8.26 The inner product of the first row of $\mathbf{X}'$ with $\mathbf{e}$ is

$$[1\ 1\ \cdots\ 1]\mathbf{e} = \sum_{i=1}^{n} e_i.$$

Setting this equal to zero yields the orthogonality condition (8.9a) of the text. Substituting in the definition (8.6) of the least squares residuals, this condition is

$$\sum_{i=1}^{n}(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_K X_{Ki}) = 0.$$

Applying the summation to each term, dividing through by n, and solving for $\hat{\beta}_1$ yields the intercept estimator (8.11):

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \cdots - \hat{\beta}_K \bar{X}_K.$$

8.27 (a) Starting with (8.69b),

$$\sigma^2_{\hat{\beta}_2} = \frac{n\sigma^2}{n\sum X_i^2 - (\sum X_i)^2} = \frac{n\sigma^2}{n\sum X_i^2 - n^2\bar{X}^2} = \frac{\sigma^2}{\sum x_i^2}.$$

The last step cancels n and uses the result of Exercise A.7.

(b) It is easiest to start with (4.27) and put the expression inside parentheses over a common denominator:

$$V(\hat{\alpha}) = \sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right) = \sigma^2\left(\frac{\sum x_i^2 + n\bar{X}^2}{n\sum x_i^2}\right) = \frac{\sigma^2 \sum X_i^2}{n\sum X_i^2 - (\sum X_i)^2}.$$

This is (8.69a). Again, the last step uses the result of Exercise A.7.

Examination Questions

1. Labor economists study the determination of labor earnings using a "statistical earnings function."

   (a) A simple example of such a regression, estimated using data for 31,093 men, is

   $$\log Y_i = 7.58 + \underset{(0.00160)}{0.070}\, X_i + e_i.$$

   Here Y denotes earnings and X is years of education; "log" denotes a natural logarithm. The value in parentheses is the estimated standard error.

   i. Using your knowledge of logarithmic functional forms, explain the interpretation of the coefficient on education.
   ii. What does this equation predict would be the earnings of a hypothetical person with no education?
   iii. Obtain a 90% confidence interval for the rate of return to education.

iv. Use a hypothesis test to establish whether these estimation results constitute compelling evidence that there is a positive return to education. State the elements of your test carefully.

(b) Consider expanding the above regression to include experience Z and its square:

$$\log Y_i = 6.20 + \underset{(0.00148)}{0.107} X_i + \underset{(0.00107)}{0.081} Z_i - \underset{(0.0000215)}{0.0012} Z_i^2 + e_i.$$

i. At what level of experience do earnings peak?

ii. Is the quadratic term on experience statistically significant? State the elements of your test procedure carefully.

(c) The following R^2 values were reported for the above regressions: 0.067 and 0.285.

i. Which of these R^2 is for which regression?

ii. Consider the hypothesis that experience is not an important determinant of earnings. Present the standard form of the test statistic that would be used to test this hypothesis, including its distribution.

iii. Show how this test statistic can be computed using the R^2's. Use this to test the hypothesis, stating carefully the elements of your test procedure.

2. Suppose that the true population model is

$$Y_i = \beta_1 + \beta_2 X_{2i} + \varepsilon_i,$$

but we mistakenly include in the regression an irrelevant regressor X_3, obtaining the sample regression line

$$Y_i = b_1 + b_2 X_{2i} + b_3 X_{3i} + e_i.$$

The slope estimates are given by the usual least squares formulas

$$\hat{\beta}_2 = \frac{\sum x_{3i}^2 \sum x_{2i} y_i - \sum x_{2i} x_{3i} \sum x_{3i} y_i}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i} x_{3i})^2}$$
$$\hat{\beta}_3 = \frac{\sum x_{2i}^2 \sum x_{3i} y_i - \sum x_{2i} x_{3i} \sum x_{2i} y_i}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i} x_{3i})^2}.$$

Answer one and only one of the following questions.

(a) Derive $E(b_1)$.

(b) Derive $E(b_2)$.

(c) Derive $E(b_3)$.