**ECON 351* -- NOTE 9**

# F-Tests and Analysis of Variance (ANOVA) in the Simple Linear Regression Model

## 1. Introduction

**1.** The simple linear regression model is given by the following **population regression equation**, or **PRE**:

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

This is the simple **Classical Normal Linear Regression Model** (**CNLRM**).

**2.** OLS estimation of the PRE (1) yields 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

$\hat{\beta}_0$ = the OLS estimator of the intercept coefficient $\beta_0$;

$\hat{\beta}_1$ = the OLS estimator of the slope coefficient $\beta_1$;

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ is the i-th estimated (or predicted) value of the dependent variable from the OLS regression, called the **OLS sample regression function** (or **OLS-SRF**);

$\hat{u}_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$ is the i-th OLS residual.

❑ **<u>Formulas</u>:**

$$\hat{\beta}_1 = \frac{\sum_i x_i y_i}{\sum_i x_i^2} = \frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_i (X_i - \overline{X})^2} = \text{ unbiased estimator of } \beta_1;$$

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

$$\hat{\sigma}^2 = \frac{\sum_i \hat{u}_i^2}{N - 2} = \text{ unbiased estimator of } \sigma^2;$$

$$\hat{\text{Var}}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\Sigma_i x_i^2} = \frac{\hat{\sigma}^2}{\Sigma_i \left(X_i - \overline{X}\right)^2} = \text{ unbiased estimator of } \text{Var}(\hat{\beta}_1);$$

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

$$\hat{\text{Var}}(\hat{\beta}_0) = \frac{\hat{\sigma}^2 \sum_i X_i^2}{N \sum_i x_i^2} = \frac{\hat{\sigma}^2 \sum_i X_i^2}{N \sum_i \left(X_i - \overline{X}\right)^2} = \text{ unbiased estimator of } \text{Var}(\hat{\beta}_0);$$

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

## 2. F-Tests of Individual Coefficient Equality Restrictions

***Note 6*** derived both the t-statistic and the F-statistic for $\hat{\beta}_1$. ***Note 8*** explained how to use the t-statistic for $\hat{\beta}_1$ to perform tests of equality restrictions on individual regression coefficients. This section outlines how the **F-statistic for $\hat{\beta}_1$** can be used to perform ***two-tail* tests of equality restrictions on *individual* regression coefficients**.

❑ **The F- and t-statistics for $\hat{\beta}_j$ (j = 0, 1)**

$$F(\hat{\beta}_j) = \frac{\left(\hat{\beta}_j - \beta_j\right)^2}{\hat{Var}(\hat{\beta}_j)} \sim F[1, N-2]; \qquad t(\hat{\beta}_j) = \frac{\hat{\beta}_j - \beta_j}{\hat{se}(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{Var}(\hat{\beta}_j)}} \sim t[N-2].$$

- **Relationship between the test statistics $F(\hat{\beta}_j)$ and $t(\hat{\beta}_j)$:**

$$F(\hat{\beta}_j) = \left(t(\hat{\beta}_j)\right)^2 \qquad \textit{or equivalently} \qquad t(\hat{\beta}_j) = \sqrt{F(\hat{\beta}_j)}\,.$$

  - The **F-statistic for $\hat{\beta}_j$ = the *square* of the t-statistic for $\hat{\beta}_j$**.
  - The **t-statistic for $\hat{\beta}_j$ = the *square root* of the F-statistic for $\hat{\beta}_j$**.

- **Relationship between the t and F distributions:**

  The square of a t-distribution with N–2 degrees of freedom has the F-distribution with 1 numerator degree of freedom and N–2 denominator degrees of freedom.

  $\left(t[N-2]\right)^2$ has the $F[1, N-2]$ distribution: $\left(t[N-2]\right)^2 \sim F[1, N-2]$.

  ***Implications:*** Let α be the chosen significance level for a two-tail hypothesis test. The ***critical values* of the F[1, N–2] and t[N–2] distributions** are related as follows:

$$F_{\alpha}[1, N-2] = \left(t_{\alpha/2}[N-2]\right)^2 \qquad \textit{or} \qquad t_{\alpha/2}[N-2] = \sqrt{F_{\alpha}[1, N-2]}\,.$$

❑ ***Two-tail* F-tests of *equality* restrictions on *individual* regression coefficients**

Like the t-statistic $t(\hat{\beta}_1)$, the F-statistic $F(\hat{\beta}_1)$ can be used to perform ***two-tail*** **tests of *equality* restrictions on *individual* regression coefficients**.

- **Null and alternative hypotheses:**

$H_0$: $\beta_1 = b_1$ *or* $\beta_1 - b_1 = 0$ where $b_1$ is a specified constant

$H_1$: $\beta_1 \neq b_1$ *or* $\beta_1 - b_1 \neq 0$ ⇐ a ***two-sided* alternative hypothesis**
⇐ a ***two-tail* test**.

- **A *feasible test statistic* for $\hat{\beta}_1$ is *either* the F-statistic *or* the t-statistic for $\hat{\beta}_1$:**

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

- **Calculate the *sample value* of the F-statistic $F(\hat{\beta}_1)$ or the t-statistic $t(\hat{\beta}_1)$** under the null hypothesis $H_0$. In the expression for $F(\hat{\beta}_1)$ or $t(\hat{\beta}_1)$, set $\beta_1$ equal to $b_1$, which is the value of $\beta_1$ specified by $H_0$. The resulting sample values of the test statistics $F(\hat{\beta}_1)$ and $t(\hat{\beta}_1)$ under $H_0$ are:

$$F_0(\hat{\beta}_1) = \frac{\left(\hat{\beta}_1 - b_1\right)^2}{\hat{Var}(\hat{\beta}_1)} \qquad \textit{and} \qquad t_0(\hat{\beta}_1) = \frac{\hat{\beta}_1 - b_1}{\hat{se}(\hat{\beta}_1)}.$$

- The ***null distributions*** **of $F_0(\hat{\beta}_1)$ and $t_0(\hat{\beta}_1)$:** If the null hypothesis $H_0$: $\beta_1 = b_1$ is true, then

$$F_0(\hat{\beta}_1) = \frac{\left(\hat{\beta}_1 - b_1\right)^2}{\hat{Var}(\hat{\beta}_1)} \sim F[1, N-2] \text{ under } H_0;$$

$$t_0(\hat{\beta}_1) = \frac{\hat{\beta}_1 - b_1}{\hat{se}(\hat{\beta}_1)} \sim t[N-2] \text{ under } H_0.$$

- **Decision Rule -- Formulation 1**

Let α = the chosen significance level for the test.

$F_{\alpha}[1, N-2]$ = the *α-level critical value* of the $F[1, N-2]$ distribution;
$t_{\alpha/2}[N-2]$ = the *α/2-level critical value* of the $t[N-2]$ distribution.

**At significance level α:**

***Retain* $H_0$: $\beta_1 = b_1$** if $F_0(\hat{\beta}_1) \le F_{\alpha}[1, N-2]$;
if $\left|t_0(\hat{\beta}_1)\right| \le t_{\alpha/2}[N-2]$.

***Reject* $H_0$: $\beta_1 = b_1$** if $F_0(\hat{\beta}_1) > F_{\alpha}[1, N-2]$;
if $\left|t_0(\hat{\beta}_1)\right| > t_{\alpha/2}[N-2]$.

- **Decision Rule -- Formulation 2: the p-value approach**

The **p-value for the calculated F-statistic** $F_0(\hat{\beta}_1) = \Pr(F > F_0)$.
The ***two-tail* p-value for the calculated t-statistic** $t_0(\hat{\beta}_1) = \Pr\left(|t| > |t_0|\right)$.

**At significance level α:**

***Retain* $H_0$: $\beta_1 = b_1$** if p-value for $F_0(\hat{\beta}_1) \ge \alpha$;
if two-tail p-value for $t_0(\hat{\beta}_1) \ge \alpha$.

***Reject* $H_0$: $\beta_1 = b_1$** if p-value for $F_0(\hat{\beta}_1) < \alpha$.
if two-tail p-value for $t_0(\hat{\beta}_1) < \alpha$.

- **Equivalence of the F-test and t-test of $H_0$: $\beta_1 = b_1$ versus $H_1$: $\beta_1 \neq b_1$** follows from two facts:

  **(1)** The **<u>sample values</u> of the two test statistics are related** according to the equality

$$F_0 = (t_0)^2 .$$

Under the null hypothesis $H_0$: $\beta_1 = b_1$, the calculated ***sample value* of the general F-statistic *equals* the *square* of $t_0$**:

$$F_0 = \frac{\left(\hat{\beta}_1 - b_1\right)^2}{\hat{Var}(\hat{\beta}_1)} = \left(\frac{\hat{\beta}_1 - b_1}{\sqrt{\hat{Var}(\hat{\beta}_1)}}\right)^2 = \left(\frac{\hat{\beta}_1 - b_1}{\hat{se}(\hat{\beta}_1)}\right)^2 = (t_0)^2 .$$

  **(2)** The **<u>null distributions</u> of the two test statistics are related** according to a similar equality:

$$\left(t[N-2]\right)^2 \sim F[1, N-2] .$$

i.e., the square of a t[N – 2] distribution has the F[1, N – 2] distribution

***Implication:*** The square of the $\alpha/2$ critical value of the t[N – 2] distribution *equals* the α-level critical value of the F[1, N – 2] distribution; i.e.,

$$\left(t_{\alpha/2}[N-2]\right)^2 = F_\alpha[1, N-2] .$$

## ❑ Usage of t-statistics and F-statistics for testing *individual* coefficient *equality* restrictions

**t-statistics** can be used to perform **both *two-tail* and *one-tail* tests** of equality restrictions on individual regression coefficients.

**F-statistics** can be used ***only*** **for *two-tail* tests** of equality restrictions on individual regression coefficients.

## 3. The ANOVA Table for the OLS SRE

- **<u>The OLS Decomposition Equation</u>:** The Analysis-of-Variance (ANOVA) table for an OLS SRE such as (2) is based on the **OLS decomposition equation**

$$\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} \hat{y}_i^2 + \sum_{i=1}^{N} \hat{u}_i^2 \,. \qquad \textbf{(3)}$$

Each of the **three terms in equation (3)** are defined as follows:

**(1)** $\sum_{i=1}^{N} y_i^2 \equiv \sum_{i=1}^{N} \left(Y_i - \overline{Y}\right)^2 \equiv$ **TSS** $\equiv$ the **Total Sum of Squares**

$=$ the total sample variation of the observed $Y_i$ values.

**(2)** $\sum_{i=1}^{N} \hat{y}_i^2 \equiv \sum_{i=1}^{N} \left(\hat{Y}_i - \overline{Y}\right)^2 \equiv$ **ESS** $\equiv$ the **Explained Sum of Squares**

$=$ the sum of squares explained by the sample regression function, i.e., by the regressor X.

**(3)** $\sum_{i=1}^{N} \hat{u}_i^2 \equiv \sum_{i=1}^{N} \left(Y_i - \hat{Y}_i\right)^2 \equiv$ **RSS** $\equiv$ the **Residual Sum of Squares**

$=$ the unexplained variation of the observed sample values $Y_i$ of the regressand Y around the sample regression line

- **The General ANOVA Table:** The **Analysis-of-Variance (ANOVA) table** for an OLS SRE takes the following general form.

| Source of variation | SS | df | MSS = SS/df |
|---|---|---|---|
| The regression function (explained) | $ESS = \sum_{i=1}^{N} \hat{y}_i^2$ | K – 1 | $\frac{ESS}{K-1} = \frac{\sum_i \hat{y}_i^2}{K-1}$ |
| The residuals (unexplained) | $RSS = \sum_{i=1}^{N} \hat{u}_i^2$ | N – K | $\frac{RSS}{N-K} = \frac{\sum_i \hat{u}_i^2}{N-K}$ |
| Total sample variation of $Y_i$ | $TSS = \sum_{i=1}^{N} y_i^2$ | N – 1 | |

***Definitions:***

**K** ≡ the **total number** of estimated **regression coefficients** in the OLS-SRE.

Thus, **K – 1** = the **number** of estimated ***slope*** **coefficients** in the OLS-SRE.

- **The ANOVA Table for a Simple OLS-SRE:** The **Analysis-of-Variance (ANOVA) table** for a simple OLS-SRE such as equation (2) takes the following form, where K = 2 and hence K – 1 = 2 – 1 = 1.

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

| Source of variation | SS | df | MSS = SS/df |
|---|---|---|---|
| The regression function (explained) | $ESS = \sum_{i=1}^{N} \hat{y}_i^2 = \hat{\beta}_1^2 \sum_{i=1}^{N} x_i^2$ | 1 | $\frac{ESS}{1} = \frac{\sum_i \hat{y}_i^2}{1}$ |
| The residuals (unexplained) | $RSS = \sum_{i=1}^{N} \hat{u}_i^2$ | N – 2 | $\frac{RSS}{N-2} = \frac{\sum_i \hat{u}_i^2}{N-2}$ |
| Total sample variation of $Y_i$ | $TSS = \sum_{i=1}^{N} y_i^2$ | N – 1 | |

***Note:*** $ESS \equiv \sum_{i=1}^{N} \hat{y}_i^2 = \hat{\beta}_1^2 \sum_{i=1}^{N} x_i^2$ where $x_i \equiv X_i - \overline{X}$ (i = 1, ..., N).

## 4. The F-Statistic for the ANOVA Table: Simple Regression

❑ **<u>Form of the ANOVA F-Statistic</u>:** For a simple linear regression model such as equation (1) and its corresponding OLS-SRE (2)

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

the ANOVA F-statistic is defined as the **ratio** of **(1)** the **MSS for the regression function** to **(2)** the **MSS for the residuals**:

(1) the MSS for the OLS sample regression function $= \dfrac{ESS}{K-1} = \dfrac{ESS}{1}$

(2) the MSS for the OLS residuals $= \dfrac{RSS}{N-K} = \dfrac{RSS}{N-2}$

The **ratio of (1) to (2)** is the **ANOVA F-statistic**.

- For the general linear regression model with K regression coefficients:

$$\text{ANOVA–F}_0 = \frac{ESS/K-1}{RSS/N-K}$$

- For the simple linear regression model with K = 2 regression coefficients:

$$\text{ANOVA–F}_0 = \frac{ESS/1}{RSS/N-2}$$

❑ **<u>Alternative Formula for the ANOVA F-Statistic</u>:** For a simple linear regression model such as equation (1) and its corresponding OLS-SRE (2), the ANOVA F-statistic can be written alternatively in terms of the OLS slope coefficient estimator $\hat{\beta}_1$ and its estimated variance $\hat{Var}(\hat{\beta}_1)$:

$$\text{ANOVA–}F_0 = \frac{ESS/1}{RSS/(N-2)}$$

$$= \frac{\sum_{i=1}^{N} \hat{y}_i^2 \Big/ 1}{\sum_{i=1}^{N} \hat{u}_i^2 \Big/ (N-2)}$$

$$= \frac{\sum_{i=1}^{N} \hat{y}_i^2}{\hat{\sigma}^2} \qquad \text{since } \frac{\sum_{i=1}^{N} \hat{u}_i^2}{(N-2)} = \hat{\sigma}^2$$

$$= \frac{\hat{\beta}_1^2 \sum_{i=1}^{N} x_i^2}{\hat{\sigma}^2} \qquad \text{since } \sum_{i=1}^{N} \hat{y}_i^2 = \hat{\beta}_1^2 \sum_{i=1}^{N} x_i^2$$

$$= \frac{\hat{\beta}_1^2}{\hat{\sigma}^2 \Big/ \sum_{i=1}^{N} x_i^2} \qquad \text{dividing by } \sum_{i=1}^{N} x_i^2$$

$$= \frac{\hat{\beta}_1^2}{\hat{Var}(\hat{\beta}_1)} \qquad \text{since } \frac{\hat{\sigma}^2}{\sum_{i=1}^{N} x_i^2} = \hat{Var}(\hat{\beta}_1).$$

- **<u>Distribution of the ANOVA F-Statistic</u>: Under the null hypothesis $H_0$: $\beta_1 = 0$** – that is if the null hypothesis **$H_0$: $\beta_1 = 0$ is *true*** – the ANOVA $F_0$-statistic has the **F-distribution with *numerator* degrees-of-freedom = 1 and *denominator* degrees-of-freedom = (N−2)**.

That is, if the null hypothesis **$H_0$: $\beta_1 = 0$ is *true***,

$$\text{ANOVA–}F_0 = \frac{ESS/1}{RSS/(N-2)} = \frac{\sum_{i=1}^{N}\hat{y}_i^2 \Big/ 1}{\sum_{i=1}^{N}\hat{u}_i^2 \Big/ (N-2)} = \frac{\sum_{i=1}^{N}\hat{y}_i^2}{\hat{\sigma}^2} \sim F[1, N-2].$$

- **ANOVA F-Test of $H_0$: $\beta_1 = 0$ against $H_1$: $\beta_1 \neq 0$:**

$H_0$: $\beta_1 = 0$

$H_1$: $\beta_1 \neq 0$ ⇐ a ***two-sided* alternative hypothesis**

1. The **calculated *sample value* of the ANOVA F-statistic** under the null hypothesis **$H_0$: $\beta_1 = 0$** is:

$$\text{ANOVA–}F_0 = \frac{ESS/1}{RSS/(N-2)} = \frac{\sum_{i=1}^{N}\hat{y}_i^2 \Big/ 1}{\sum_{i=1}^{N}\hat{u}_i^2 \Big/ (N-2)} = \frac{\sum_{i=1}^{N}\hat{y}_i^2}{\hat{\sigma}^2}.$$

2. **Decision Rule**

Let $F_\alpha[1, N-2]$ = the ***α-level critical value*** of the $F[1, N-2]$ distribution.

**At significance level α:**

***Retain* $H_0$: $\beta_1 = 0$** if $\text{ANOVA–}F_0 \le F_\alpha[1, N-2]$;
if p-value for ANOVA–$F_0 \ge \alpha$.

***Reject* $H_0$: $\beta_1 = 0$** if $\text{ANOVA–}F_0 > F_\alpha[1, N-2]$;
if p-value for ANOVA–$F_0 < \alpha$.

- **Equivalence of ANOVA F-Test and General F-Test of $H_0$: $\beta_1 = 0$ against $H_1$: $\beta_1 \neq 0$:**

- **Applies *only* in the case of the *simple* linear regression model** that has just one slope coefficient and one regressor.

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

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

- Under the null hypothesis $H_0$: $\beta_1 = 0$, the ***sample value* of the general F-statistic $F(\hat{\beta}_1)$** is:

$$F_0(\hat{\beta}_1) = \frac{\hat{\beta}_1^2}{\hat{Var}(\hat{\beta}_1)} \sim F[1, N-2] \quad \text{under } H_0\text{: } \beta_1 = 0.$$

- For the ***simple* linear regression model**, the **ANOVA F-statistic *equals* the general F-statistic for $\hat{\beta}_1$ under the null hypothesis $H_0$: $\beta_1 = 0$**. That is, only for the simple linear regression model is it true that

$$\text{ANOVA–}F_0 = \frac{ESS/1}{RSS/(N-2)} = \frac{\hat{\beta}_1^2}{\hat{Var}(\hat{\beta}_1)} = F_0(\hat{\beta}_1 : \beta_1 = 0).$$