0,1) random variables** such that

$$Z_i \sim N(0,1) \qquad i = 1, ..., m,$$

then the random variable

$$V = Z_1^2 + Z_2^2 + \cdots + Z_m^2 = \sum_{i=1}^{m} Z_i^2 \sim \chi^2[m]$$

where $\chi^2[m]$ denotes the chi-square distribution with m degrees of freedom.

Properties of the Chi-Square Distribution: The **degrees of freedom** parameter of the chi-square distribution equals the **number of *independent* N(0,1) random variables** that are squared and summed to form the chi-square distributed variable V.

(1) The mean and variance of a chi-square distribution are determined entirely by the value of *m*, the degrees-of-freedom parameter:

$$E(V) = E(\chi^2[m]) = m;$$

$$Var(V) = Var(\chi^2[m]) = 2m.$$

(2) The value of *m* also completely determines the shape of the chi-square distribution.

(3) **Additive (Reproductive) Property of the Chi-Square Distribution:** If V_1, V_2, . . . , V_n are n independent random variables each of which has a chi-square distribution with m_i degrees of freedom such that

$$V_i \sim \chi^2[m_i] \qquad i = 1, ..., n,$$

then the sum of these random variables $V = V_1 + V_2 + \ldots + V_n$ has a chi-square distribution with $k = m_1 + m_2 + \ldots + m_n$ degrees of freedom. That is,

$$V = V_1 + V_2 + \ldots + V_n = \sum_{i=1}^{n} V_i \sim \chi^2[k], \qquad k = \sum_{i=1}^{n} m_i .$$

5.2 The t-Distribution

Definition: A random variable 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*

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

***Formally*:**

If (1) $Z \sim N(0,1)$
(2) $V \sim \chi^2[m]$
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**.

<u>**Properties of the t-Distribution**</u>**:** The **degrees of freedom parameter *m*** completely determines the shape of the t-distribution.

(1) The t-distribution has a mean equal to zero and a variance that is determined completely by the value of m:

$$E(t) = E(t[m]) = 0;$$

$$Var(t) = Var(t[m]) = \frac{m}{m-2}.$$

(2) The t-distribution is **symmetric about its mean** of zero.

(3) The limiting distribution of the t[m]-distribution is the standard normal N(0,1) distribution: that is,

$$\text{As } m \rightarrow \infty, \quad t[m] \rightarrow N(0,1).$$

5.3 The F-Distribution

<u>Definition</u>: A random variable constructed by forming the ratio of two independent chi-square random variables, each of which has been divided by its degrees of freedom, has the F-distribution with specified numerator and denominator degrees of freedom.

***<u>Formally</u>*:**

If (1) $V_1 \sim \chi^2[m_1]$
(2) $V_2 \sim \chi^2[m_2]$
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**.

<u>Properties of the F-Distribution</u>: The values of the **degrees of freedom parameters *m_1* and *m_2*** determine the shape of the F-distribution. The range of the F-distribution is $0 \le F < \infty$.

<u>Relationship Between the t-Distribution and the F-Distribution</u>:

The square of a random variable that has the t-distribution with k degrees of freedom equals a random variable that has the F-distribution with $m_1 = 1$ numerator degrees of freedom and $m_2 = k$ denominator degrees of freedom. That is,

$$\left(t[k]\right)^2 = F[1, k].$$

<u>NOTE</u>: This equality holds ***only*** **for F variables that have *numerator* degrees of freedom equal to 1**.