Testing Hypotheses in Multiple Linear Regression: overall test of significance and partial F test

Overall test of significance: $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$.

- If the multiple regression model holds,

$$E(MSE) = \sigma^2$$

- while in general
-

$$E(MSR) = \sigma^2 + \frac{\lambda}{k}$$

where

- $\lambda = \frac{1}{\sigma^2} \boldsymbol{\beta}^{*'} \boldsymbol{X}_c^{'} \boldsymbol{X}_c \boldsymbol{\beta}^*$
- $\boldsymbol{\beta}^{*'} = (\beta_1, \beta_2, \ldots, \beta_k)$
- and $\boldsymbol{X}_c$ is the centered $n \times k$ matrix with ij'th element $x_{ij} - \bar{x}_j$

- Under the null hypothesis $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$,

$$F = \frac{SSR/k}{SSE/(n-p)} = \frac{MSR}{MSE}$$

  has an $F_{k,n-p}$ distribution.

- When testing against the alternative H_A : at least one of $\beta_1, \beta_2, \ldots, \beta_k \neq 0$, the p-value for the overall test of significance of the regression is given by

$$P(F_{k,n-p} > F_{obs})$$

computational details

- the error sum of squares $SSE = \boldsymbol{y^T y} - \hat{\boldsymbol{\beta}}^{\boldsymbol{T}} \boldsymbol{X}^T \boldsymbol{y}$ has $n-p$ degrees of freedom
- the total sum of squares $SST = \boldsymbol{y^T y} - n\bar{y}^2$ has $n-1$ degrees of freedom
- the regression sum of squares $SSR = \hat{\boldsymbol{\beta}}^{\boldsymbol{T}} \boldsymbol{X}^T \boldsymbol{y} - n\bar{y}^2$ has k degrees of freedom.
- $SST = SSE + SSR$
- total degrees of freedom (n-1)= regression degrees of freedom (k) + error degrees of freedom (n-p)
- An ANOVA table is used to organize the calculation of the test statistics F
- the p-value is $P(F_{k,n-p} > F_{obs})$.

General regression significance test/Partial F test

(Section 3.3 in Montgomery *et al*)

- There are k independent variables in all, so that $\boldsymbol{\beta}$ is a $p = k + 1$ vector.
- partition $\boldsymbol{\beta}$ as

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}$$

- where $\boldsymbol{\beta}_2$ is an r vector, and $\boldsymbol{\beta}_1$ is a $p - r$ vector
- partition $\mathbf{X}$, accordingly, as

$$\mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_2]$$

  where $\mathbf{X}_1$ consists of the first $p - r$ columns of $\mathbf{X}$ and $\mathbf{X}_2$ consists of the last r columns of $\mathbf{X}$.
- In terms of these partitioned arrays, the regression model $\mathbf{y} = \mathbf{X}\beta + \epsilon$ can be written as $\mathbf{y} = \mathbf{X_1}\beta_\mathbf{1} + \mathbf{X_2}\beta_\mathbf{2} + \epsilon$

- The goal is to test $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$, against the two sided alternative $H_A : \boldsymbol{\beta}_2 \neq \mathbf{0}$
- Note that the null hypothesis sets r parameters equal to 0, which means that r variables (formally $x_{k-r+1}, x_{k-r+2}, \ldots, x_k$ are not part of the regression model, which leaves the other $k - r$ variables in the regression model.).
- The model under the alternative hypothesis is referred to as the **full model** and the model under the null hypothesis is called the **reduced model**.
- The reduced model is $\mathbf{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \epsilon$.

For the full model

- LSE of β is $\hat{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$
- regression sum of squares is

$$SSR_{Full}(\beta) = \hat{\beta}^T\boldsymbol{X}^T\boldsymbol{y} - n\bar{y}^2$$

- the error sum of squares is

$$SSE_{Full}(\beta) = \boldsymbol{y}^T\boldsymbol{y} - \hat{\beta}^T\boldsymbol{X}^T\boldsymbol{y}$$

- the error mean square is

$$MSE_{Full} = \frac{SSE_{Full}(\beta)}{n-p}$$

For the reduced model

- LSE of $\boldsymbol{\beta_1}$ is $\hat{\boldsymbol{\beta}}_{\mathbf{1}} = (\boldsymbol{X_1}^T \boldsymbol{X}_1)^{-1} \boldsymbol{X_1}^T \boldsymbol{y}$
- regression sum of squares is

$$SSR_{Red}(\boldsymbol{\beta_1}) = \hat{\boldsymbol{\beta}}_{\mathbf{1}}^T \boldsymbol{X_1}^T \boldsymbol{y} - n\bar{y}^2$$

- error sum of squares is

$$SSE_{Red}(\boldsymbol{\beta_1}) = \boldsymbol{y}^T \boldsymbol{y} - \hat{\boldsymbol{\beta}}_{\mathbf{1}}^T \boldsymbol{X_1}^T \boldsymbol{y}$$

- **The regression sum of squares due to β_2 given that β_1 is already in the model** is given by

$$SSR(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = SSR_{Full}(\boldsymbol{\beta}) - SSR_{Red}(\boldsymbol{\beta}_1)$$

  This is more appropriately phrased as **the regression sum of squares due to $\boldsymbol{X}_2$ given that $\boldsymbol{X}_1$ is already in the model**.
- Note that

$$SSR(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = SSE_{Red}(\boldsymbol{\beta}_1) - SSE_{Full}(\boldsymbol{\beta})$$

  This equivalent form is more typically used.

General regression significance test/partial F test

- $SSR(\beta_2|\beta_1)$ is independent of MSE_{Full}.
- the test statistic
$$F = \frac{SSR(\beta_2|\beta_1)/r}{MSE_{Full}}$$
equals
$$F = \frac{(SSE_{Red}(\beta_1) - SSE_{Full}(\beta))/r}{MSE_{Full}}$$
  - has an F distribution with r numerator and $n - p$ denominator degrees of freedom under the null hypothesis $H_0 : \beta_2 = \mathbf{0}$.
  - has a noncentral F distribution with r numerator and $n - p$ denominator degrees of freedom under the alternative hypothesis. (More about the noncentral F distribution later.)
  - the p-value for the test is
$$P(F_{r,n-p} > F_{Obs})$$

Partial F test - you are not responsible for remembering the algebraic derivations

- Suppose we have the model
$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}$$
and want to add the r predictors $\boldsymbol{X}_2$.
- For example, we may wish to test the hypotheses
$H_0 : \boldsymbol{\beta}_2 = 0$
$H_A : \boldsymbol{\beta}_{2_j} \neq 0$ for some j
- Then we want to compare the fit of the reduced model under H_0 to that of the full model under H_1.
- In total there are k predictors, so $\boldsymbol{X}_1$ consists of the column of 1's and $k - r$ columns of predictors.
- Write $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ where $\boldsymbol{X}_1$ is $n \times (k + 1 - r)$, $\boldsymbol{X}_2$ is $n \times r$ and $\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}$ to conform, so $\boldsymbol{\beta}_1$ is $(k + 1 - r) \times 1$ and $\boldsymbol{\beta}_2$ is $r \times 1$.

- Then the model containing $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ can be written

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}.$$

Case 1: Predictors orthogonal

- If the new predictors $\boldsymbol{X}_2$ are orthogonal to the old ones $\boldsymbol{X}_1^T\boldsymbol{X}_2 = \boldsymbol{0}$ and

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1^T\boldsymbol{X}_1 & 0 \\ 0 & \boldsymbol{X}_2^T\boldsymbol{X}_2 \end{pmatrix}$$

  which has inverse

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \begin{pmatrix} (\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1} & 0 \\ 0 & (\boldsymbol{X}_2^T\boldsymbol{X}_2)^{-1} \end{pmatrix}.$$

- The least squares estimates are

$$\begin{aligned} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} &= \begin{pmatrix} (\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1} & 0 \\ 0 & (\boldsymbol{X}_2^T\boldsymbol{X}_2)^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{X}_1^T\boldsymbol{y} \\ \boldsymbol{X}_2^T\boldsymbol{y} \end{pmatrix} \\ &= \begin{pmatrix} (\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T\boldsymbol{y} \\ (\boldsymbol{X}_2^T\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^T\boldsymbol{y} \end{pmatrix}. \end{aligned}$$

- The estimates of $\boldsymbol{\beta}_1$ are unchanged and $\boldsymbol{\beta}_2$ is estimated separately from the new columns.
- The regression sum of squares is

$$\begin{aligned} SSR(\boldsymbol{\beta}) &= \hat{\boldsymbol{\beta}}^T \boldsymbol{X}^T \boldsymbol{y} - n\bar{y}^2 \\ &= \hat{\boldsymbol{\beta}}_1^T \boldsymbol{X}_1^T \boldsymbol{y} - n\bar{y}^2 + \hat{\boldsymbol{\beta}}_2^T \boldsymbol{X}_2^T \boldsymbol{y} \\ &= SSR(\boldsymbol{\beta}_1) + SSR(\boldsymbol{\beta}_2) \end{aligned}$$

  and factors into two parts depending on $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ separately.
- The extra regression sum of squares for $\boldsymbol{X}_2$ given that $\boldsymbol{X}_1$ is already in the model can be written

$$\begin{aligned} SSR(\boldsymbol{\beta}_2) &= \hat{\boldsymbol{\beta}}_2^T \boldsymbol{X}_2^T \boldsymbol{y} \\ &= \boldsymbol{y}^T \boldsymbol{X}_2 (\boldsymbol{X}_2^T \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^T \boldsymbol{y} \\ &= \boldsymbol{y}^T \boldsymbol{H}_2 \boldsymbol{y} \end{aligned}$$

  where $\boldsymbol{H}_2 = \boldsymbol{X}_2 (\boldsymbol{X}_2^T \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^T$ is the projection onto the subspace spanned by the columns of $\boldsymbol{X}_2$ (which is orthogonal to $\boldsymbol{X}_1$).

- Under the null hypothesis that $\boldsymbol{\beta}_2 = \mathbf{0}$

$$\frac{SSR(\boldsymbol{\beta}_2)}{\sigma^2} \sim \chi^2_r$$

and

$$F = \frac{MSR(\boldsymbol{\beta}_2)}{MSE_{full}} \sim F_{r,n-1-k}$$

and large F gives evidence against H_0.

Case 2: Predictors not orthogonal

- When the new predictors are not orthogonal to the old ones, $\boldsymbol{X}_1^T \boldsymbol{X}_2 \neq \mathbf{0}$, the situation is more complicated.
- The model can be written as before, and then manipulated to create new predictors which are orthogonal

$$
\begin{aligned}
\boldsymbol{y} &= \boldsymbol{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\epsilon} \\
&= \boldsymbol{X}_1 \boldsymbol{\beta}_1 + (\boldsymbol{H}_1 + \boldsymbol{I} - \boldsymbol{H}_1) \boldsymbol{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\epsilon} \\
&= \boldsymbol{X}_1 \boldsymbol{\theta} + (\boldsymbol{I} - \boldsymbol{H}_1) \boldsymbol{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\epsilon},
\end{aligned}
$$

where

$$\boldsymbol{H}_1 = \boldsymbol{X}_1 (\boldsymbol{X}_1^T \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^T$$

is the projection on the subspace spanned by the predictors $\boldsymbol{X}_1$, and

$$\boldsymbol{\theta} = \boldsymbol{\beta}_1 + (\boldsymbol{X}_1^T \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^T \boldsymbol{X}_2 \boldsymbol{\beta}_2 \tag{1}$$

is a new parameter created from $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$.

- The matrices $\boldsymbol{X}_1$ and $(\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{X}_2$ are orthogonal, so estimates of $\boldsymbol{\theta}$ and $\beta_{\mathbf{2}}$ can be obtained separately, as above:

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{X}_1^T \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^T \boldsymbol{y} \tag{2}$$

  and

$$\hat{\boldsymbol{\beta}}_2 = [\boldsymbol{X}_2^T (\boldsymbol{I} - \boldsymbol{H}_1) \boldsymbol{X}_2]^{-1} \boldsymbol{X}_2^T (\boldsymbol{I} - \boldsymbol{H}_1) \boldsymbol{y}. \tag{3}$$

- Rearranging (1) gives

$$\hat{\boldsymbol{\beta}}_1 = \hat{\boldsymbol{\theta}} - (\boldsymbol{X}_1^T \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^T \boldsymbol{X}_2 \hat{\boldsymbol{\beta}}_2$$

  or

$$\hat{\boldsymbol{\beta}}_1 = [\boldsymbol{X}_1^T \boldsymbol{X}_1]^{-1} \boldsymbol{X}_1^T (\boldsymbol{y} - \boldsymbol{X}_2 \hat{\boldsymbol{\beta}}_2). \tag{4}$$

- From (3) we see that $\hat{\boldsymbol{\beta}}_2$ is the result of regressing one set of residuals, $(\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{y}$ on another $(\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{X}_2$.

- The latter is a matrix of residuals obtained by regressing each column of $\boldsymbol{X}_2$ on $\boldsymbol{X}_1$.
- It contains the information from $\boldsymbol{X}_2$ not already explained by $\boldsymbol{X}_1$.

back to the overall significance of the regression

- Suppose that $\boldsymbol{X}_2$ is all of $\boldsymbol{X}$ except for the initial column of ones.
- In which case $\boldsymbol{\beta}_1 = \beta_0$, $\boldsymbol{X}_1 = \mathbf{1}$, and the reduced model is

$$\mathbf{y} = \beta_0 \mathbf{1} + \epsilon$$

which just says that

$$y_i = \beta_0 + \epsilon_i$$

In this reduced model

- $\hat{\beta}_0 = \bar{y}$
- and the error sum of squares is

$$SSE_{Red} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

(which we recognize as the usual total sum of squares, SST).

- so the partial F statistic is given by

$$F = \frac{(SSE_{Red}(\boldsymbol{\beta_1}) - SSE_{Full}(\boldsymbol{\beta}))/r}{MSE_{Full}}$$

$$= \frac{(SST - SSE_{Full}(\boldsymbol{\beta}))/r}{MSE_{Full}}$$

- but this is just the F statistic used to test

$$H_0 : \beta_1 = \beta_2 = \ldots \beta_k = 0$$

  the overall test of significance of the regression.
  so the overall test of significance is just a particular application of the partial F test.