Example:

Test the following hypothesis using the data from the previous example:

$H_0 : \beta_0 = 100$

$H_1 : \beta_0 \neq 100$

Solution:

Determine the table value as follows:

$t(0.025, 8) = 2.306$

Calculate the statistical test t as follows:

$$t_{cal.} = \frac{\hat{\beta}_0 - \beta_{00}}{S_{\hat{\beta}_0}}$$

$$= \frac{85.043 - 100}{9.97}$$

$t = -1.5$

The calculated t value is compared with the table value as follows:

$|t_{cal.}| = |-1.5| < t(0.025, 8) = 2.306$

Since the calculated t value is less than the table value, the null hypothesis is accepted and the alternative hypothesis H_1 is rejected, i.e. $\beta_0 = 100$.

B- Testing the hypotheses related to the regression coefficient β_1

There are two methods:

First: Using the statistical test t:

1- Test the following hypothesis:

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

$$t_{cal.} = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}}$$

Whereas if the calculated t is greater than the tabular t X, we reject the null hypothesis and accept the alternative hypothesis.

As for the hypothesis that $\beta_1$equals a certain value, for example β_{00}, the hypothesis will be as follows:

$$H_0 : \beta_1 = \beta_{00}$$
$$H_1 : \beta_1 \neq \beta_{00}$$

The statistical test t will be shown in the following formula:

$$t_{cal.} = \frac{\hat{\beta}_1 - \beta_{00}}{S_{\hat{\beta}_1}}$$

If $|t_{cal.}| > t(\frac{\alpha}{2}, n-2)$ rejects the null hypothesis H_0 and accepts the alternative hypothesis H_1.

<u>Note</u>: If we accept the null hypothesis that states:

$$H_0 : \beta_1 = 0$$

This indicates that there is no linear relationship between x and y. This may result from two reasons:

1- Either X does not explain the changes that occur in y for any value of x, as shown in Figure (2).
2- Either the relationship between x and y is not a linear relationship, i.e. it is a non-linear relationship, and thus the drawing is in Figure (3).

But if we reject the null hypothesis $H_0 : \beta_1 = 0$.
This indicates that X is important in explaining the differences with y, which means that the relationship between X and y is represented by a straight line, meaning that the linear model represents the data well, as in Figure (4).

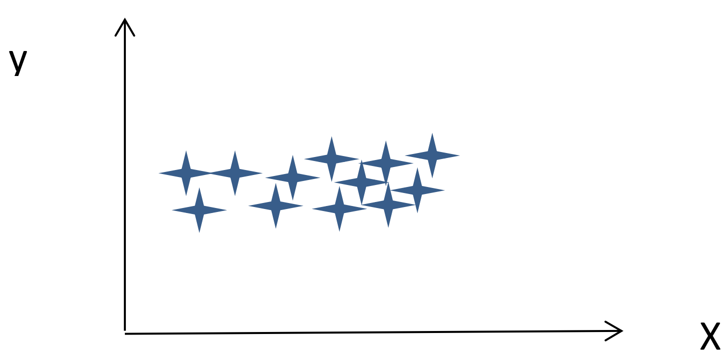

Figure (2) shows that the variable X does not explain the changes that occur in Y.

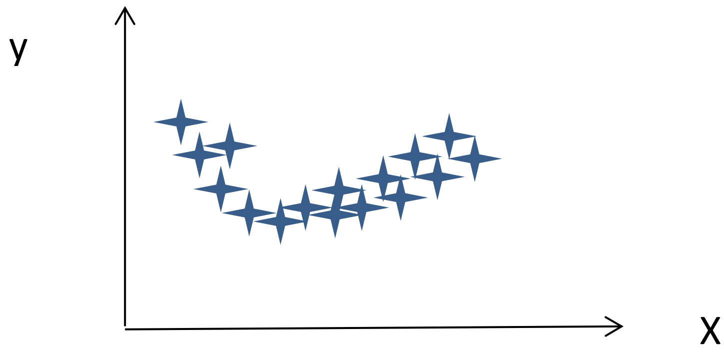

Figure (3) shows the relationship between x and y. It is a non-linear relationship.

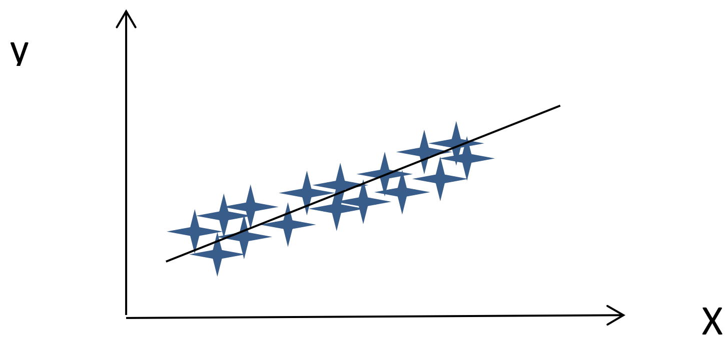

Figure (4) shows the relationship between x and y represented by a straight line.

Second: Using the statistical laboratory, i.e. the analysis of variance table, as follows:

Analysis of variance

The deviation of the original values y_i from the arithmetic mean $\bar{y}$ consists of two elements:

$$(y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$$

Taking the square of both sides, we get:

$$(y_i - \bar{y})^2 = [(\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)]^2$$

Taking the sum of both sides, we get:

$$\sum(y_i-\bar{y})^2=\sum[(\hat{y}_i-\bar{y})+(y_i-\hat{y}_i)]^2$$

Open the square bracket on the right side:

$$\sum(y_i-\bar{y})^2=\sum[(\hat{y}_i-\bar{y})^2+2(\hat{y}_i-\bar{y})(y_i-\hat{y}_i)+(y_i-\hat{y}_i)^2]$$

By entering the sum in the parentheses on the right side:

$$\sum(y_i-\bar{y})^2=\sum(\hat{y}_i-\bar{y})^2+2\sum(\hat{y}_i-\bar{y})(y_i-\hat{y}_i)+\sum(y_i-\hat{y}_i)^2 \qquad \ldots \quad (*)$$

The second term on the right-hand side can be proven to be equal to zero as follows:

$$\sum(\hat{y}_i-\bar{y})(y_i-\hat{y}_i)=\sum(\hat{y}_i y_i-\bar{y}y_i-\hat{y}_i^2+\bar{y}\hat{y}_i)$$

By entering the sum in parentheses:

$$\sum(\hat{y}_i-\bar{y})(y_i-\hat{y}_i)=\sum\hat{y}_i y_i-\bar{y}\sum y_i-\sum\hat{y}_i^2+\bar{y}\sum\hat{y}_i$$

$$\boxed{\begin{aligned}\sum\hat{y}_i y_i &= \sum\hat{y}_i y_i\\ &=\sum\hat{y}_i(\beta_0+\beta_1 X_i+e_i)\\ &=\sum\hat{y}_i(\hat{y}_i+e_i)\\ &=\sum\hat{y}_i\hat{y}_i+\sum\hat{y}_i e_i\\ &=\sum\hat{y}_i^2+\sum\hat{y}_i e_i\end{aligned}}$$

Thus:

$$\sum(\hat{y}_i-\bar{y})(y_i-\hat{y}_i)=\sum\hat{y}_i^2+\sum\hat{y}_i e_i-\bar{y}\sum y_i-\sum\hat{y}_i^2+\bar{y}\sum\hat{y}_i$$

In short, we get:

$$\sum(\hat{y}_i-\bar{y})(y_i-\hat{y}_i)=\cancel{\sum\hat{y}_i^2}+\sum\hat{y}_i e_i-\bar{y}\cancel{\sum}y_i-\cancel{\sum\hat{y}_i^2}+\bar{y}\sum\cancel{\hat{y}_i}$$

$$\boxed{\sum\hat{y}_i e_i=0}$$

Thus, the result will be equal to zero, i.e.:

$$\sum(\hat{y}_i - \bar{y})(y_i - \hat{y}_i) = 0$$

Thus, the equation (*) will be as follows:

$$\sum(y_i - \bar{y})^2 = \sum(\hat{y}_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2$$

We have $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ so:

$$\sum(y_i - \bar{y})^2 = \sum(\hat{\beta}_0 + \hat{\beta}_1 X_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2$$

We have $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{X}$ so:

$$\sum(y_i - \bar{y})^2 = \sum(\bar{y} - \hat{\beta}_1 \bar{X} + \hat{\beta}_1 X_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2$$

Taking $\hat{\beta}_1$ as a common factor we get:

$$\sum(y_i - \bar{y})^2 = \sum(\bar{y} + \hat{\beta}_1(X_i - \bar{X}) - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2$$

By entering the sum on the first bracket:

$$\sum(y_i - \bar{y})^2 = \hat{\beta}_1^2 \sum(X_i - \bar{X})^2 + \sum(y_i - \hat{y}_i)^2$$

According to the symbols we have, we get:

$$S_{yy} = \hat{\beta}_1^2 S_{XX} + SSe$$

$$\therefore SSe = S_{yy} - \hat{\beta}_1^2 S_{XX}$$

Thus, the ANOVA table can be formed as follows:

Source of variation مصادر التباين S.O.V	Degree of Fredom درجات الحرية	Sum of squares مجموع المربعات	Mean of squares متوسط المربعات	Fcal. قيمة F المحسوبة
Due to Regression مصدر يعود الى الانحدار	1 عدد المتغيرات المستقلة في النموذج	$SSR(X_1)=\hat{\beta}_1^2 S_{XX}$ $Or\ \ SSR(X_1)=\hat{\beta}_1 S_{Xy}$	$M.S.R(X_1)=\dfrac{SSR(X_1)}{d.f.R}$	$F_{cal.}=\dfrac{M.S.R}{M.S.e}$
About Regression Error مصدر يعود للخطأ	n-2 n مطروح منها عدد المعلمات في النموذج	$SSe=SST-SSR(X_1)$ $SSe=S_{yy}-\hat{\beta}_1^2 S_{XX}$ $or\ \ SSe=S_{yy}-\hat{\beta}_1 S_{Xy}$	$M.S.e=\dfrac{SSe}{d.f.e}$	
Total المجموع	n-1	$SST=S_{yy}$		

The tabular F value is extracted from the F distribution tables by knowing the degree of freedom of the source of the regression, the degree of freedom of the error, and the level of significance. This can be expressed in symbolic form, i.e.:

$$F_{tab.}=F(\alpha,v_1,v_2)$$

In the case of simple linear regression, which consists of one independent variable, the tabular F value is as follows:

$$F_{tab.}=F(\alpha,1,n-2)$$

The ANOVA table tests the same hypothesis that the t-test tests about the parameter, which is:

$$H_0: \beta_1 = 0$$
$$H_1: \beta_1 \neq 0$$