Multiple linear regression: Hypothesis test for several parameters

- Approach is to compare two models:
  - Full model: regression fitted with all 4 independent variables
  - Reduced model: regression fitted with land size and year built removed
  - Construct an F-test or ANOVA
- This approach does not indicate which coefficient is 0 even if H_0 is rejected, but only at least of these coefficients is linearly related to the response variable

Setup

Suppose now we wish to test a hypothesis of the form

$$H_0: \beta_p = \beta_{p-1} = \cdots = \beta_{p-k+1} = 0.$$

That is, the last k components of the parameter vector $\boldsymbol{\beta}$ are all zero.

$$\beta = \Big[\overbrace{\underbrace{\beta_0\ \beta_1\ \cdots\ \beta_{r-k}}_{\beta_0}}^{p\text{-}k\text{ elements}} \ \vdots\ \overbrace{\underbrace{\beta_{r-k+1}\ \cdots\ \beta_r}_{\beta_1}}^{k\text{ elements}}\Big]^T$$

$$X = \left[\begin{array}{ccccc:cccc} 1 & x_{11} & x_{12} & \cdots & x_{1,r-k} & x_{1,r-k+1} & \cdots & x_{1r} \\ 1 & x_{21} & x_{22} & \cdots & x_{2,r-k} & x_{2,r-k+1} & \cdots & x_{2r} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{n,r-k} & x_{n,r-k+1} & \cdots & x_{nr} \end{array}\right]$$

$\underbrace{\qquad\qquad\qquad}_{X_0}$ $n \times (p-k)$ matrix

To test

$H_0: \beta_1 = 0$ vs $H_a: \beta_1 \neq 0$

we compare

full model: $y = X\beta + \varepsilon$

reduced model: $y = X_0\beta_0 + \varepsilon$

Let $\boldsymbol{X}_{\mathbf{0}}$ be the matrix containing the first $p - k$ columns of $\boldsymbol{X}$ and let

$$\boldsymbol{\beta}_0 = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{r-k} \end{bmatrix}.$$

Observe that H_0 can be expressed equivalently as

$$H_0 \colon \underline{\boldsymbol{\eta} = \boldsymbol{X}_0 \boldsymbol{\beta}_0}$$

Now let

full model

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$
$$\widehat{\boldsymbol{\eta}} = \boldsymbol{X} \widehat{\boldsymbol{\beta}}$$

reduced model

$$\widehat{\boldsymbol{\beta}}_0 = (\boldsymbol{X}_0^\top \boldsymbol{X}_0)^{-1} \boldsymbol{X}_0^\top \boldsymbol{y}$$
$$\widehat{\boldsymbol{\eta}}_0 = \boldsymbol{X}_0 \widehat{\boldsymbol{\beta}}_0$$

Observe that

$$y - \hat{\eta}_0 = \underbrace{y - \hat{\eta}} + \underbrace{\hat{\eta} - \hat{\eta}_0}$$

Lemma 8

$$\underbrace{\sum_{i=1}^{n}(y_i-\hat{\eta}_{0i})^2}_{\substack{SSE_R \\ \text{(SSE for reduced model)}}}=\underbrace{\sum_{i=1}^{n}(y_i-\hat{\eta}_i)^2}_{\substack{SSE_F \\ \text{(SSE for full model)}}}+\underbrace{\sum_{i=1}^{n}(\hat{\eta}_i-\hat{\eta}_{0i})^2}_{SSE_D}.$$

That is,

$$\underline{\left\|\boldsymbol{y}-\boldsymbol{X}_0\widehat{\boldsymbol{\beta}}_0\right\|^2=\left\|\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right\|^2+\left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}}-\boldsymbol{X}_0\widehat{\boldsymbol{\beta}}_0\right\|^2.}$$

Proof of Lemma 8

$$SSE_R = \|y - X_0\hat{\beta}_0\|^2$$

$$= \|\underbrace{y - X\hat{\beta}} + \underbrace{X\hat{\beta} - X_0\hat{\beta}_0}\|^2$$

$$= \|y - X\hat{\beta}\|^2 + \|X\hat{\beta} - X_0\hat{\beta}_0\|^2 + \underbrace{2(y - X\hat{\beta})^T(X\hat{\beta} - X_0\hat{\beta}_0)}_{\text{show this equals } 0}$$

$$(y - X\hat{\beta})^T(X\hat{\beta} - X_0\hat{\beta}_0) = \left[y - X(X^TX)^{-1}X^Ty\right]^T(X\hat{\beta} - X_0\hat{\beta}_0)$$

$$= \left[(I - \underbrace{X(X^TX)^{-1}X^T}_{H})y\right]^T(X\hat{\beta} - X_0\hat{\beta}_0)$$

$$= y^T(I-H)^T(X\hat{\beta} - X_0\hat{\beta}_0)$$

$$= y^T\underbrace{(I-H)^TX}_{0}\hat{\beta} - y^T\underbrace{(I-H)^TX_0}_{0}\hat{\beta}_0$$

$$= 0$$

Recall during the proof of Theorem 10, we mentioned that $(I-H)^TX = 0$.
Recall X_0 is the first $p-k$ columns of X. It follows that $(I-H)^TX_0$ is the first $p-k$ columns of $(I-H)^TX$. Hence, $(I-H)^TX_0 = 0$.

Expected values

$Se_0^2 = \frac{\|y - X_0\hat{\beta}_0\|^2}{n-p_0} = \frac{SSE_R}{n-p_0}$

Under H_0, $\frac{(n-p_0)\, Se_0^2}{\sigma^2} \sim \chi^2_{n-p_0}$

$E\left[\frac{(n-p_0)\, Se_0^2}{\sigma^2}\right] = n-p_0$

If H_0 is true, then

$$E\left[\underbrace{\boxed{\frac{1}{n-p_0}\left\|\boldsymbol{y} - \boldsymbol{X}_0\widehat{\boldsymbol{\beta}}_0\right\|^2}}_{Se_0^2}\right] = \sigma^2,$$

where $p_0 = \underline{p - k}$.

If H_0 is true, then so is the full regression model $\boldsymbol{\eta} = \boldsymbol{X\beta}$, and so

$$E\left[\underbrace{\boxed{\frac{1}{n-p_0}\left\|\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right\|^2}}_{Se^2}\right] = \sigma^2,$$

Hence what is

$$E\left[\frac{1}{p-p_0}\left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{X}_0\widehat{\boldsymbol{\beta}}_0\right\|^2\right]? = \sigma^2$$

Recall $SSE_R = SSE_F + SSE_D$. So $E[SSE_R] = E[SSE_F] + E[SSE_D]$.

$\therefore E[SSE_D] = E[SSE_R] - E[SSE_F] = (n-p_0)\sigma^2 - (n-p)\sigma^2 = (p-p_0)\sigma^2$.

Null not correct

If the full model is correct, but H_0 is not, then it can be shown that

$$E\left[\frac{1}{p-p_0}\left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}}-\boldsymbol{X}_0\widehat{\boldsymbol{\beta}}_0\right\|^2\right] > \sigma^2$$

Test statistic

Hence we can test H_0 by calculating

$$F = \frac{\left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{X}_0\widehat{\boldsymbol{\beta}}_0\right\|^2 / (p - p_0)}{\left\|\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right\|^2 / (n - p)}$$

and rejecting it if F is ‘large’.

Theorem 12

Suppose $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$ with $\epsilon_i \sim N(0, \sigma^2)$ independently for $i = 1,2,\ldots,n$. If $H_0: \boldsymbol{\eta} = \boldsymbol{X}_0\boldsymbol{\beta}_0$ is true, then

$$\boxed{F = \frac{\left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{X}_0\widehat{\boldsymbol{\beta}}_0\right\|^2 / (p - p_0)}{\left\|\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right\|^2 / (n-p)} \sim F_{p-p_0, n-p}.}$$

$$Se^2 = \frac{\|y - X\hat{\beta}\|^2}{n-p}, \qquad \frac{(n-p)Se^2}{\sigma^2} \sim \chi^2_{n-p}$$

$$Se_0^2 = \frac{\|X\hat{\beta} - X_0\hat{\beta}_0\|^2}{p-p_0}, \qquad \frac{(p-p_0)Se_0^2}{\sigma^2} \sim \chi^2_{p-p_0}$$

independent

$$F = \frac{\frac{\cancel{(p-p_0)}\, Se_0^2}{\cancel{(p-p_0)}\, \cancel{\sigma^2}}}{\frac{\cancel{(n-p)}\, Se^2}{\cancel{(n-p)}\, \cancel{\sigma^2}}} = \frac{Se_0^2}{Se^2} \sim F_{p-p_0,\, n-p}$$

We reject H_0 if $F \geq F_{p-p_0, n-p, \alpha}$

ANOVA table

Source	SS	df	MS	F
H_0 vs M	$Q_0 - Q$ $= SSE_D$	$p - p_0$	$MSE_D = \dfrac{Q_0 - Q}{p - p_0}$ $(*)$	$F = \dfrac{*}{\dagger} = \dfrac{MSE_D}{MSE_F}$
Error	$SSE_F = Q$	$n - p$	$MSE_F = \dfrac{Q}{n - p}$ $(\dagger)$	
Total	$SSE_R = Q_0$	$n - p_0$		

where

$$\underbrace{Q = \left\|\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right\|^2}_{SSE_F} \text{ and } \underbrace{Q_0 = \left\|\boldsymbol{y} - \boldsymbol{X}_0\widehat{\boldsymbol{\beta}}_0\right\|^2}_{SSE_R}.$$

and

$$H_0: \boldsymbol{\eta} = \boldsymbol{X}_0\boldsymbol{\beta}_0 \quad \text{(reduced model)}$$
$$M: \boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} \quad \text{(full model)}$$

Remarks:

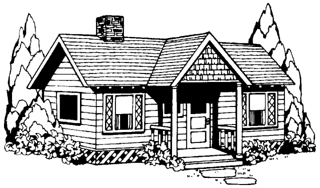
① P-value can also be calculated for this test and are provided by most statistical packages.

② This F-test may appears to be a one-sided test, but it is actually sensitive to any departure from

$$H_0: \mu_p = \mu_{p-1} = \dots = \mu_{p-k+1} = 0$$

③ When testing a single parameter, we can use either a test or an F-test. It can be proved that they are the same in this case.

Example 3.7



The price (y) in $\$1000\text{AUD}$, age ($x_1$) in decades, and area ($x_2$) in $1000\ m^2$, of five randomly chosen houses in a regional South Australian town are shown in the table below.

y	100	80	104	94	130
x_1	1	5	5	10	20
x_2	1	1	2	2	3

a) Fit a multiple linear regression model

$$y = \beta_0 + \beta_1 x_1 + \beta x_2 + \epsilon$$

to this data.

b) Test the hypothesis $H_0: \beta_1 = \beta_2 = 0$ against H_1: at least one of the $\beta_i \neq 0$, $i = 1{,}2$. Use $\alpha = 0.05$.

Example 3.7 Solution

a)

$$Y = \begin{bmatrix} 100 \\ 80 \\ 104 \\ 94 \\ 130 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 5 & 1 \\ 1 & 5 & 2 \\ 1 & 10 & 2 \\ 1 & 20 & 3 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 5 & 41 & 9 \\ 41 & 551 & 96 \\ 9 & 96 & 19 \end{bmatrix}, \quad X^T Y = \begin{bmatrix} 508 \\ 4560 \\ 966 \end{bmatrix}$$

$$(X^T X)^{-1} = \frac{1}{543} \begin{bmatrix} 1253 & 85 & -1023 \\ 85 & 14 & -111 \\ -1023 & -111 & 1074 \end{bmatrix}$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \begin{bmatrix} 66.1252 \\ -0.3794 \\ 21.4365 \end{bmatrix}$$

Example 3.7 Solution

b)

To find the SST, we need to fit the reduced model first.

The reduced model is $\boldsymbol{Y} = \boldsymbol{x}_0 \boldsymbol{\beta}_0 + \boldsymbol{\epsilon}$, where

$$\boldsymbol{x}_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \text{ and } \boldsymbol{\beta}_0 = [\beta_0].$$

Observe that $\boldsymbol{x}_0^\top \boldsymbol{x}_0 = 5$, so $(\boldsymbol{x}_0^\top \boldsymbol{x}_0)^{-1} = 1/5$.

Hence, $\widehat{\boldsymbol{\beta}}_0 = (\boldsymbol{x}_0^\top \boldsymbol{x}_0)^{-1} \boldsymbol{x}_0 \boldsymbol{y} = \frac{1}{5} \sum_{i=1}^{5} y_i = \bar{y} = 101.60.$

If follows that $\boldsymbol{y} - \boldsymbol{x}_0 \widehat{\boldsymbol{\beta}}_0 = \boldsymbol{y} - \bar{y} = \begin{bmatrix} 100 \\ 80 \\ 104 \\ 194 \\ 130 \end{bmatrix} - \begin{bmatrix} 101.60 \\ 101.60 \\ 101.60 \\ 101.60 \\ 101.60 \end{bmatrix}.$

Hence, $SSE_R = \left\| \boldsymbol{y} - \boldsymbol{x}_0 \widehat{\boldsymbol{\beta}}_0 \right\|^2 = \left(\boldsymbol{y} - \boldsymbol{x}_0 \widehat{\boldsymbol{\beta}}_0 \right)^\top \left(\boldsymbol{y} - \boldsymbol{x}_0 \widehat{\boldsymbol{\beta}}_0 \right) = 1339.20.$

Example 3.7 Solution

$p = 3, \quad p_0 = 1, \quad n = 5$

$SSE_R = \| Y - X_0 \hat{\beta}_0 \|^2 = 1339.20$

$SSE_F = \| Y - X \hat{\beta} \|^2 = 382.7$

$SSE_D = SSE_R - SSE_F = 956.5$

$MSE_F = \dfrac{SSE_F}{n-p} = 191.4$

$MSE_D = \dfrac{SSE_D}{p - p_0} = 478.2$

test statistic $\quad F = \dfrac{MSE_D}{MSE_F} = 2.5$

critical region $\quad F \geq F_{2,2,0.05} = 19 \quad$ (using `qf(0.95, 2, 2)` in R)

As F is not in the critical region, there is insufficient evidence to reject H_0.