Analysis of variance table

If the linear model consists of several independent variables or m independent variables that are in the form of matrices, the analysis of variance table in this case is as follows:

S.O.V	d.f	S.S	M.S	Fcal.
$R(X_1, X_2, ..., X_m)$	m	$\hat{\beta}'X'y - n\bar{y}^2$	$M.S.R$	$\frac{M.S.R}{M.S.e}$
$Error\,(X_1, X_2, ..., X_m)$	$n-m-1$	$SST - SSR$	$M.S.e$	
$Total$	$n-1$	$y'y - n\bar{y}^2$		

The hypothesis to be tested is:

$$H_0 : \beta_1 = \beta_2 = ... = \beta_m = 0$$
$$H_1 : \beta_1 \neq \beta_2 \neq ... \neq \beta_m \neq 0$$

Interpretation of the null hypothesis: There is no variable that explains the changes occurring in y.
Interpretation of the alternative hypothesis: At least one variable is important in explaining the changes occurring in y.

Example: Using the data from the previous example, create an analysis of variance table to test the following hypothesis:

$$H_0 : \beta_1 = \beta_2 = 0$$
$$H_1 : \beta_1 \neq \beta_2 \neq 0$$

X_2	X_1	y
1	3	6
2	5	5
5	4	8
7	9	7
9	2	6
8	1	2

Solution: First, we write the analysis of variance table with its mathematical formulas, as follows:

S.O.V	d.f	S.S	M.S	Fcal.
$R(X_1, X_2, ..., X_m)$	m	$\hat{\beta}'X'y - n\bar{y}^2$	$M.S.R$	$\frac{M.S.R}{M.S.e}$
$Error\,(X_1, X_2, ..., X_m)$	$n-m-1$	$SST - SSR$	$M.S.e$	
$Total$	$n-1$	$y'y - n\bar{y}^2$		

The regression sum of squares is calculated as follows:

$SSR(X_1X_2) = \hat{\beta}'X'y - n\bar{y}^2$

We find the product of multiplying the variables together as follows:

y^2	X_2y	X_1y	X_2^2	X_1^2	X_1X_2	X_2	X_1	y	المشاهدة
36	6	18	1	9	3	1	3	6	1
25	10	25	4	25	10	2	5	5	2
64	40	32	25	16	20	5	4	8	3
49	49	63	49	81	63	7	9	7	4
36	54	12	81	4	18	9	2	6	5
4	16	2	64	1	8	8	1	2	6
214	175	152	224	136	122	32	24	34	المجموع

$$SSR(X_1X_2) = \hat{\beta}'X'y = [\hat{\beta}_0 \quad \hat{\beta}_1 \quad \hat{\beta}_2]\begin{bmatrix} \sum y_i \\ \sum X_i y_i \\ \sum X_i y_i \end{bmatrix} - n\bar{y}^2$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{34}{6} = 5.66$$

$$SSR(X_1X_2) = \hat{\beta}'X'y = [4.5117 \quad 0.387 \quad -0.075]\begin{bmatrix} 34 \\ 152 \\ 175 \end{bmatrix} - (6)(5.66)^2$$

$$= 6.7017$$

$SST = y'y - n\bar{y}^2$

whereas:

$y'y = \sum y_i^2 = 214$

Thus, the total sum of squares will be:

$$SST = y'y - n\bar{y}^2$$
$$= 214 - (6)(5.66)^2$$
$$= 21.34$$

The sum of squares of the error will be:

$$SSe = SST - SSR(X_1X_2)$$
$$= 21.34 - 6.7017$$
$$= 14.6383$$

Thus, the analysis of variance table is as follows:

S.O.V	d.f	S.S	M.S	$F_{cal.}$
$R(X_1, X_2)$	$m = 2$	6.7017	3.35	0.6866
$Error\,(X_1, X_2)$	$n-m-1$ $6-2-1=3$	14.638	4.879	
$Total$	$n-1$ $6-1=5$	21.34		

Thus, the calculated F value was found to be 0.6866. The tabular F value was:

$F(0.05, 2, 3) = 9.55$

Comparing the two values we get:

$F_{cal.} = 0.6866 < F(0.05, 2, 3) = 9.55$

So, the null hypothesis is accepted and the alternative hypothesis is rejected, that is, both variables X_1X_2 do not affect the variable y.

Multiple correlation coefficient

The multiple correlation coefficient is a measure of the relationship between the observed y values and the expected y values, i.e. $\hat{y}$, and it is symbolized by R, meaning:

$$R = r_{y\hat{y}} = \frac{S_{y\hat{y}}}{\sqrt{S_{yy}S_{\hat{y}\hat{y}}}} = \frac{\sum(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum(y_i - \bar{y})^2 \sum(\hat{y}_i - \bar{\hat{y}})^2}}$$

The correlation coefficient between y and $\hat{y}$ is a simple correlation coefficient between y and $\hat{y}$, but it is called multiple because $\hat{y}$ came from several values of X.

The square of the multiple correlation coefficient is the same as the coefficient of determination R^2, where R^2 takes the following formula:

$$R^2 = \frac{SS\ due\ to\ Regression}{SS\ Total}$$

It can be proven that $R^2 = r_{y\hat{y}}^2$ as follows:

$$r_{y\hat{y}}^2 = \left[\frac{S_{y\hat{y}}}{\sqrt{S_{yy}S_{\hat{y}\hat{y}}}}\right]^2 = \frac{S_{y\hat{y}}^2}{S_{yy}S_{\hat{y}\hat{y}}}$$

Through the numerator:

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

Since $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$, then:

$$S_{y\hat{y}}^2 = \left[\sum(y_i - \bar{y})\left(\hat{\beta}_0 + \hat{\beta}_1 X_i - \bar{\hat{y}}\right)\right]^2$$

Since $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{X}$ then:

$$S_{y\hat{y}}^2 = \left[\sum(y_i - \bar{y})\left(\bar{y} - \hat{\beta}_1 \bar{X} + \hat{\beta}_1 X_i - \bar{\hat{y}}\right)\right]^2$$