Applied Linear Statistical Models, Part 1

Section 2.1. Inferences Concerning β_1—Proofs of Theorems

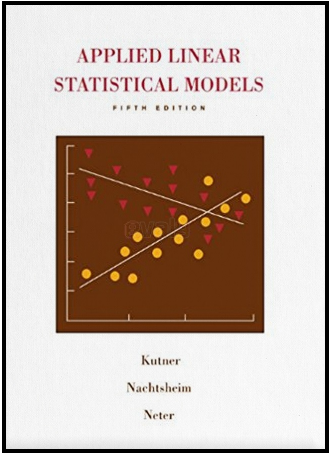

Table of contents

Lemma 2.1.A

Lemma 2.1.A. Statistic b_1 is a linear combination of the observations Y_i:

$$b_1 = \sum_{i=1}^{n} k_i Y_i \text{ where } k_i = \frac{X_i - \overline{X}}{\sum_{i=1}^{n}(X_i - \overline{X})^2}.$$

Proof. First, we have

$$\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) = \sum_{i=1}^{n}(X_i - \overline{X})Y_i - \sum_{i=1}^{n}(X_i - \overline{X})\overline{Y}.$$

But

$$\sum_{i=1}^{n}(X_i - \overline{X})\overline{Y} = \overline{Y}\sum_{i=1}^{n}(X_i - \overline{X}) = \overline{Y}\sum_{i=1}^{n}\left(X_i - \sum_{i=1}^{n} X_i/n\right)$$

$$= \overline{Y}\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} X_i\right) = 0.$$

Lemma 2.1.A

Lemma 2.1.A. Statistic b_1 is a linear combination of the observations Y_i:

$$b_1 = \sum_{i=1}^{n} k_i Y_i \text{ where } k_i = \frac{X_i - \overline{X}}{\sum_{i=1}^{n}(X_i - \overline{X})^2}.$$

Proof. First, we have

$$\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) = \sum_{i=1}^{n}(X_i - \overline{X})Y_i - \sum_{i=1}^{n}(X_i - \overline{X})\overline{Y}.$$

But

$$\sum_{i=1}^{n}(X_i - \overline{X})\overline{Y} = \overline{Y}\sum_{i=1}^{n}(X_i - \overline{X}) = \overline{Y}\sum_{i=1}^{n}\left(X_i - \sum_{i=1}^{n} X_i/n\right)$$

$$= \overline{Y}\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} X_i\right) = 0.$$

Lemma 2.1.A (continued)

Proof (continued). So

$$\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) = \sum_{i=1}^{n}(X_i - \overline{X})Y_i - \sum_{i=1}^{n}(X_i - \overline{X})\overline{Y} = \sum_{i=1}^{n}(X_i - \overline{X})Y_i.$$

Then we have

$$b_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})Y_i}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \sum_{i=1}^{n} k_i Y_i,$$

where $k_i = \dfrac{X_i - \overline{X}}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$, as claimed. □

Theorem 1.11

Theorem 1.11. The Gauss-Markov Theorem for the Normal Error Regression Model.
Consider the data points (X_i, Y_i) for $i = 1, 2, \ldots, n$ and the normal error linear regression model $Y_i = \beta_0 + \beta_1 X_1 + \varepsilon_i$ given in (1.1) with the added hypothesis that each error term has a $N(0, \sigma^2)$ distribution. The least squares estimators

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

are unbiased (that is, $E\{b_0\} = \beta_0$ and $E\{b_1\} = \beta_1$) and have minimum variance among all unbiased linear estimators (i.e., linear combinations of the Y_i).

Proof. First we have by Lemma 2.1.A and the linearity of expectation (see my online notes for Mathematical Statistics 1 [STAT 4047/5047] on Section 1.8. Expectation of a Random Variables; notice Theorem 1.8.2) we have: ...

Theorem 1.11

Theorem 1.11. The Gauss-Markov Theorem for the Normal Error Regression Model.
Consider the data points (X_i, Y_i) for $i = 1, 2, \ldots, n$ and the normal error linear regression model $Y_i = \beta_0 + \beta_1 X_1 + \varepsilon_i$ given in (1.1) with the added hypothesis that each error term has a $N(0, \sigma^2)$ distribution. The least squares estimators

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

are unbiased (that is, $E\{b_0\} = \beta_0$ and $E\{b_1\} = \beta_1$) and have minimum variance among all unbiased linear estimators (i.e., linear combinations of the Y_i).

Proof. First we have by Lemma 2.1.A and the linearity of expectation (see my online notes for Mathematical Statistics 1 [STAT 4047/5047] on Section 1.8. Expectation of a Random Variables; notice Theorem 1.8.2) we have: ...

Theorem 1.11 (continued 1)

Proof (continued). ...

$$E\{b_1\} = E\left\{\sum_{i=1}^{n} k_i Y_i\right\} = \sum_{i=1}^{n} k_i E\{Y_i\}$$

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

because $\sum_{i=1}^{n} k_i = 0$ by (2.5) and $\sum_{i=1}^{n} k_i X_i = 1$ by Exercise 2.50, as claimed.

Recall that for independent Y_i, we have $\sigma^2\left(\sum_{i=1}^{n} a_i Y_i\right) = \sum_{i=1}^{n} a_1^2 \sigma^2(Y_i)$; see equation A.31 in Appendix A or my online notes on Mathematical Statistics 1 on Section 1.8. Expectation of a Random Variables (see Theorem 1.8.2) and on Section 1.9. Some Special Expectations (see Theorem 1.9.1). We use this to find $\sigma\{b_1\}$.

Theorem 1.11 (continued 1)

Proof (continued). ...

$$E\{b_1\} = E\left\{\sum_{i=1}^{n} k_i Y_i\right\} = \sum_{i=1}^{n} k_i E\{Y_i\}$$

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

because $\sum_{i=1}^{n} k_i = 0$ by (2.5) and $\sum_{i=1}^{n} k_i X_i = 1$ by Exercise 2.50, as claimed.

Recall that for independent Y_i, we have $\sigma^2\left(\sum_{i=1}^{n} a_i Y_i\right) = \sum_{i=1}^{n} a_1^2 \sigma^2(Y_i)$; see equation A.31 in Appendix A or my online notes on Mathematical Statistics 1 on Section 1.8. Expectation of a Random Variables (see Theorem 1.8.2) and on Section 1.9. Some Special Expectations (see Theorem 1.9.1). We use this to find $\sigma\{b_1\}$.

Theorem 1.11 (continued 2)

Proof (continued). We now have by Lemma 2.1.A and (2.7) of Note 2.1.A that

$$\sigma^2\{b_1\} = \sigma^2\left\{\sum_{i=1}^{n} k_i Y_i\right\} = \sum_{i=1}^{n} k_i^2 \sigma^2\{Y_i\}$$

$$= \sum_{i=1}^{n} k_i^2 \sigma^2 = \sigma^2 \frac{1}{\sum_{i=1}^{n}(X_i - \overline{X})^2}.$$

Next, we need to show that among all unbiased linear estimators of β_1 of the form $\hat{\beta}_1 = \sum_{i=1}^{n} c_i Y_i$, the one of minimum variance is the one for which $c_i = k_i$ for $i = 1, 2, \ldots, n$. Since $\hat{\beta}_1$ must be unbiased, then

$$E\{\hat{\beta}_1\} = E\left\{\sum_{i=1}^{n} c_i Y_i\right\} = \sum_{i=1}^{n} c_i E\{Y_i\} = \beta_1.$$

Theorem 1.11 (continued 3)

Proof (continued). We know $E\{Y_i\} = \beta_0 + \beta_1 X_i$ by Note 1.3.A, so we now have

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

$$= \beta_0 \sum i = 1^n c_i + \beta_1 \sum_{i=1}^{n} c_i X_i = \beta_1.$$

For this to hold for arbitrary normal error linear regression (and hence to hold for all β_0 and β_1) we must have $\sum_{i=1}^{n} c_i = 0$ and $\sum_{i=1}^{n} c_i X_i = 1$.

As described above, the variance of $\hat{\beta}_1$ satisfies:

$$\sigma^2\{\hat{\beta}_1\} = \sigma^2 \left\{ \sum_{i=1}^{n} c_i Y_i \right\} = \sum_{i=1}^{n} c_i^2 \sigma^2\{Y_i\} = \sigma^2 \sum_{i=1}^{n} c_i^2.$$

Define d_i so that it satisfies $c_i = k_i + d_i$, where the k_i are defined in Lemma 2.1.A.

Theorem 1.11 (continued 3)

Proof (continued). We know $E\{Y_i\} = \beta_0 + \beta_1 X_i$ by Note 1.3.A, so we now have

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

$$= \beta_0 \sum i = 1^n c_i + \beta_1 \sum_{i=1}^{n} c_i X_i = \beta_1.$$

For this to hold for arbitrary normal error linear regression (and hence to hold for all β_0 and β_1) we must have $\sum_{i=1}^{n} c_i = 0$ and $\sum_{i=1}^{n} c_i X_i = 1$.

As described above, the variance of $\hat{\beta}_1$ satisfies:

$$\sigma^2\{\hat{\beta}_1\} = \sigma^2 \left\{ \sum_{i=1}^{n} c_i Y_i \right\} = \sum_{i=1}^{n} c_i^2 \sigma^2\{Y_i\} = \sigma^2 \sum_{i=1}^{n} c_i^2.$$

Define d_i so that it satisfies $c_i = k_i + d_i$, where the k_i are defined in Lemma 2.1.A.

Theorem 1.11 (continued 4)

Proof (continued). The variance of $\hat{\beta}_1$ then becomes

$$\sigma^2\{\hat{\beta}_1\} = \sigma^2 \sum_{i=1}^{n} c_i^2 = \sigma^2 \sum_{i=1}^{n} (k_i + d_i)^2 = \sigma^2 \left(\sum_{i=1}^{n} k_i^2 + \sum_{i=1}^{n} d_i^2 + 2 \sum_{i=1}^{n} k_i d_i \right).$$

Now
$$\begin{aligned}
\sum_{i=1}^{n} k_i d_i &= \sum_{i=1}^{n} k_i (c_i - k_i) = \sum_{i=1}^{n} c_i k_i - \sum_{i=1}^{n} k_i^2 \\
&= \sum_{i=1}^{n} c_i \left(\frac{X_i - \overline{X}}{\sum_{i=1}^{n} (X_i - \overline{X})^2} \right) - \frac{1}{\sum_{i=1}^{n} (X_1 - \overline{X})^2} \\
&\qquad \text{by Lemma 2.1.A and (2.7) of Note 2.1.A} \\
&= \frac{\sum_{i=1}^{n} c_i X_i - \overline{X} \sum_{i=1}^{n} c_i}{\sum_{i=1}^{n} (X_i - \overline{X})^2} - \frac{1}{\sum_{i=1}^{n} (X_i - \overline{X})^2} = 0
\end{aligned}$$

since $\sum_{i=1}^{n} c_i = 0$ and $\sum_{i=1}^{n} c_i X_i = 1$, as shown above.

Theorem 1.11 (continued 5)

Proof (continued). So the variance of $\hat{\beta}_1$ becomes

$$\sigma^2\{\hat{\beta}_1\} = \sigma^2 \left(\sum_{i=1}^{n} k_i^2 + \sum_{i=1}^{n} d_i^2 \right).$$

We have $\sigma^2\{b_1\} = \sigma^2 \sum_{i=1}^{n} k_i^2$ as shown above. Therefore $\sigma^2\{\hat{\beta}_1\} = \sigma^2\{b_1\} + \sum_{i=1}^{n} d_i^2$. So the variance of $\hat{\beta}_1$ is minimized with all $d_i = 0$, and hence when $c_i = k_i$ for all $i = 1, 2, \ldots, n$. That is, the variance of $\hat{\beta}_1$ is minimized when $\hat{\beta}_1 = b_1$, as claimed.

In Exercise 2.51 it is to be shown that b_0 is an unbiased estimator of β_0 (as claimed): $E\{b_0\} = \beta_0$.

Theorem 1.11 (continued 5)

Proof (continued). So the variance of $\hat{\beta}_1$ becomes

$$\sigma^2\{\hat{\beta}_1\} = \sigma^2 \left(\sum_{i=1}^{n} k_i^2 + \sum_{i=1}^{n} d_i^2 \right).$$

We have $\sigma^2\{b_1\} = \sigma^2 \sum_{i=1}^{n} k_i^2$ as shown above. Therefore $\sigma^2\{\hat{\beta}_1\} = \sigma^2\{b_1\} + \sum_{i=1}^{n} d_i^2$. So the variance of $\hat{\beta}_1$ is minimized with all $d_i = 0$, and hence when $c_i = k_i$ for all $i = 1, 2, \ldots, n$. That is, the variance of $\hat{\beta}_1$ is minimized when $\hat{\beta}_1 = b_1$, as claimed.

In Exercise 2.51 it is to be shown that b_0 is an unbiased estimator of β_0 (as claimed): $E\{b_0\} = \beta_0$.