

Tutorial 5

1. Suppose we have $n = 10$ observations (X_i, Y_i) and fit the data with model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

with $\varepsilon_i, i = 1, \dots, 10$ are IID $N(0, \sigma^2)$. We have the following calculations.

$$\bar{X} = 0.5669, \quad \bar{Y} = 0.9624, \quad \sum_{i=1}^n Y_i^2 = 10.2695,$$

$$\sum_{i=1}^n X_i^2 = 4.0169, \quad \sum_{i=1}^n X_i Y_i = 6.2841.$$

- (a) Write down the estimated model

$$b_1 = \frac{\sum_{i=1}^n X_i Y_i - n \bar{X} \bar{Y}}{\sum_{i=1}^n X_i^2 - n \bar{X}^2} = 1.030974, \quad b_0 = \bar{Y} - b_1 \bar{X} = 0.3780$$

$$SST = \sum_{i=1}^n (Y_i^2 - \bar{Y})^2 = \sum_{i=1}^n Y_i^2 - n \bar{Y}^2 = 1.0069$$

$$SSR = b_1^2 \sum_{i=1}^n (X_i^2 - \bar{X})^2 = b_1^2 (\sum_{i=1}^n X_i^2 - n \bar{X}^2) = 0.85427$$

$$SSE = SST - SSR = 0.15269, \quad MSE = SSE/(n - 2) = 0.01908625$$

Thus

$$s(b_1) = \sqrt{MSE / \sum_{i=1}^n (X_i - \bar{X})^2} = 0.1541026$$

and

$$s(b_0) = \left\{ MSE \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right] \right\}^{1/2} = 0.09767$$

$$R^2 = SSR/SST = 0.8484,$$

$$F = \frac{SSR/p}{SSE/(n - p - 1)} = 44.76, \quad (\text{with } DF = 1 \text{ and } 8)$$

The model is

$$\begin{aligned} \hat{Y} &= 0.3780 + 1.030974X \\ SE &\quad (0.0977) \quad (0.15410) \\ R^2 &= 0.8484, \quad \hat{\sigma}^2 = 0.01908625, \quad F = 44.76 \end{aligned}$$

- (b) Test $H_0 : \beta_1 = 1$ with $\alpha = 0.05$

$$|t| = \left| \frac{b_1 - 1}{s(b_1)} \right| = 0.2010014 < t(1 - 0.025, 8) = 2.7515$$

We accept H_0 .

- (c) For a new $X = 1$, find the 95% CI for its expected response

$$\hat{Y} = 0.3780 + 1.030974 * 1 = 1.408974$$

$$s(\hat{Y}) = \left\{MSE \left[\frac{1}{n} + \frac{(X - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right] \right\}^{1/2} = 0.07977546$$

The CI is

$$\hat{Y} \pm s(\hat{Y}) * t(1 - 0.025, 8) = [1.2250, 1.5929]$$

- (d) For a new $X = 0.5$, find the 95% prediction interval for its possible response

$$\hat{Y} = 0.8934994,$$

$$s(pred) = \left\{MSE \left[1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right] \right\}^{1/2} = 0.1595317$$

The PI is

$$\hat{Y} \pm s(pred) * t(1 - 0.025, 8) = [0.5585214, 1.228477]$$

2. For the least square estimator b_0, b_1 of simple linear regression model, find $Cov(b_0, b_1)$

$$\begin{aligned} \mathbf{Var}\{\mathbf{b}\} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{Var}\{\mathbf{Y}\} \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2 \begin{bmatrix} \frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} & \frac{-\bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2} \\ \frac{-\bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2} & \frac{1}{\sum_{i=1}^n (X_i - \bar{X})^2} \end{bmatrix} \end{aligned}$$

$$\text{so, } Cov(b_0, b_1) = -\sigma^2 \frac{\bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

3. Suppose $A : m \times n$ is a constant matrix and $Y : n \times 1$, is a random vector. Then

$$\mathbf{Var}(AY) = A \mathbf{Var}(Y) A'$$

Please give your proof for $m = 2, n = 3$.

By definition,

$$\begin{aligned} Var(AY) &= E[\{AY - E(AY)\}\{AY - E(AY)\}^T] \\ &= E[\{A(Y - EY)\}\{A(Y - EY)\}^T] \\ &= E[A\{(Y - EY)\}\{(Y - EY)\}^T A^T] \\ &= AE[\{(Y - EY)\}\{(Y - EY)\}^T A^T] \\ &= AE[\{(Y - EY)\}\{(Y - EY)\}^T] A^T \\ &= A \mathbf{Var}(Y) A' \end{aligned}$$

4. For multiple linear regression, the normal equations are

$$\begin{aligned} \sum_{i=1}^n e_i &= 0 \\ \sum_{i=1}^n e_i X_{i1} &= 0 \\ &\vdots \\ \sum_{i=1}^n e_i X_{ip} &= 0 \end{aligned}$$

Prove that

$$\sum_{i=1}^n \hat{Y}_i e_i = 0$$

By $1st.Eq \times b_0 + 2nd.Eq \times b_1 + \dots + (p+1)th.Eq \times b_p$, we have

$$b_0 \sum_{i=1}^n e_i + b_1 \sum_{i=1}^n e_i X_{i1} + \dots + b_p \sum_{i=1}^n e_i X_{ip} = 0$$

i.e.

$$\sum_{i=1}^n e_i b_0 + \sum_{i=1}^n e_i b_1 X_{i1} + \dots + \sum_{i=1}^n e_i b_p X_{ip} = 0$$

i.e.

$$\sum_{i=1}^n e_i (b_0 + b_1 X_{i1} + \dots + b_p X_{ip}) = 0$$

i.e.

$$\sum_{i=1}^n \hat{Y}_i e_i = 0$$

5. Consider the simple regression model $Y = \beta_0 + \beta_1 X + \varepsilon$. We have n observations (X_i, Y_i) with sample correlation coefficient r_{XY} . Standardize

$$\tilde{X}_i = \frac{X_i - \bar{X}}{\sqrt{\sum_{j=1}^n (X_j - \bar{X})^2 / (n-1)}}, \quad \tilde{Y}_i = \frac{Y_i - \bar{Y}}{\sqrt{\sum_{j=1}^n (Y_j - \bar{Y})^2 / (n-1)}}$$

What is the estimated model for $\tilde{Y}_i = \tilde{\beta}_0 + \tilde{\beta}_1 \tilde{X}_i + \varepsilon_i$?

$$b_1 = r_{XY}, \quad \text{and} \quad b_0 = 0.$$