Applied Linear Statistical Models, Part 1

Section 1.6. Estimation of Regression Function—Proofs of Theorems

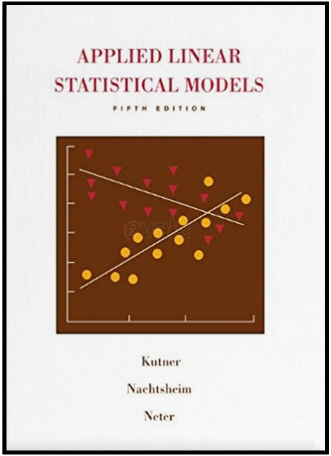

Table of contents

1. Theorem 1.6.A
2. Theorem 1.6.B

Theorem 1.6.A

Theorem 1.6.A. For data points (X_i, Y_i) where $i = 1, 2, \ldots, n$, the values of β_0 and β_1 which minimize

$$Q = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$

are

$$b_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \text{ and } b_0 = \frac{1}{n}\left(\sum_{i=1}^{n} Y_i - b_1 \sum_{i=1}^{n} X_i\right) = \overline{Y} - b_1\overline{X},$$

respectively.

Proof. We define Q as a function of β_0 and β_1 (in terms of the given data (X_i, Y_i)):

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2.$$

Theorem 1.6.A

Theorem 1.6.A. For data points (X_i, Y_i) where $i = 1, 2, \ldots, n$, the values of β_0 and β_1 which minimize

$$Q = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$

are

$$b_1 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n} (X_i - \overline{X})^2} \text{ and } b_0 = \frac{1}{n}\left(\sum_{i=1}^{n} Y_i - b_1 \sum_{i=1}^{n} X_i\right) = \overline{Y} - b_1 \overline{X},$$

respectively.

Proof. We define Q as a function of β_0 and β_1 (in terms of the given data (X_i, Y_i)):

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2.$$

Theorem 1.6.A (continued 1)

Proof. To find critical points of Q, we consider the partial derivatives:

$$\frac{\partial Q}{\partial \beta_0} = \sum_{i=1}^{n} -2(Y_i - \beta_0 - \beta_1 X_i) \text{ and } \frac{\partial Q}{\partial \beta_1} = \sum_{i=1}^{n} -2X_i(Y_i - \beta_0 - \beta_1 X_i).$$

Expanding the partial derivatives:

$$\frac{\partial Q}{\partial \beta_0} = \sum_{i=1}^{n} -2(Y_i - \beta_0 - \beta_1 X_i) = \sum_{i=1}^{n}(-2Y_i + 2\beta_0 + 2\beta_1 X_i)$$

$$= -2\sum_{i=1}^{n} Y_i + 2n\beta_0 + 2\beta_1 \sum_{i=1}^{n} X_i$$

and

$$\frac{\partial Q}{\partial \beta_1} = \sum_{i=1}^{n} -2X_i(Y_i - \beta_0 - \beta_1 X_i) = -2\sum_{i=1}^{n} X_i Y_i + 2\beta_0 \sum_{i=1}^{n} X_i + 2\beta_1 \sum_{i=1}^{n} X_i^2.$$

Theorem 1.6.A (continued 2)

Proof. Setting each partial derivative equal to 0 gives the following two equations in unknowns β_0 and β_1:

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

and

$$\beta_0 \sum_{i=1}^{n} X_i + \beta_1 \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Y_i. \qquad (**)$$

(These are called the *normal equations*.) This is a system of two linear equations in two unknowns and has solution $\beta_0 = b_0$ and $\beta_1 = b_1$ where:

$$b_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \text{ and } b_0 = \frac{1}{n}\left(\sum_{i=1}^{n} Y_i - b_1 \sum_{i=1}^{n} X_i\right) = \overline{Y} - b_1\overline{X}.$$

Theorem 1.6.A (continued 2)

Proof. Setting each partial derivative equal to 0 gives the following two equations in unknowns β_0 and β_1:

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

and

$$\beta_0 \sum_{i=1}^{n} X_i + \beta_1 \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Y_i. \qquad (**)$$

(These are called the *normal equations*.) This is a system of two linear equations in two unknowns and has solution $\beta_0 = b_0$ and $\beta_1 = b_1$ where:

$$b_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \text{ and } b_0 = \frac{1}{n}\left(\sum_{i=1}^{n} Y_i - b_1 \sum_{i=1}^{n} X_i\right) = \overline{Y} - b_1\overline{X}.$$

Theorem 1.6.A (continued 3)

Proof. Finally, we need to check a second partial derivative of $Q(\beta_0, \beta_1)$ at the critical point (b_0, b_1). We have

$$\frac{\partial^2 Q}{\partial \beta_1^2} = \frac{\partial}{\partial \beta_1}\left[\frac{\partial Q}{\partial \beta_1}\right] = \frac{\partial}{\partial \beta_1}\left[-2\sum_{i=1}^{n} X_i Y_i + 2\beta_0 \sum_{i=1}^{n} X_i + 2\beta_1 \sum_{i=1}^{n} X_i^2\right]$$

$$= 2\beta_0 \sum_{i=1}^{n} X_i^2 > 0.$$

So by the Second Derivative Test for Local Extreme Values (see my online Calculus 3 [MATH 2110] notes on Section 14.7. Extreme Values and Saddle Points, the critical point $(\beta_0, \beta_1) = (b_0, b_1)$ yields a local minimum. Since there is only one critical point, this must yield an absolute (or "global") minimum. □

Theorem 1.6.B

Theorem 1.6.B. For data points (X_i, Y_i) where $i = 1, 2, \ldots, n$, estimated regression model $\hat{Y} = b_0 + b_1 X$, and residuals $e_i = Y_i - \hat{Y}_i$, we have the following properties.

1. The sum of the residuals is zero: $\sum_{i=1}^{n} e_i = 0$.
2. The sum of the squared residuals, $\sum_{i=1}^{n} e_i^2$, is a minimum.
3. The sum of the observed values Y_i equals the sum of the fitted values $\hat{Y}_i$: $\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i$.
4. The sum of the weighted residuals is zero when the residual in the ith trial is weighted by the level of the predictor variable in the ith trial. That is, $\sum_{i=1}^{n} X_i e_i = 0$.
5. The sum of the weighted residuals is zero when the ith trial is weighted by the fitted value of the response variable for the ith trial. That is, $\sum_{i=1}^{n} \hat{Y}_i e_i = 0$.
6. The regression line always goes through the point $(\overline{X}, \overline{Y})$.

Theorem 1.6.B (continued 1)

Theorem 1.6.B. For data points (X_i, Y_i) where $i = 1, 2, \ldots, n$, estimated regression model $\hat{Y} = b_0 + b_1 X$, and residuals $e_i = Y_i - \hat{Y}_i$, we have the following properties.

1. The sum of the residuals is zero: $\sum_{i=1}^n e_i = 0$.

Proof. (1) The ith residual is (by definition) $e_i = Y_i - \hat{Y}_i$ and the estimated regression model is $\hat{Y} = b_0 + b_1 X$, so

$$\sum_{i=1}^n e_i = \sum_{i=1}^n (Y_i - \hat{Y}_i) = \sum_{i=1}^n (Y_i - (b_0 - b_1 X_i))$$

$$= \sum_{i=1}^n Y_i - \sum_{i=1}^n b_0 + b_1 \sum_{i=1}^n X_i = \sum_{i=1}^n Y_i - nb_0 - b_1 \sum_{i=1}^n X_i = 0,$$

as claimed, since $\sum_{i=1}^n Y_i = nb_0 + b_1 \sum_{i=1}^n X_i$ by the first normal equation given as $(*)$ in the proof of Theorem 1.6.A above.

Theorem 1.6.B (continued 2)

Theorem 1.6.B. For data points (X_i, Y_i) where $i = 1, 2, \ldots, n$, estimated regression model $\hat{Y} = b_0 + b_1 X$, and residuals $e_i = Y_i - \hat{Y}_i$, we have the following properties.

2. The sum of the squared residuals, $\sum_{i=1}^{n} e_i^2$, is a minimum.
3. The sum of the observed values Y_i equals the sum of the fitted values $\hat{Y}_i$: $\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i$.

Proof (continued). (2) By Theorem 1.6.A, the values of β_0 and β_1 that minimize quantity $Q = \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2$ are b_0 and b_1, respectively. When we replace β_0 and β_1 with b_0 and b_1, respectively, we get that the minimum quantity is $\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)^2$. Now $\hat{Y}_i = b_0 - b_1 X_i$, so the minimum quantity is $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i-1}^{n} e_i^2$, as claimed.

(3) The proof of this is to be given in Exercise 1.35.

Theorem 1.6.B (continued 2)

Theorem 1.6.B. For data points (X_i, Y_i) where $i = 1, 2, \ldots, n$, estimated regression model $\hat{Y} = b_0 + b_1 X$, and residuals $e_i = Y_i - \hat{Y}_i$, we have the following properties.

2. The sum of the squared residuals, $\sum_{i=1}^n e_i^2$, is a minimum.
3. The sum of the observed values Y_i equals the sum of the fitted values $\hat{Y}_i$: $\sum_{i=1}^n Y_i = \sum_{i=1}^n \hat{Y}_i$.

Proof (continued). (2) By Theorem 1.6.A, the values of β_0 and β_1 that minimize quantity $Q = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2$ are b_0 and b_1, respectively. When we replace β_0 and β_1 with b_0 and b_1, respectively, we get that the minimum quantity is $\sum_{i=1}^n (Y_i - b_0 - b_1 X_i)^2$. Now $\hat{Y}_i = b_0 - b_1 X_i$, so the minimum quantity is $\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i-1}^n e_i^2$, as claimed.

(3) The proof of this is to be given in Exercise 1.35.

Theorem 1.6.B (continued 3)

Theorem 1.6.B. For data points (X_i, Y_i) where $i = 1, 2, \ldots, n$, estimated regression model $\hat{Y} = b_0 + b_1 X$, and residuals $e_i = Y_i - \hat{Y}_i$, we have the following properties.

4. The sum of the weighted residuals is zero when the residual in the ith trial is weighted by the level of the predictor variable in the ith trial. That is, $\sum_{i=1}^n X_i e_i = 0$.

Proof (continued). (4) Since $e_i = Y_i - \hat{Y}_i$, then $e_i = Y_i - (b_0 + b_1 X_i)$ and

$$\sum_{i=1}^n X_i e_i = \sum_{i=1}^n X_i(Y_i - b_0 - b_1 X_i)) = \sum_{i=1}^n X_i Y_i - b_0 \sum_{i=1}^n X_i - b_1 \sum_{i=1}^n X_i^2.$$

Now $\sum_{i=1}^n X_i Y_i = b_0 \sum_{i=1}^n X_i + \beta_1 \sum_{i=1}^n X_i^2$ by the second normal equation given as $(**)$ in the proof of Theorem 1.6.A above. Therefore, $\sum_{i=1}^n X_i e_i = 0$, as claimed.

Theorem 1.6.B (continued 4)

Theorem 1.6.B. For data points (X_i, Y_i) where $i = 1, 2, \ldots, n$, estimated regression model $\hat{Y} = b_0 + b_1 X$, and residuals $e_i = Y_i - \hat{Y}_i$, we have the following properties.

5. The sum of the weighted residuals is zero when the *i*th trial is weighted by the fitted value of the response variable for the *i*th trial. That is, $\sum_{i=1}^{n} \hat{Y}_i e_i = 0$.
6. The regression line always goes through the point $(\overline{X}, \overline{Y})$.

Proof (continued). (5) The proof of this is to be given in Exercise 1.36.

(6) By Note 1.6.B, the alternative form of the estimated regression model (i.e., the regression line) is $\hat{Y} = \overline{Y} + b_1(X - \overline{X})$. So when the predictor variable X takes on the value $\overline{X}$, we have

$$\hat{Y} = \overline{Y} + b_1(\overline{X} - \overline{X}) = \overline{Y}.$$

That is, the regression line goes through the point $(\overline{X}, \overline{Y})$, as claimed. □

Theorem 1.6.B (continued 4)

Theorem 1.6.B. For data points (X_i, Y_i) where $i = 1, 2, \ldots, n$, estimated regression model $\hat{Y} = b_0 + b_1 X$, and residuals $e_i = Y_i - \hat{Y}_i$, we have the following properties.

5. The sum of the weighted residuals is zero when the ith trial is weighted by the fitted value of the response variable for the ith trial. That is, $\sum_{i=1}^{n} \hat{Y}_i e_i = 0$.
6. The regression line always goes through the point $(\overline{X}, \overline{Y})$.

Proof (continued). (5) The proof of this is to be given in Exercise 1.36.

(6) By Note 1.6.B, the alternative form of the estimated regression model (i.e., the regression line) is $\hat{Y} = \overline{Y} + b_1(X - \overline{X})$. So when the predictor variable X takes on the value $\overline{X}$, we have

$$\hat{Y} = \overline{Y} + b_1(\overline{X} - \overline{X}) = \overline{Y}.$$

That is, the regression line goes through the point $(\overline{X}, \overline{Y})$, as claimed. □