Inference for Multiple Linear Regression

Linear combinations of β

Two important special cases:

Subset of regression coefficients:

If $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_r \end{bmatrix}$ and $\boldsymbol{\lambda} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$, then $\boldsymbol{\lambda}^{\mathsf{T}}\boldsymbol{\beta} = \beta_j$.

Prediction:

If $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_r \end{bmatrix}$ and $\boldsymbol{\lambda} = \overbrace{\begin{bmatrix} 1 \\ x_{01} \\ x_{02} \\ \vdots \\ x_{0r} \end{bmatrix}}^{x_0}$, then $\boldsymbol{\lambda}^{\mathsf{T}}\boldsymbol{\beta} = \underline{\beta_0 + \beta_1 x_{01} + \cdots + \beta_r x_{0r}}$.

Inference

In addition to showing that the least squares estimates are unbiased and providing standard errors for estimates, Theorem 11 provides the basis for inference when data are normally distributed.

$\lambda^\top \hat{\beta} \sim N(\lambda^\top \beta, \sigma^2 \lambda^\top (X^\top X)^{-1} \lambda)$

In particular,

$$T = \frac{Z}{\sqrt{\frac{V}{n-p}}} = \boxed{\frac{\boldsymbol{\lambda}^\top \widehat{\boldsymbol{\beta}} - \boldsymbol{\lambda}^\top \boldsymbol{\beta}}{s_e \sqrt{\boldsymbol{\lambda}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{\lambda}}}} \sim t_{n-p}$$

is a pivotal quantity for $\lambda^\top \beta$.

$$Z = \frac{\lambda^\top \hat{\beta} - \lambda^\top \beta}{\sigma \sqrt{\lambda^\top (X^\top X)^{-1} \lambda}} \sim N(0,1)$$ independent of $V = \frac{(n-p) S_e^2}{\sigma^2} \sim \chi^2_{n-p}$.

Confidence interval

The $100(1-\alpha)\%$ confidence interval for $\boldsymbol{\lambda}^{\top}\boldsymbol{\beta}$ is given by

$$\boldsymbol{\lambda}^{\top}\widehat{\boldsymbol{\beta}} \pm t_{n-p,\alpha/2} s_e \sqrt{\boldsymbol{\lambda}^{\top}(\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\boldsymbol{\lambda}}$$

Prediction interval for $x_0^T\beta$

$x_0 = [1 \;\; x_{01} \;\; x_{02} \ldots \;\; x_{0r}]^T$

$Y_0 \sim N(x^T\beta, \sigma^2)$

$\hat{Y}_0 = x_0^T\hat{\beta} \sim N(x_0^T\beta, \sigma^2 x_0^T (X^TX)^{-1} x_0)$ (Take $\lambda = x_0$ in Theorem 11)

$Y_0 - \hat{Y}_0 \sim N(0, \sigma^2(1 + x_0^T(X^TX)^{-1}x_0))$ as Y_0 and $\hat{Y}_0$ are independent

$Z = \dfrac{(Y_0 - \hat{Y}_0) - 0}{\sigma\sqrt{1 + x_0^T(X^TX)^{-1}x_0}} \sim N(0,1)$ and $V = \dfrac{(n-p)Se^2}{\sigma^2} \sim \chi^2_{n-p}$ are independent

$t = \dfrac{Y_0 - \hat{Y}_0}{Se\sqrt{1 + x_0^T(X^TX)^{-1}x_0}} \sim t_{n-p}$

$\therefore$ The PI is $\hat{y}_0 \pm t_{n-p,\frac{\alpha}{2}}\, Se\sqrt{1 + x_0^T(X^TX)^{-1}x_0}$.

Hypothesis test

To test H_0: $\boldsymbol{\lambda}^\top \boldsymbol{\beta} = \delta_0$ at the α level of significance, calculate

$$\underbrace{t = \frac{\boldsymbol{\lambda}^\top \widehat{\boldsymbol{\beta}} - \delta_0}{s_e \sqrt{\boldsymbol{\lambda}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{\lambda}}}}$$

Under H_0, $t \sim t_{n-p}$

and reject H_0 if

$$|t| \geq t_{n-p,\alpha/2} \quad \text{for } H_a: \lambda^\top\beta \neq \delta_0$$

$$t > t_{n-p,\alpha} \quad \text{for } H_a: \lambda^\top\beta > \delta_0$$

$$t < -t_{n-p,\alpha} \quad \text{for } H_a: \lambda^\top\beta < \delta_0$$

P-value

The P-value is given by

$$\text{P-value} = P(|T| \geq |t|)$$

where t is the observed value of the test statistic and $T \sim t_{n-p}$.

Example 3.5

Consider the Example 3.1 again. The fitted SLR model is

$$\hat{y} = -3.1011 + 2.0266\,x.$$

$\hat{\beta} = \begin{bmatrix} -3.1011 \\ 2.0266 \end{bmatrix}$

Other useful information are

$$X^{\top}X = \begin{bmatrix} 10 & 38 \\ 38 & 408 \end{bmatrix}, S_e^2 = 0.9768$$

Using the matrix approach:

a) Obtain the 95% confidence interval for $x = 5$. $x_0 = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$

b) Obtain the 95% prediction interval for $x = 5$.

c) Test the hypothesis $H_0: \beta_1 = 2$ versus$H_1: \beta_1 \neq 2$ using $\alpha = 0.05$ level of significance $\lambda = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

Example 3.5 Solution

$$\boldsymbol{x}_0^\top \widehat{\boldsymbol{\beta}} = \begin{bmatrix} 1 & 5 \end{bmatrix} \begin{bmatrix} -3.101 \\ 2.0266 \end{bmatrix} \approx 7.0319$$

$$\boldsymbol{x}_0^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_0 = \frac{1}{2636} \begin{bmatrix} 1 & 5 \end{bmatrix} \begin{bmatrix} 408 & -38 \\ -38 & 10 \end{bmatrix} \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \frac{278}{2636} \approx 0.1055$$

(a) CI is $\boldsymbol{x}_0^\top \widehat{\boldsymbol{\beta}} \pm t_{8,0.025}\, S_e \sqrt{\boldsymbol{x}_0^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_0}$

$$\approx 7.0319 \pm 2.306 \sqrt{0.9708\,(0.1055)}$$

$$\approx (6.29, 7.77)$$

(b) PI is $\boldsymbol{x}_0^\top \widehat{\boldsymbol{\beta}} \pm t_{8,0.025}\, S_e \sqrt{1 + \boldsymbol{x}_0^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_0}$

$$\approx 7.0319 \pm 2.306 \sqrt{0.9708\,(1 + 0.1055)}$$

$$\approx (4.65, 9.43)$$

Example 3.5 Solution

(c)

We set $\boldsymbol{\lambda} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ so that $\boldsymbol{\lambda}^\top \boldsymbol{\beta} = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \beta_1$. Also, $\delta_0 = 2$.

Now $\boldsymbol{\lambda}^\top \widehat{\boldsymbol{\beta}} = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} -3.101 \\ 2.0266 \end{bmatrix} = 2.0266$, and

$$\boldsymbol{\lambda}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{\lambda} = \frac{1}{2636} \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 408 & -38 \\ -38 & 10 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{10}{2636}.$$

The test statistic is $t = \dfrac{\boldsymbol{\lambda}^\top \widehat{\boldsymbol{\beta}} - \delta_0}{S_e \sqrt{\boldsymbol{\lambda}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{\lambda}}} = \dfrac{2.0266 - 2}{\sqrt{0.9708 \left(\frac{10}{2636} \right)}} \approx 0.44$.

We reject H_0 if $|t| \geq t_{8,0.025} \approx 2.306$.

As $t = 0.44 < 2.306$, there is insufficient evidence to reject H_0.