Matrix notation
For a set of \( n \) observations the linear regression can be written as
\[
y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \varepsilon_i \quad i = 1, \ldots, n
\]
This can be written as
\[
Y = X\beta + \varepsilon
\]  
(1)
where
\[
Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & \ldots & x_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \ldots & x_{nk} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}
\]
\( Y \in \mathbb{R}^n, \ X \in \mathbb{R}^{n \times (k+1)}, \beta \in \mathbb{R}^{k+1} \) and \( \varepsilon \in \mathbb{R}^n \).
\( X \) is called the design matrix.

Estimation of \( \sigma^2 \)
In order to obtain an estimate of \( \sigma^2 \) recall that
\[
SSE = ||Y - \hat{Y}||^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]
corresponds to the ‘size’ of the vector of residuals and gives information about the variability in the data after accounting for the regressors.
Recalling the figure we note that \( Y - \hat{Y} \) is orthogonal to \( V = \text{gen}(X) \) and thus belongs to an \( n - k - 1 \) dimensional space.
So the ‘average size’ of the error is given by
\[
s^2 = \frac{SSE}{n - k - 1} = \frac{||Y||^2 - ||\hat{Y}||^2}{n - k - 1}
\]
this is an unbiased estimator of \( \sigma^2 \).

Hypothesis testing
A t test can be done using the statistics
\[
t = \frac{\hat{\beta}_i}{s\sqrt{c_{ii}}}
\]
using the two tailed region \(|t| > t_{\alpha/2}\) to test
\[
H_0 : \beta_i = 0 \quad H_1 : \beta_i \neq 0
\]
or the one tailed region \( t > t_\alpha \) (\( t < -t_\alpha \)) to test
\[
H_0 : \beta_i = 0 \quad H_1 : \beta_i > 0 \quad (\beta_i < 0)
\]
Model Adequacy

The multiple coefficient of determination is defined as an overall measure of model adequacy:

\[ R^2 = 1 - \frac{\sum (y_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{SSE}{SS_y} \]

Since \( R^2 \) can be inflated by a large number of regressors the adjusted version of it is preferred:

\[ R_a^2 = 1 - \frac{(n - 1)(1 - R^2)}{n - k - 1} \]

It can be seen that \( R_a^2 \leq R^2 \).

Prediction Intervals

An estimate of the mean response at a particular value of the regressors \( x_{01}, \ldots, x_{0k} \) is given by

\[ \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \ldots + \hat{\beta}_k x_{0k} = x_0' \hat{\beta} \]

In general, let \( l = \sum_{i=0}^k \hat{\beta}_i = a' \hat{\beta} \).

A 100(1 - \( \alpha \))% confidence interval for \( E(l) \) (the mean response of a linear combination of \( \beta_i \)'s) is given by

\[ l \pm t_{\alpha/2} \sqrt{a'(X'X)^{-1}a} \]

where \( t_{\alpha/2} \) corresponds to a Student density with \( n - k - 1 \) DF.

A 100(1 - \( \alpha \))% confidence interval for a future value of \( y \) at a given point \( (x_{01}, \ldots, x_{0k}) \) is given by

\[ \hat{y} \pm t_{\alpha/2} \sqrt{1 + a'(X'X)^{-1}a} \]

where \( a = (1, x_{01}, \ldots, x_{0k})' \).

F Test

Complete model:

\[ y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_g x_{ig} + \beta_{g+1} x_{i(g+1)} + \ldots + \beta_k x_{ik} + \varepsilon_i \]

Reduced model:

\[ y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_g x_{ig} + \varepsilon_i \]

The test statistics is given by with a rejection region given by \( F > F_\alpha \) where \( F_\alpha \) is based on a \( f \) distribution with \( \nu_1 = k - g \) and \( \nu_2 = n - k - 1 \) DF.

\( SSE_C \) is the \( SSE \) under the complete model, while \( SSE_R \) is the \( SSE \) under the reduced model. Note that \( SSE_R \geq SSE_C \), so \( F \geq 0 \).