The analysis of variance table for regression analysis.

A test on whether the regression is significant (or if the regression equation differs from a constant) is given by

$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$ against H_1: at least one β_i, $i = 1, 2, \ldots, k$ is different from 0.

With $\underline{\varepsilon} \sim N(\underline{0}, \sigma^2 \underline{I})$, we know that $\frac{SS_E}{\sigma^2} \sim \chi^2(n-k-1)$. Under H_0 we also have that $\frac{SS_R}{\sigma^2}$ is $\chi^2(k)$ and since they are independent $\frac{SS_R}{k} \Big/ \frac{SS_E}{n-k-1} \sim F_{k, n-k-1}$. Since the expected SS_R denominator is independent of H_0 and $(E[SS_R])$ is increasing if H_0 is not true, we reject H_0 if $f \geq f_{\alpha, k, n-(k+1)}$ where f is the observed value of $F_{k, n-k-1}$

It is normal to collect the sum of squares values from a regression analysis in a table as follows.

The analysis of variance table for a regression analysis,

Source	Sum of squares	DF	Mean sum of squares	F
Regression	$SS_R = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$	k	$MSS_R = \frac{SS_R}{k}$	$\frac{SS_R/k}{SS_E/n-(k+1)}$
Error	$SS_E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$	$n-(k+1)$	$MSS_E = \frac{SS_E}{n-(k+1)}$	
Total	$SS_T = \sum_{i=1}^{n} (y_i - \bar{y})^2$	$n-1$		

The partial F-test

The partial F-test is useful if we want to test the significance of a group of regression variables.

Assume we want to test the following hypothesis

$H_0: \beta_{r+1}, \beta_{r+2}, \dots, \beta_k = 0$ $\quad H_1:$ at least one of $\beta_{r+1}, \dots, \beta_k \neq 0$

Let $X = [\underline{1}, \underline{x}_1, \underline{x}_2, \dots, \underline{x}_k]$ and $X_1 = [\underline{1}, \underline{x}_1, \underline{x}_2, \dots, \underline{x}_r]$

and define $\underline{H} = \underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T$ and $\underline{H}_1 = \underline{X}_1(\underline{X}_1^T\underline{X}_1)^{-1}\underline{X}_1^T$ and thereby idempotent and symmetric.

We get for all $\underline{y}$:

$$\underline{H}\,\underline{H}_1\,\underline{y} = \underline{H}\,\underline{X}_1\underline{b}_1 = \underline{H}\,\underline{X}\begin{bmatrix} b_1 \\ 0 \end{bmatrix} = \underline{X}\begin{bmatrix} b_1 \\ 0 \end{bmatrix} = \underline{X}_1\underline{b}_1$$

Thereby $\underline{H}\,\underline{H}_1 = \underline{H}_1$ and by transposing we get $(\underline{H}\,\underline{H}_1)^T = \underline{H}_1\underline{H} = \underline{H}_1$

It follows that $(\underline{H} - \underline{H}_1)^2 = \underline{H} - \underline{H}\,\underline{H}_1 - \underline{H}_1\underline{H} + \underline{H}_1^2 = \underline{H} - \underline{H}_1$ i.e

$\underline{H} - \underline{H}_1$ is idempotent.

Let $\underline{\beta} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_n \end{bmatrix}$ and $\underline{\beta}_1 = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_r \end{bmatrix}$

$$SS_R(\underline{\beta}) - SS_R(\underline{\beta}_1) = SS_T - \underline{y}^T(\underline{I} - \underline{H})\underline{y} - SS_T + \underline{y}^T(\underline{I} - \underline{H}_1)\underline{y}$$

$$= \underline{y}^T(\underline{H} - \underline{H}_1)\underline{y} = (\underline{y} - \underline{X}_1\underline{\beta}_1)^T(\underline{H} - \underline{H}_1)(\underline{y} - \underline{X}_1\underline{\beta}_1)$$

since $(\underline{H} - \underline{H}_1)\underline{X}_1\underline{\beta}_1 = \underline{X}_1\underline{\beta}_1 - \underline{X}_1\underline{\beta}_1$

Rank$(\underline{H} - \underline{H}_1) = tr(\underline{H}) - tr(\underline{H}_1) = k + 1 - (r + 1) = k - r$ and

$$\frac{SS_R(\underline{\beta}) - SS_R(\underline{\beta}_1)}{\sigma^2} = \frac{(\underline{y} - \underline{X}_1\underline{\beta}_1)^T(\underline{H} - \underline{H}_1)(\underline{y} - \underline{X}_1\underline{\beta}_1)}{\sigma^2} \sim \chi^2(k - r)$$

Also $(\underline{H}-\underline{H}_1)(\underline{I}-\underline{H}) = H - H_1 - H + H_1 = \underline{0}$

and $\underline{Y}^T(\underline{H}-\underline{H}_1)\underline{Y}$ and $\hat{\sigma}^2$ are independent.

The test statistic for H_0 is $F = \dfrac{\dfrac{SS_R(\underline{\beta}) - SS_R(\underline{\beta}_1)}{k-r}}{\dfrac{SS_E(\underline{\beta})}{n-(k+1)}} \sim F_{(k-r)(n-k-1)}$

and we reject H_0 if $f_{obs} \geq f_{\alpha,(k-r)(n-k-1)}$

Note: A partial F-test can be used to test on single variables given that the rest of the variables are in the model, or groups of variables given that the rest of the variables are in the model and if the regression is significant.

Test on significant impact of variables

In a multiple linear regression model we have

$$E[Y] = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

Thus β_i, $i = 1, 2, \dots, n$ is the change in $E[Y]$ if x_i is changed with one unit and all the rest of the variables are kept unchanged.

We want to test. Does a variable have a significant impact on the response given that the other variables are in the model.

Such a test for x_j, $j = 1, 2, \dots, k$ is

$H_0: \beta_j = 0$ against $H_1: \beta_j \neq 0$

We know $\underline{\hat{\beta}}$ is independent of $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n-k-1}$ if

$\underline{\varepsilon} \sim N(\underline{0}, \sigma^2 \underline{I})$

Let c_{jj} be the $(j+1)$th diagonal element in $(\underline{X}^T \underline{X})^{-1}$

Then $$T = \frac{\hat{\beta}_j}{\hat{\sigma}\sqrt{c_{jj}}} = \frac{\frac{\hat{\beta}_j}{\sigma\sqrt{c_{jj}}}}{\frac{\hat{\sigma}}{\sigma}} = \frac{\frac{\hat{\beta}_j}{\sigma\sqrt{c_{jj}}}}{\sqrt{\frac{\hat{\sigma}^2 (n-k-1)}{(n-k-1)\sigma^2}}} \sim \frac{N(0,1)}{\sqrt{\frac{\chi^2(n-k-1)}{n-k-1}}}$$ if H_0 is true

and therefore t-distributed with $n-k-1$ degrees of freedom. We reject if $|T_{obs}| \geq t_{\frac{\alpha}{2}, n-k-1}$

For the test $H_0: \beta_j = \beta_{j0}$ against $H_1: \beta_j \gtrless \beta_{j0}$, we use