STA302/1001: Quiz #1, 10:20–11:00am, October 1, 2013

Let x_i denote the predictor variable and y_i denote the response variable. The simple linear regression model is given by $y_i = \beta_0+\beta_1 x_i+e_i$, $i = 1, \ldots, n$, where the error e_i is independently and identically (i.i.d.) distributed with mean zero and variance σ^2.

1. [12 mks] Write down the form of the residual sum of squares (RSS), and derive the ordinary least square (OLS) estimates of β_0 and β_1.

   $RSS(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$, differentiate w.r.t. to β_0 and β_1,

   $$\frac{\partial RSS(\beta_0,\beta_1)}{\partial \beta_0} = -2\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = 0$$
   $$\frac{\partial RSS(\beta_0,\beta_1)}{\partial \beta_1} = -2\sum_{i=1}^n x_i(y_i - \beta_0 - \beta_1 x_i) = 0$$
   $$\Rightarrow \beta_0 n + \beta_1 \sum_i x_i = \sum_i y_i, \quad \beta_0 \sum_i x_i + \beta_1 \sum_i x_i^2 = \sum_i x_i y_i.$$

   Solve these equations, denote $\bar{x} = \frac{1}{n}\sum_i x_i$, $\bar{y} = \frac{1}{n}\sum_i y_i$, $SXY = \sum_i (x_i - \overline{x})(y_i - \overline{y}) = \sum_i x_i y_i - n\bar{x}\bar{y}$, $SXX = \sum_i (x_i - \overline{x})^2 = \sum_i x_i^2 - n\bar{x}^2$,

   $$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \quad \hat{\beta}_1 = \frac{SXY}{SXX}$$

2. [4 mks] Write down the estimate of σ^2.

   $\hat{\sigma}^2 = \frac{1}{n-2} RSS$, where RSS is given by plugging in the OLS estimates $\hat{\beta}_0$ and $\hat{\beta}_1$.

3. [10 mks] Show that the OLS estimate $\hat{\beta}_1$ is an unbiased estimate of β_1.

   Denote $c_i = \frac{x_i - \bar{x}}{SXX}$ and $\mathbb{X} = \{x_1, \ldots, x_n\}$, then $\hat{\beta}_1 = \frac{SXY}{SXX} = \sum_i (\frac{x_i-\bar{x}}{SXX}) y_i = \sum_i c_i y_i$ and

   $$E(\hat{\beta}_1|\mathbb{X}) = E(\sum_i c_i y_i | X = x_i) = \sum_i c_i E(y_i|X = x_i)$$
   $$= \sum_i c_i(\beta_0 + \beta_1 x_i) = \beta_0 \sum_i c_i + \beta_1 \sum_i c_i x_i.$$

   Since $\sum_i c_i = \frac{\sum_i (x_i - \bar{x})}{SXX} = 0$, $\sum_i c_i x_i = \frac{\sum_i (x_i-\bar{x})x_i}{SXX} = 1$, we have $E(\hat{\beta}_1|\mathbb{X}) = \beta_1$

4. [10 mks] Derive the expression of $\mathrm{Var}(\hat{\beta}_1|\mathbb{X})$, where $\mathbb{X} = \{x_1, \ldots, x_n\}$.

   Note y_i's are assumed independent given x_i's and $SXX = \sum_{i=1}^n (x_i - \bar{x})^2$,

   $$\begin{aligned} Var(\hat{\beta}_1|\mathbb{X}) &= Var(\sum_i c_i y_i|\mathbb{X}) = \sum_i c_i^2 Var(y_i|X = x_i) = \sigma^2 \sum_i c_i^2 \\ &= \sigma^2 \sum_i (x_i - \bar{x})^2 / SXX^2 = \sigma^2/SXX \end{aligned}$$

5. [14 mks] Rewrite the model as $y_i = \beta_0^* + \beta_1^*(x_i - \bar{x}) + e_i$, where $\bar{x} = n^{-1}\sum_{i=1}^n x_i$. What is the relationship between β_0 and β_0^*, β_1 and β_1^*? Derive the OLS estimates of β_0^* and β_1^*, and the covariance between $\hat{\beta}_0^*$ and $\hat{\beta}_1^*$.

   $y_i = \beta_0 + \beta_1 x_i + e_i = \beta_0 + \beta_1\bar{x} + \beta_1(x_i - \bar{x}) + e_i$, thus $\beta_0^* = \beta_0 + \beta_1\bar{x}$ and $\beta_1^* = \beta_1$. To get the OLS estimates of $\beta_0^a st$ and β_1^*, differentiating the following w.r.t. β_0^* and β_1^*,

   $RSS(\beta_0^*, \beta_1^*) = \sum_{i=1}^n [y_i - \beta_0^* - \beta_1^*(x_i - \bar{x})]^2$, we get

   $\sum_i [y_i - \beta_0^* - \beta_1^*(x_i - \bar{x})] = 0 \Rightarrow \hat{\beta}_0^* = \bar{y}$, since $\sum_i (x_i - \bar{x}) = 0$.

   $\sum_i [y_i - \beta_0^* - \beta_1^*(x_i - \bar{x})](x_i - \bar{x}) = 0 \Rightarrow \hat{\beta}_1^* = SXY/SXX$.

   Find the covariance between $\hat{\beta}_0^*$ and $\hat{\beta}_1^*$,

$$\begin{aligned} Cov(\bar{y}, \hat{\beta}_1^*|\mathbb{X}) &= Cov(\frac{1}{n}\sum_i y_i, \sum_i c_i y_i|\mathbb{X}) = \frac{1}{n}\sum_i c_i Cov(y_i, y_i|\mathbb{X}) \\ &= \frac{\sigma^2}{n}\sum_i c_i = \frac{\sigma^2}{n}\sum_i (x_i - \bar{x}) = 0 \end{aligned}$$