Chapter 2
Inferences in Regression and Correlation Analysis

In the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \ i = 1, 2, ..., n$$
$$E(\varepsilon_i) = 0, Var(\varepsilon_i) = \sigma^2 \quad and \quad Cov(\varepsilon_i, \varepsilon_j) = 0 \ for \ all \ i \neq j.$$

Then

$$E(Y_i) = \beta_0 + \beta_1 X_i \qquad and \qquad Var(Y_i) = \sigma^2.$$

Let's introduce some more notations:

$$S_{xx} = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2$$
$$S_{yy} = \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2$$
$$S_{xy} = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) = \sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}$$

The point estimates of β_0, β_1 are

$$\hat{\beta}_1 = b_1 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} = \frac{S_{xy}}{S_{xx}} \tag{2.1}$$

$$\hat{\beta}_0 = b_0 = \bar{Y} - b_1\bar{X}$$

Properties of Point estimation of β_1, β_0

The point estimation of the coefficients of the simple linear regression model in (2.1) can be written in linear combination forms of Y_i as follows:

$$\widehat{\beta}_1 = b_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})Y_i - \bar{Y}\sum_{i=1}^{n}(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$= \frac{\sum_{i=1}^{n}(X_i - \bar{X})Y_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2},$$

$$= \sum_{i=1}^{n}\left(\frac{(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)Y_i$$

$$= \sum_{i=1}^{n} K_i Y_i$$

$$\widehat{\beta}_1 = b_1 = \sum_{i=1}^{n} K_i Y_i \qquad (2.2)$$

where

$$K_i = \frac{\left(X_i - \bar{X}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} \tag{2.3}$$

As we can see, the form of the point estimation of slope the simple liner model is given in a liner combination form of Y_i .

Similarly,

$$\widehat{\beta}_0 = b_0 = \bar{Y} - b_1\bar{X} = \widehat{\beta}_1 = \bar{Y} - \bar{X}\sum_{i=1}^{n} k_i Y_i$$

$$= \sum_{i=1}^{n} \frac{Y_i}{n} - \bar{X}\sum_{i=1}^{n} k_i Y_i$$

$$= \sum_{i=1}^{n} \left(\frac{1}{n} - \bar{X}k_i\right) Y_i$$

$$= \sum_{i=1}^{n} L_i Y_i$$

$$\widehat{\beta}_0 = b_0 = \sum_{i=1}^{n} L_i Y_i \tag{2.4}$$

where

$$L_i = \frac{1}{n} - \bar{X} K_i \tag{2.5}$$

and K_i is given in (2.3).

The coefficients K_i and L_i satisfy the following properties:

Lemma

The coefficients K_i given in (2.3) satisfies the following properties

$$\sum k_i = 0$$

$$\sum k_i X_i = 1$$

$$\sum k_i^2 = \frac{1}{\sum (X_i - \bar{X})^2}$$

Proof.

$$\sum k_i = \sum \left[\frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2} \right] = \frac{1}{\sum (X_i - \bar{X})^2} \sum (X_i - \bar{X}) = \frac{0}{\sum (X_i - \bar{X})^2} = 0$$

$$K_i = \frac{\left(X_i - \bar{X}\right)}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2} \implies K_i X_i = \frac{\left(X_i^2 - \bar{X} X_i\right)}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}$$

$$\implies \sum_{i=1}^{n} K_i X_i = \frac{\sum_{i=1}^{n} \left(X_i^2 - \bar{X} X_i\right)}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2} = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2} = 1$$

$$\sum k_i^2 = \sum \left[\frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2} \right]^2 = \frac{1}{\left[\sum (X_i - \bar{X})^2\right]^2} \sum (X_i - \bar{X})^2 = \frac{1}{\sum (X_i - \bar{X})^2}$$

Lemma

The coefficients L_i given in (2.5) satisfies the following properties

$$\sum_{i=1}^{n} L_i = 1$$

$$\sum_{i=1}^{n} L_i X_i = 0$$

$$\sum_{i=1}^{n} L_i^2 = \frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}$$

Proof.

$$\sum_{i=1}^{n} L_i = \sum_{i=1}^{n} \left(\frac{1}{n} - \bar{X} K_i\right) = \frac{n}{n} - \bar{X} \sum K_i = 1 - 0 = 1$$

$$\sum_{i=1}^{n} L_i X_i = \sum_{i=1}^{n} \left(\frac{1}{n} - \bar{X} K_i\right) X_i = \frac{1}{n} \sum_{i=1}^{n} X_i - \bar{X} \sum_{i=1}^{n} K_i X_i$$

$$= \bar{X} - \bar{X} = 0$$

$$\sum_{i=1}^{n} L_i^2 = \sum_{i=1}^{n} \left(\frac{1}{n} - \bar{X} K_i\right)^2 = \sum_{i=1}^{n} \left(\frac{1}{n^2} - \frac{2K_i \bar{X}}{n} + K_i^2 \bar{X}^2\right)$$

$$= \sum_{i=1}^{n} \frac{1}{n^2} - \frac{2\bar{X}\sum_{i=1}^{n} K_i}{n} + \bar{X}^2 \sum_{i=1}^{n} K_i^2$$

$$= \frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2}$$

1- Unbiasedness of Point estimation of β_1, β_0

Lemma

The point estimators of β_1 and β_0 are unbiased

Proof.

From (2.2), we have

$\hat{\beta}_1 = b_1 = \sum_{i=1}^{n} K_i Y_i$ then

$$E(\widehat{\beta}_1) = E(b_1) = \sum_{i=1}^{n} K_i E(Y_i) = \sum_{i=1}^{n} K_i (\beta_0 + \beta_1 X_i)$$

$$= \sum_{i=1}^{n} K_i (\beta_0 + \beta_1 X_i)$$

$$= \beta_0 \sum_{i=1}^{n} K_i + \beta_1 \sum_{i=1}^{n} K_i X_i$$

$$= \beta_1.$$

Similarly, from (2.4), we have

$$E(\widehat{\beta}_0) = E(b_0) = \sum_{i=1}^{n} L_i E(Y_i) = \sum_{i=1}^{n} L_i (\beta_0 + \beta_1 X_i)$$

$$= \sum_{i=1}^{n} L_i (\beta_0 + \beta_1 X_i)$$

$$= \beta_0 \sum_{i=1}^{n} L_i + \beta_1 \sum_{i=1}^{n} L_i X_i$$

$$= \beta_0.$$

2- Variances

Lemma

The point estimators of β_1 and β_0 have the following variances, respectively

$$Var(\widehat{\beta}_1) = Var(b_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}.$$

and

$$Var(\widehat{\beta}_0) = Var(b_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right].$$

Proof.

From (2.2), we have

$\widehat{\beta}_1 = b_1 = \sum_{i=1}^{n} K_i Y_i$ then

$$Var(\hat{\beta}_1) = Var(b_1) = \sum_{i=1}^{n} K_i^2 Var(Y_i) = \sum_{i=1}^{n} K_i^2 \sigma^2$$

$$= \frac{\sigma^2}{\sum (X_i - \bar{X})^2}.$$

Similarly, from (2.4), we have

$$Var(\hat{\beta}_0) = Var(b_0) = \sum_{i=1}^{n} L_i^2 Var(Y_i) = \sum_{i=1}^{n} L_i^2 \sigma^2$$

$$= \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right].$$

Example

Consider the Toluca Company example, the variance of β_1, β_0 are:

$$Var(\hat{\beta}_1) = Var(b_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}$$

$$= \frac{MSE}{\sum (X_i - \bar{X})^2} = \frac{2384}{19800} = .12040$$

Hence the squared error of $\hat{\beta}_1$ is

$$S.E(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)} = \sqrt{.12040} = .3470$$

Similarly,

$$Var(\widehat{\beta}_0) = Var(b_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right]$$

$$= MSE \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right]$$

$$= 2{,}384 \left[\frac{1}{25} + \frac{70^2}{19800} \right]$$

$$= 685.34$$

Hence the squared error of $\widehat{\beta}_0$ is

$$S.E(\widehat{\beta}_0) = \sqrt{Var(\widehat{\beta}_0)} = \sqrt{685.34} = 26.18$$