ANOVA for S.L.R Model

The total deviation $Y_i - \overline{Y}$ can be viewed as the sum of two components: -

1. The deviation of the estimated value $\widehat{Y}_i$ from the mean $\overline{Y}$
2. The deviation of the observed value Y_i from the estimated regression line, i.e.: -

$$(Y_i - \overline{Y}) = (\widehat{Y}_i - \overline{Y}) + (Y_i - \widehat{Y}_i)$$

$$(Y_i - \overline{Y})^2 = [(\widehat{Y}_i - \overline{Y}) + (Y_i - \widehat{Y}_i)]^2$$

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n}[(\widehat{Y}_i - \overline{Y}) + (Y_i - \widehat{Y}_i)]^2$$

$$= \sum_{i=1}^{n}[(\widehat{Y}_i - \overline{Y})^2 + 2(\widehat{Y}_i - \overline{Y})(Y_i - \widehat{Y}_i) + (Y_i - \widehat{Y}_i)^2]$$

$$= \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2 + 2\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})(Y_i - \widehat{Y}_i) + \sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2$$

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2 + \sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2$$

$\rightarrow S_{yy}$ where $\sum(Y_i - \overline{Y})^2$ sum of squares about the mean

see $\rightarrow$ where $\sum(\widehat{Y}_i - \overline{Y})^2$ sum of squares Due to Regression

see $\rightarrow$ where $\sum(Y_i - \widehat{Y})^2$ sum of squares about the Regression

$$\sum(Y_i - \overline{Y})^2 = \sum(\widehat{Y}_i - \overline{Y})^2 + \sum(Y_i - \widehat{Y}_i)^2$$

$S_{st} = S_{yy} = SSR(X_1) + SSe$

Where :-

$$\sum(Y_i - \overline{Y})^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n}$$

$$\sum(Y_i - \overline{Y})^2 = \sum(\overline{Y} + \hat{\beta}_1(x_i - \overline{x}) - \overline{Y})^2$$

$$= \hat{\beta}_1^2 \sum (x_i - \bar{x})^2 = \hat{\beta}_1^2\, S_{xx} = \hat{\beta}_1\, S_{xx}$$

$$\sum \left(Y_i - \bar{Y}\right)^2 = SSe = S_{yy} - \hat{\beta}_1\, S_{xx}$$

ANOVA TABLE

$H_0 : \beta_1 = 100$
$H_1 : \beta_1 \neq 100$

S.O.V.	D.F.	SS	MS	Cal. F
Due to Regression R(x_1)	1	$SS_R(X_1) = \hat{\beta}_1 S_{xy}$	$\frac{SS_R(X_1)}{1 = d.f.}$	$\frac{MSR}{MSE}$
About Regression or (Error)	n-2	$SS_e = SS_T - SS_R(X_1)$ $= S_{yy} - \hat{\beta}_1^2\, S_{xx}$	$\frac{SS_e}{n-2} = Mse = \hat{\delta}^2$	
About the Mean (Total)	n-1	$SS_T = S_{yy}$		

The null hypothesis is rejected in the case of:

$Cal.\,F \geq tab.\,F = F(\propto, v_1, v_2)$

$v_1 = d.f.\,R(x_1) \qquad v_1 = d.f.\,Error$

Equivalence between F-test and t-test:-

In simple linear regression and at a significance level of (α), the F-test for the hypothesis $H_0 : \beta_1 = 0$ and $H_1 : \beta_1 \neq 0$ is equivalent to a two-tailed t-test as well, and this can be explained as follows:

$SSR = \hat{\beta}_1^2\, S_{xx}$

$$S^2\left(\hat{\beta}_1\right) = \frac{\hat{\delta}^2}{S_{xx}} \Longrightarrow \hat{\delta}^2 = S^2(\hat{\beta}_1) S_{xx} = MSe$$

$$\therefore Cal.\,F = \frac{MSR}{MSE} = \frac{SSR/1}{SSE/n-2} = \frac{\hat{\beta}_1^2\, S_{xx}}{S^2\left(\hat{\beta}_1\right) S_{xx}} = \left[\frac{\hat{\beta}_1}{S(\hat{\beta}_1)}\right]^2$$

For the same hypothesis and using the t-test we have:

$$Cal.\,t = \frac{\hat{\beta}_1 - 0}{S(\hat{\beta}_1)} = \frac{\hat{\beta}_1}{S(\hat{\beta}_1)}$$

$$\therefore Cal.\,F = [Cal.\,t]^2$$

Also, the tabulated value of both F and t are algebraically equivalent, i.e.:

$$tab.\mathrm{F}(\propto, 1, \mathrm{n}-2) = \mathrm{t}^2\left(\frac{\propto}{2}, \mathrm{n}-2\right)$$

The F and t tests can be used to test the previous hypothesis.

The superiority of the t test over the f test:

This is shown by the following points:

1. It is more flexible than the F-test because it is used to test one-tailed hypotheses:

$\mathrm{H}_0 : \beta_1 \leq 0$
$\mathrm{H}_1 : \beta_1 > 0$
او
$\mathrm{H}_0 : \beta_1 \geq 0$
$\mathrm{H}_1 : \beta_1 < 0$

The F-test is not valid in this case.

2. The t-test can be used to test the hypothesis $H_0 : \beta_1 = \beta_{10}$ *and* $H_1 : \beta_1 \neq \beta_{10}$.
3. The t-test can be used to find the confidence interval for β_1
4. The t-test can be used to test $\mathrm{H}_0 : \beta_0 = 0$ and $\mathrm{H}_1 : \beta_1 \neq 0$

Coefficient of Determination

We have SST=SSR+SSE

$$(\mathrm{SSR} = 0 \text{ iff } \mathrm{SST} = \mathrm{SSE}\,)\, 0 \leq \mathrm{SSR} \leq \mathrm{SST}(= \mathrm{SSR} \text{ iff } \mathrm{SSE} = 0)$$
$$\hat{\beta}_1 = 0 \;\&\; \hat{Y}_i = \bar{Y} \qquad\qquad \hat{Y}_i = Y_i$$

In the estimation process, we always try to reduce the SSE value as much as possible. By dividing the above inequality by SST, we get:

$$0 \leq \frac{\mathrm{SSR}}{\mathrm{SST}} \leq 1$$

That is: -

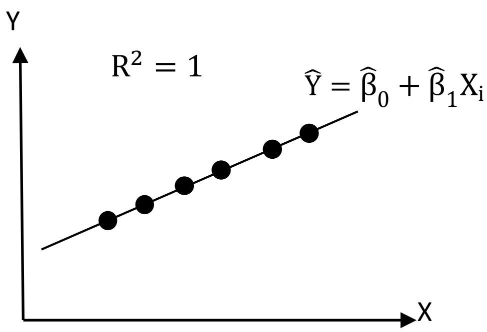
$$\text{if } \mathrm{R}^2 = \frac{\mathrm{SSR}}{\mathrm{SST}} \qquad 0 \leq \mathrm{R}^2 \leq 1$$

R^2 is called the coefficient of determination and represents the percentage of total deviations explained or explained by the estimated regression equation.

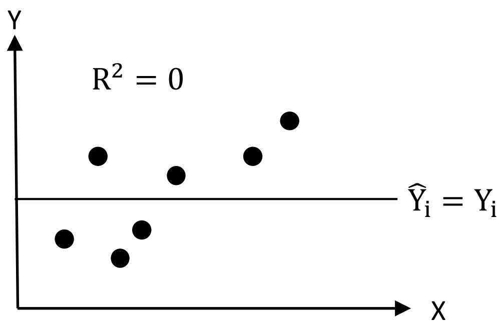
Or the percentage of the estimated regression equation's contribution to explaining or interpreting the total deviations in the values of Y around the arithmetic mean $\overline{Y}$.

Or it is the amount that the regression equation explains through X_i (the independent variable) of the changes that occur in Y.

As we mentioned, the value of R^2 ranges between zero and one, as follows:



We notice from the figure that all values lie on the regression line, i.e. $\widehat{Y}_i = Y_i$, and thus SSR = SST $\Longleftarrow$ SSE = 0, and $R^2 = 1$



If $\widehat{\beta}_1 = 0 \Longleftarrow R^2 = 0$ then the regression line is applied to $\overline{Y}$, i.e. $SSR = 0 \Longleftarrow \widehat{Y}_i = Y_i$ and in this case there is no relationship between the dependent variable Y and the independent variable X in the sample data.

It is worth noting that the value of R^2 is not necessarily sufficient evidence of the type of relationship between X and Y, as is clear from the following figures: