8 Multiple Regression: Tests of Hypotheses and Confidence Intervals

In this chapter we consider hypothesis tests and confidence intervals for the parameters $\beta_0, \beta_1, \ldots, \beta_k$ in $\boldsymbol{\beta}$ in the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. We also provide a confidence interval for $\sigma^2 = \text{var}(y_i)$. We will assume throughout the chapter that $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where $\mathbf{X}$ is $n \times (k+1)$ of rank $k+1 < n$.

8.1 TEST OF OVERALL REGRESSION

We noted in Section 7.9 that the problems associated with both overfitting and underfitting motivate us to seek an optimal model. Hypothesis testing is a formal tool for, among other things, choosing between a reduced model and an associated full model. The hypothesis H_0, expresses the reduced model in terms of values of a subset of the β_j's in $\boldsymbol{\beta}$. The alternative hypothesis, H_1, is associated with the full model.

To illustrate this tool we begin with a common test, the test of the overall regression hypothesis that none of the x variables predict y. This hypothesis (leading to the reduced model) can be expressed as $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$, where $\boldsymbol{\beta}_1 = (\beta_1, \beta_2, \ldots, \beta_k)'$. Note that we wish to test $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$, not $H_0 : \boldsymbol{\beta} = \mathbf{0}$, where

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \boldsymbol{\beta}_1 \end{pmatrix}.$$

Since β_0 is usually not zero, we would rarely be interested in including $\beta_0 = 0$ in the hypothesis. Rejection of H_0: $\boldsymbol{\beta} = \mathbf{0}$ might be due solely to β_0, and we would not learn whether the x variables predict y. For a test of $H_0 : \boldsymbol{\beta} = \mathbf{0}$, see Problem 8.6.

We proceed by proposing a test statistic that is distributed as a central F if H_0 is true and as a noncentral F otherwise. Our approach to obtaining a test statistic is somewhat

Linear Models in Statistics, *Second Edition*, by Alvin C. Rencher and G. Bruce Schaalje

simplified if we use the centered model (7.32)

$$\mathbf{y} = (\mathbf{j}, \mathbf{X}_c)\begin{pmatrix} \alpha \\ \boldsymbol{\beta}_1 \end{pmatrix} + \boldsymbol{\varepsilon},$$

where $\mathbf{X}_c = [\mathbf{I} - (1/n)\mathbf{J}]\mathbf{X}_1$ is the centered matrix [see (7.33)] and $\mathbf{X}_1$ contains all the columns of $\mathbf{X}$ except the first [see (7.19)]. The corrected total sum of squares $\mathrm{SST} = \sum_{i=1}^n (y_i - \bar{y})^2$ can be partitioned as

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \hat{\boldsymbol{\beta}}_1' \mathbf{X}_c' \mathbf{y} + \left[\textstyle\sum_{i=1}^n (y_i - \bar{y})^2 - \hat{\boldsymbol{\beta}}_1' \mathbf{X}_c' \mathbf{y} \right] \quad \text{[by (7.53)]}$$

$$= \hat{\boldsymbol{\beta}}_1' \mathbf{X}_c' \mathbf{X}_c \hat{\boldsymbol{\beta}}_1 + \mathrm{SSE} = \mathrm{SSR} + \mathrm{SSE} \quad \text{[by (7.54)]}, \tag{8.1}$$

where SSE is as given in (7.39). The regression sum of squares $\mathrm{SSR} = \hat{\boldsymbol{\beta}}_1' \mathbf{X}_c' \mathbf{X}_c \hat{\boldsymbol{\beta}}_1$ is clearly due to $\boldsymbol{\beta}_1$.

In order to construct an F test, we first express the sums of squares in (8.1) as quadratic forms in $\mathbf{y}$ so that we can use theorems from Chapter 5 to show that SSR and SSE have chi-square distributions and are independent. Using $\sum_i (y_i - \bar{y})^2 = \mathbf{y}'[\mathbf{I} - (1/n)\mathbf{J}]\mathbf{y}$ in (5.2), $\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_c'\mathbf{y}$ in (7.37), and $\mathrm{SSE} = \sum_{i=1}^n (y_i - \bar{y})^2 - \hat{\boldsymbol{\beta}}_1' \mathbf{X}_c' \mathbf{y}$ in (7.39), we can write (8.1) as

$$\mathbf{y}'\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\mathbf{y} = \mathrm{SSR} + \mathrm{SSE}$$

$$= \mathbf{y}'\mathbf{X}_c(\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_c'\mathbf{y} + \mathbf{y}'\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\mathbf{y} - \mathbf{y}'\mathbf{X}_c(\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_c'\mathbf{y}$$

$$= \mathbf{y}'\mathbf{H}_c\mathbf{y} + \mathbf{y}'\left(\mathbf{I} - \frac{1}{n}\mathbf{J} - \mathbf{H}_c\right)\mathbf{y}, \tag{8.2}$$

where $\mathbf{H}_c = \mathbf{X}_c(\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_c'$.

In the following theorem we establish some properties of the three matrices of the quadratic forms in (8.2).

Theorem 8.1a. The matrices $\mathbf{I} - (1/n)\,\mathbf{J}$, $\mathbf{H}_c = \mathbf{X}_c(\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_c'$, and $\mathbf{I} - (1/n)\,\mathbf{J} - \mathbf{H}_c$ have the following properties:

(i) $\mathbf{H}_c[\mathbf{I} - (1/n)\,\mathbf{J}] = \mathbf{H}_c$. (8.3)

(ii) $\mathbf{H}_c$ is idempotent of rank k.

(iii) $\mathbf{I} - (1/n)\,\mathbf{J} - \mathbf{H}_c$ is idempotent of rank $n - k - 1$.

(iv) $\mathbf{H}_c[\mathbf{I} - (1/n)\,\mathbf{J} - \mathbf{H}_c] = \mathbf{O}$. (8.4)

Proof. Part (i) follows from $\mathbf{X}_c'\mathbf{j} = \mathbf{0}$, which was established in Problem 7.16. Part (ii) can be shown by direct multiplication. Parts (iii) and (iv) follow from (i) and (ii). □

The distributions of SSR/σ^2 and SSE/σ^2 are given in the following theorem.

Theorem 8.1b. If $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, then $\mathrm{SSR}/\sigma^2 = \hat{\boldsymbol{\beta}}_1'\mathbf{X}_c'\mathbf{X}_c\hat{\boldsymbol{\beta}}_1/\sigma^2$ and $\mathrm{SSE}/\sigma^2 = \left[\sum_{i=1}^n (y_i - \bar{y})^2 - \hat{\boldsymbol{\beta}}_1'\mathbf{X}_c'\mathbf{X}_c\hat{\boldsymbol{\beta}}_1\right]/\sigma^2$ have the following distributions:

(i) SSR/σ^2 is $\chi^2(k, \lambda_1)$, where $\lambda_1 = \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}/2\sigma^2 = \boldsymbol{\beta}_1'\mathbf{X}_c'\mathbf{X}_c\boldsymbol{\beta}_1/2\sigma^2$.
(ii) SSE/σ^2 is $\chi^2(n - k - 1)$.

Proof. These results follow from (8.2), Theorem 8.1a(ii) and (iii), and Corollary 2 to Theorem 5.5. □

The independence of SSR and SSE is demonstrated in the following theorem.

Theorem 8.1c. If $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, then SSR and SSE are independent, where SSR and SSE are defined in (8.1) and (8.2).

Proof. This follows from Theorem 8.1a(iv) and Corollary 1 to Theorem 5.6b. □

We can now establish an F test for H_0: $\boldsymbol{\beta}_1 = \mathbf{0}$ versus H_1: $\boldsymbol{\beta}_1 \neq \mathbf{0}$.

Theorem 8.1d. If $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, the distribution of

$$F = \frac{\mathrm{SSR}/(k\sigma^2)}{\mathrm{SSE}/[(n-k-1)\sigma^2]} = \frac{\mathrm{SSR}/k}{\mathrm{SSE}/(n-k-1)} \tag{8.5}$$

is as follows:

(i) If $H_0: \boldsymbol{\beta}_1 = \mathbf{0}$ is false, then

$$F \text{ is distributed as } F(k, n-k-1, \lambda_1),$$

where $\lambda_1 = \boldsymbol{\beta}_1'\mathbf{X}_c'\mathbf{X}_c\boldsymbol{\beta}_1/2\sigma^2$.
(ii) If $H_0: \boldsymbol{\beta}_1 = \mathbf{0}$ is true, then $\lambda_1 = 0$ and

$$F \text{ is distributed as } F(k, n-k-1).$$

Proof

(i) This result follows from (5.30) and Theorems 8.1b and 8.1c.
(ii) This result follows from (5.28) and Theorems 8.1b and 8.1c. □

Note that $\lambda_1 = 0$ if and only if $\boldsymbol{\beta}_1 = \mathbf{0}$, since $\mathbf{X}_c'\mathbf{X}_c$ is positive definite (see Corollary 1 to Theorem 2.6b).

TABLE 8.1 ANOVA Table for the F Test of $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$

Source of Variation	df	Sum of Squares	Mean Square	Expected Mean Square
Due to $\boldsymbol{\beta}_1$	k	$\text{SSR} = \hat{\boldsymbol{\beta}}_1'\mathbf{X}_c'\mathbf{y} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{y}^2$	SSR/k	$\sigma^2 + \frac{1}{k}\boldsymbol{\beta}_1'\mathbf{X}_c'\mathbf{X}_c\boldsymbol{\beta}_1$
Error	$n-k-1$	$\text{SSE} = \sum_i (y_i - \bar{y})^2 - \hat{\boldsymbol{\beta}}_1'\mathbf{X}_c'\mathbf{y} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$	$\text{SSE}/(n-k-1)$	σ^2
Total	$n-1$	$\text{SST} = \sum_i (y_i - \bar{y})^2$		

The test for $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ is carried out as follows. Reject H_0 if $F \geq F_{\alpha,k,n-k-1}$, where $F_{\alpha,k,n-k-1}$ is the upper α percentage point of the (central) F distribution. Alternatively, a p value can be used to carry out the test. A p value is the tail area of the central F distribution beyond the calculated F value, that is, the probability of exceeding the calculated F value, assuming $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ to be true. A p value less than α is equivalent to $F > F_{\alpha,k,n-k-1}$.

The analysis-of-variance (ANOVA) table (Table 8.1) summarizes the results and calculations leading to the overall F test. Mean squares are sums of squares divided by the degrees of freedom of the associated chi-square (χ^2) distributions.

The entries in the column for expected mean squares in Table 8.1 are simply $E(\text{SSR}/k)$ and $E[\text{SSE}/(n-k-1)]$. The first of these can be established by Theorem 5.2a or by (5.20). The second was established by Theorem 7.3f.

If $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ is true, both of the expected mean squares in Table 8.1 are equal to σ^2, and we expect F to be near 1. If $\boldsymbol{\beta}_1 \neq \mathbf{0}$, then $E(\text{SSR}/k) > \sigma^2$ since $\mathbf{X}_c'\mathbf{X}_c$ is positive definite, and we expect F to exceed 1. We therefore reject H_0 for large values of F.

The test of $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ in Table 8.1 has been developed using the centered model (7.32). We can also express SSR and SSE in terms of the noncentered model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ in (7.4):

$$\text{SSR} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{y}^2, \quad \text{SSE} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}. \tag{8.6}$$

These are the same as SSR and SSE in (8.1) [see (7.24), (7.39), (7.54), and Problems 7.19, 7.25].

Example 8.1. Using the data in Table 7.1, we illustrate the test of $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ where, in this case, $\boldsymbol{\beta}_1 = (\beta_1, \beta_2)'$. In Example 7.3.1(a), we found $\mathbf{X}'\mathbf{y} = (90, 482, 872)'$ and $\hat{\boldsymbol{\beta}} = (5.3754, \ 3.0118, -1.2855)'$. The quantities $\mathbf{y}'\mathbf{y}$, $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$, and $n\bar{y}^2$ are given by

$$\mathbf{y}'\mathbf{y} = \sum_{i=1}^{12} y_i^2 = 2^2 + 3^2 + \cdots + 14^2 = 840,$$

$$\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = (5.3754, 3.0118, -1.2855)\begin{pmatrix} 90 \\ 482 \\ 872 \end{pmatrix} = 814.5410,$$

TABLE 8.2 ANOVA for Overall Regression Test for Data in Table 7.1

Source	df	SS	MS	F
Due to β_1	2	139.5410	69.7705	24.665
Error	9	25.4590	2.8288	
Total	11	165.0000		

$$n\bar{y}^2 = n\left(\frac{\sum_i y_i}{n}\right)^2 = 12\left(\frac{90}{12}\right)^2 = 675.$$

Thus, by (8.6), we obtain

$$\begin{aligned}
\mathrm{SSR} &= \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{y}^2 = 139.5410, \\
\mathrm{SSE} &= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = 25.4590, \\
\sum_{i=1}^{n}(y_i - \bar{y})^2 &= \mathbf{y}'\mathbf{y} - n\bar{y}^2 = 165.
\end{aligned}$$

The F test is given in Table 8.2. Since $24.665 > F_{.05,2,9} = 4.26$, we reject $H_0: \boldsymbol{\beta}_1 = \mathbf{0}$ and conclude that at least one of β_1 or β_2 is not zero. The p value is .000223. □

8.2 TEST ON A SUBSET OF THE β'S

In more generality, suppose that we wish to test the hypothesis that a subset of the x's is not useful in predicting y. A simple example is $H_0: \beta_j = 0$ for a single β_j. If H_0 is rejected, we would retain $\beta_j x_j$ in the model. As another illustration, consider the model in (7.2)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \varepsilon,$$

for which we may wish to test the hypothesis $H_0: \beta_3 = \beta_4 = \beta_5 = 0$. If H_0 is rejected, we would choose the full second-order model over the reduced first-order model.

Without loss of generality, we assume that the β's to be tested have been arranged last in $\boldsymbol{\beta}$, with a corresponding arrangement of the columns of $\mathbf{X}$. Then $\boldsymbol{\beta}$ and $\mathbf{X}$ can be partitioned accordingly, and by (7.78), the model for all n observations becomes

$$\begin{aligned}
\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} &= (\mathbf{X}_1, \mathbf{X}_2)\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \boldsymbol{\varepsilon} \\
&= \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}, \qquad (8.7)
\end{aligned}$$

where $\boldsymbol{\beta}_2$ contains the β's to be tested. The intercept β_0 would ordinarily be included in $\boldsymbol{\beta}_1$.

The hypothesis of interest is $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$. If we designate the number of parameters in $\boldsymbol{\beta}_2$ by h, then $\mathbf{X}_2$ is $n \times h$, $\boldsymbol{\beta}_1$ is $(k-h+1) \times 1$, and $\mathbf{X}_1$ is $n \times (k-h+1)$. Thus $\boldsymbol{\beta}_1 = (\beta_0, \beta_1, \cdots, \beta_{k-h})'$ and $\boldsymbol{\beta}_2 = (\beta_{k-h+1}, \cdots, \beta_k)'$. In terms of the illustration at the beginning of this section, we would have $\boldsymbol{\beta}_1 = (\beta_0, \beta_1, \beta_2)'$ and $\boldsymbol{\beta}_2 = (\beta_3, \beta_4, \beta_5)'$. Note that $\boldsymbol{\beta}_1$ in (8.7) is different from $\boldsymbol{\beta}_1$ in Section 8.1, in which $\boldsymbol{\beta}$ was partitioned as $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \boldsymbol{\beta}_1 \end{pmatrix}$ and $\boldsymbol{\beta}_1$ constituted all of $\boldsymbol{\beta}$ except β_0.

To test $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ versus $H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$, we use a full–reduced-model approach. The full model is given by (8.7). Under $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$, the reduced model becomes

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1^* + \boldsymbol{\varepsilon}^*. \tag{8.8}$$

We use the notation $\boldsymbol{\beta}_1^*$ and $\boldsymbol{\varepsilon}^*$ as in Section 7.9, because in the reduced model, $\boldsymbol{\beta}_1^*$ and $\boldsymbol{\varepsilon}^*$ will typically be different from $\boldsymbol{\beta}_1$ and $\boldsymbol{\varepsilon}$ in the full model (unless $\mathbf{X}_1$ and $\mathbf{X}_2$ are orthogonal; see Theorem 7.9a and its corollary). The estimator of $\boldsymbol{\beta}_1^*$ in the reduced model (8.8) is $\hat{\boldsymbol{\beta}}_1^* = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}$, which is, in general, not the same as the first $k-h+1$ elements of $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ from the full model (8.7) (unless $\mathbf{X}_1$ and $\mathbf{X}_2$ are orthogonal; see Theorem 7.10).

In order to compare the fit of the full model (8.7) to the fit of the reduced model (8.8), we add and subtract $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$ and $\hat{\boldsymbol{\beta}}_1^{*\prime}\mathbf{X}_1'\mathbf{y}$ to the total corrected sum of squares $\sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})^2 = \mathbf{y}'\mathbf{y} - n\bar{y}^2$ so as to obtain the partitioning

$$\mathbf{y}'\mathbf{y} - n\bar{y}^2 = (\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}\mathbf{y}) + (\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \hat{\boldsymbol{\beta}}_1^{*\prime}\mathbf{X}_1'\mathbf{y}) + (\hat{\boldsymbol{\beta}}_1^{*\prime}\mathbf{X}_1\mathbf{y} - n\bar{y}^2) \tag{8.9}$$

or

$$\text{SST} = \text{SSE} + \text{SS}(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) + \text{SSR(reduced)}, \tag{8.10}$$

where $\text{SS}(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \hat{\boldsymbol{\beta}}_1^{*\prime}\mathbf{X}_1'\mathbf{y}$ is the "extra" regression sum of squares due to $\boldsymbol{\beta}_2$ after adjusting for $\boldsymbol{\beta}_1$. Note that $\text{SS}(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)$ can also be expressed as

$$\begin{aligned} \text{SS}(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) &= \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{y}^2 - (\hat{\boldsymbol{\beta}}_1^{*\prime}\mathbf{X}_1'\mathbf{y} - n\bar{y}^2) \\ &= \text{SSR(full)} - \text{SSR(reduced)}, \end{aligned}$$

which is the difference between the overall regression sum of squares for the full model and the overall regression sum of squares for the reduced model [see (8.6)].

If $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ is true, we would expect $\text{SS}(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)$ to be small so that SST in (8.10) is composed mostly of SSR(reduced) and SSE. If $\boldsymbol{\beta}_2 \neq \mathbf{0}$, we expect $\text{SS}(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)$ to be larger and account for more of SST. Thus we are testing $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ in the full model in which there are no restrictions on $\boldsymbol{\beta}_1$. We are not *ignoring* $\boldsymbol{\beta}_1$ (assuming $\boldsymbol{\beta}_1 = \mathbf{0}$) but are testing $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ *in the presence of* $\boldsymbol{\beta}_1$, that is, above and beyond whatever $\boldsymbol{\beta}_1$ contributes to SST.

To develop a test statistic based on $\text{SS}(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)$, we first write (8.9) in terms of quadratic forms in $\mathbf{y}$. Using $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ and $\hat{\boldsymbol{\beta}}_1^* = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}$ and (5.2), (8.9) becomes

$$\begin{aligned}
\mathbf{y}'\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\mathbf{y} &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} + \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&\quad - \mathbf{y}'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y} + \mathbf{y}'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y} - \mathbf{y}'\frac{1}{n}\mathbf{J}\mathbf{y} \\
&= \mathbf{y}'\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{y} + \mathbf{y}'[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1']\mathbf{y} \\
&\quad + \mathbf{y}'\left[\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1' - \frac{1}{n}\mathbf{J}\right]\mathbf{y} \qquad (8.11) \\
&= \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} + \mathbf{y}'(\mathbf{H} - \mathbf{H}_1)\mathbf{y} + \mathbf{y}'\left(\mathbf{H}_1 - \frac{1}{n}\mathbf{J}\right)\mathbf{y}, \qquad (8.12)
\end{aligned}$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{H}_1 = \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$. The matrix $\mathbf{I} - \mathbf{H}$ was shown to be idempotent in Problem 5.32a, with rank $n - k - 1$, where $k + 1$ is the rank of $\mathbf{X}$ ($k + 1$ is also the number of elements in $\boldsymbol{\beta}$). The matrix $\mathbf{H} - \mathbf{H}_1$ is shown to be idempotent in the following theorem.

Theorem 8.2a. The matrix $\mathbf{H} - \mathbf{H}_1 = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$ is idempotent with rank h, where h is the number of elements in $\boldsymbol{\beta}_2$.

PROOF. Premultiplying $\mathbf{X}$ by $\mathbf{H}$, we obtain

$$\mathbf{H}\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{X}$$

or

$$\mathbf{X} = \left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{X}. \qquad (8.13)$$

Partitioning $\mathbf{X}$ on the left side of (8.13) and the last $\mathbf{X}$ on the right side, we obtain [by an extension of (2.28)]

$$\begin{aligned}
(\mathbf{X}_1, \mathbf{X}_2) &= \left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right](\mathbf{X}_1, \mathbf{X}_2) \\
&= \left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}_1,\ \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}_2\right].
\end{aligned}$$

Thus

$$\begin{aligned}
\mathbf{X}_1 &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}_1, \\
\mathbf{X}_2 &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}_2.
\end{aligned} \qquad (8.14)$$

Simplifying $\mathbf{HH}_1$ and $\mathbf{H}_1\mathbf{H}$ by (8.14) and its transpose, we obtain

$$\mathbf{HH}_1 = \mathbf{H}_1 \quad \text{and} \quad \mathbf{H}_1\mathbf{H} = \mathbf{H}_1. \tag{8.15}$$

The matrices $\mathbf{H}$ and $\mathbf{H}_1$ are idempotent (see Problem 5.32). Thus

$$\begin{aligned}(\mathbf{H} - \mathbf{H}_1)^2 &= \mathbf{H}^2 - \mathbf{HH}_1 - \mathbf{H}_1\mathbf{H} + \mathbf{H}_1^2 \\ &= \mathbf{H} - \mathbf{H}_1 - \mathbf{H}_1 + \mathbf{H}_1 \\ &= \mathbf{H} - \mathbf{H}_1,\end{aligned}$$

and $\mathbf{H} - \mathbf{H}_1$ is idempotent. For the rank of $\mathbf{H} - \mathbf{H}_1$, we have (by Theorem 2.13d)

$$\begin{aligned}\text{rank}(\mathbf{H} - \mathbf{H}_1) &= \text{tr}(\mathbf{H} - \mathbf{H}_1) \\ &= \text{tr}(\mathbf{H}) - \text{tr}(\mathbf{H}_1) && \text{[by (2.86)]} \\ &= \text{tr}\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right] - \text{tr}\left[\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\right] \\ &= \text{tr}\left[\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right] - \text{tr}\left[\mathbf{X}_1'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\right] && \text{[by (2.87)]} \\ &= \text{tr}(\mathbf{I}_{k+1}) - \text{tr}(\mathbf{I}_{k-h+1}) = k + 1 - (k - h + 1) = h.\end{aligned}$$

□

We now find the distributions of $\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}$ and $\mathbf{y}'(\mathbf{H} - \mathbf{H}_1)\mathbf{y}$ in (8.12) and show that they are independent.

Theorem 8.2b. If $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ and $\mathbf{H}$ and $\mathbf{H}_1$ are as defined in (8.11) and (8.12), then

(i) $\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}/\sigma^2$ is $\chi^2(n - k - 1)$.
(ii) $\mathbf{y}'(\mathbf{H} - \mathbf{H}_1)\mathbf{y}/\sigma^2$ is $\chi^2(h, \lambda_1)$, $\lambda_1 = \boldsymbol{\beta}_2'\left[\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\right]\boldsymbol{\beta}_2/2\sigma^2$.
(iii) $\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}$ and $\mathbf{y}'(\mathbf{H} - \mathbf{H}_1)\mathbf{y}$ are independent.

PROOF. Adding $\mathbf{y}'(1/n)\mathbf{J}\mathbf{y}$ to both sides of (8.12), we obtain the decomposition $\mathbf{y}'\mathbf{y} = \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} + \mathbf{y}'(\mathbf{H} - \mathbf{H}_1)\mathbf{y} + \mathbf{y}'\mathbf{H}_1\mathbf{y}$. The matrices $\mathbf{I} - \mathbf{H}$, $\mathbf{H} - \mathbf{H}_1$, and $\mathbf{H}_1$ were shown to be idempotent in Problem 5.32 and Theorem 8.2a. Hence by Corollary 1 to Theorem 5.6c, all parts of the theorem follow. See Problem 8.9 for the derivation of λ_1. □

If $\lambda_1 = 0$ in Theorem 8.2b(ii), then $\mathbf{y}'(\mathbf{H} - \mathbf{H}_1)\mathbf{y}/\sigma^2$ has the central chi-square distribution $\chi^2(h)$. Since $\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$ is positive definite (see Problem 8), $\lambda_1 = 0$ if and only if $\boldsymbol{\beta}_2 = \mathbf{0}$.

An F test for $H_0: \boldsymbol{\beta}_2 = \mathbf{0}$ versus $H_1: \boldsymbol{\beta}_2 \neq \mathbf{0}$ is given in the following theorem.

Theorem 8.2c. Let $\mathbf{y}$ be $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ and define an F statistic as follows:

$$F = \frac{\mathbf{y}'(\mathbf{H} - \mathbf{H}_1)\mathbf{y}/h}{\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}/(n-k-1)} = \frac{\text{SS}(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)/h}{\text{SSE}/(n-k-1)} \tag{8.16}$$

$$= \frac{(\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \hat{\boldsymbol{\beta}}_1^{*\prime}\mathbf{X}_1'\mathbf{y})/h}{(\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y})/(n-k-1)}, \tag{8.17}$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is from the full model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and $\hat{\boldsymbol{\beta}}_1^* = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}$ is from the reduced model $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1^* + \boldsymbol{\varepsilon}^*$. The distribution of F in (8.17) is as follows:

(i) If $H_0: \boldsymbol{\beta}_2 = \mathbf{0}$ is false, then

$$F \text{ is distributed as } F(h, n-k-1, \lambda_1),$$

where $\lambda_1 = \boldsymbol{\beta}_2'\left[\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\right]\boldsymbol{\beta}_2/2\sigma^2$.

(ii) If $H_0: \boldsymbol{\beta}_2 = \mathbf{0}$ is true, then $\lambda_1 = 0$ and

$$F \text{ is distributed as } F(h, n-k-1).$$

PROOF

(i) This result follows from (5.30) and Theorem 8.2b.

(ii) This result follows from (5.28) and Theorem 8.2b. □

The test for $H_0: \boldsymbol{\beta}_2 = \mathbf{0}$ is carried out as follows: Reject H_0 if $F \geq F_{\alpha,h,n-k-1}$, where $F_{\alpha,h,n-k-1}$ is the upper α percentage point of the (central) F distribution. Alternatively, we reject H_0 if $p < \alpha$, where p is the p value. Since $\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$ is positive definite (see Problem 8.10), $\lambda_1 > 0$ if $H_0: \boldsymbol{\beta}_2 = \mathbf{0}$ is false. This justifies rejection of H_0 for large values of F.

Results and calculations leading to this F test are summarized in the ANOVA table (Table 8.3), where $\boldsymbol{\beta}_1$ is $(k-h+1) \times 1$, $\boldsymbol{\beta}_2$ is $h \times 1$, $\mathbf{X}_1$ is $n \times (k-h+1)$, and $\mathbf{X}_2$ is $n \times h$.

The entries in the column for expected mean squares are $E[\text{SS}(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)/h]$ and $E[\text{SSE}/(n-k-1)]$. For $E[\text{SS}(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)/h]$, see Problem 8.11. Note that if H_0 is true, both expected mean squares (Table 8.3) are equal to σ^2, and if H_0 is false, $E[\text{SS}(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)/h] > E[\text{SSE}/(n-k-1)]$ since $\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$ is positive definite. This inequality provides another justification for rejecting H_0 for large values of F.

TABLE 8.3 ANOVA Table for F-Test of $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$

Source of Variation	df	Sum of Squares	Mean Square	Expected Mean Square
Due to $\boldsymbol{\beta}_2$ adjusted for $\boldsymbol{\beta}_1$	h	$\mathrm{SS}(\beta_2\vert\beta_1) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \hat{\boldsymbol{\beta}}_1^{*}\mathbf{X}_1'\mathbf{y}$	$\mathrm{SS}(\boldsymbol{\beta}_2\vert\beta_1)/h$	$\sigma^2 + \frac{1}{h}\boldsymbol{\beta}_2'[\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2]\boldsymbol{\beta}_2$
Error	$n - k - 1$	$\mathrm{SSE} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$	$\mathrm{SSE}/(n - k - 1)$	σ^2
Total	$n - 1$	$\mathrm{SST} = \mathbf{y}'\mathbf{y} - n\bar{y}^2$		

Example 8.2a. Consider the dependent variable y_2 in the chemical reaction data in Table 7.4 (see Problem 7.52 for a description of the variables). In order to check the usefulness of second-order terms in predicting y_2, we use as a full model, $y_2 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1^2 + \beta_5 x_2^2 + \beta_6 x_3^2 + \beta_7 x_1 x_2 + \beta_8 x_1 x_3 + \beta_9 x_2 x_3 + \varepsilon$, and test $H_0: \beta_4 = \beta_5 = \ldots = \beta_9 = 0$. For the full model, we obtain $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{y}^2 = 339.7888$, and for the reduced model $y_2 = \beta_0^* + \beta_1^* x_1 + \beta_2^* x_2 + \beta_3^* x_3 + \varepsilon*$, we have $\hat{\boldsymbol{\beta}}_1^{*'}\mathbf{X}_1'\mathbf{y} - n\bar{y}^2 = 151.0022$. The difference is $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \hat{\boldsymbol{\beta}}_1^{*'}\mathbf{X}_1'\mathbf{y} = 188.7866$. The error sum of squares is SSE $= 60.6755$, and the F statistic is given by (8.16) or Table 8.3 as

$$F = \frac{188.7866/6}{60.6755/9} = \frac{31.4644}{6.7417} = 4.6671,$$

which has a p value of .0198. Thus the second-order terms are useful in prediction of y_2. In fact, the overall F in (8.5) for the reduced model is 3.027 with $p = .0623$, so that x_1, x_2, and x_3 are inadequate for predicting y_2. The overall F for the full model is 5.600 with $p = .0086$. □

In the following theorem, we express $\mathrm{SS}(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)$ as a quadratic form in $\hat{\boldsymbol{\beta}}_2$ that corresponds to λ_1 in Theorem 8.2b(ii).

Theorem 8.2d. If the model is partitioned as in (8.7), then $\mathrm{SS}(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \hat{\boldsymbol{\beta}}_1^{*'}\mathbf{X}_1'\mathbf{y}$ can be written as

$$\mathrm{SS}(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = \hat{\boldsymbol{\beta}}_2'\left[\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\right]\hat{\boldsymbol{\beta}}_2, \tag{8.18}$$

where $\hat{\boldsymbol{\beta}}_2$ is from a partitioning of $\hat{\boldsymbol{\beta}}$ in the full model:

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \tag{8.19}$$

Proof. We can write $\mathbf{X}\hat{\boldsymbol{\beta}}$ in terms of $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ as $\mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{X}_1, \mathbf{X}_2)\begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} = \mathbf{X}_1\hat{\boldsymbol{\beta}}_1 + \mathbf{X}_2\hat{\boldsymbol{\beta}}_2$. To write $\hat{\boldsymbol{\beta}}_1^*$ in terms of $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$, we note that by (7.80), $E(\hat{\boldsymbol{\beta}}_1^*) = \boldsymbol{\beta}_1 + \mathbf{A}\boldsymbol{\beta}_2$, where $\mathbf{A} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$ is the alias matrix defined in Theorem 7.9a. This can be estimated by $\hat{\boldsymbol{\beta}}_1^* = \hat{\boldsymbol{\beta}}_1 + \mathbf{A}\hat{\boldsymbol{\beta}}_2$, where $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ are from the full model, as in (8.19). Then $\mathrm{SS}(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)$ in (8.10) or Table 8.3 can be written as

$$\begin{aligned}
\mathrm{SS}(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) &= \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \hat{\boldsymbol{\beta}}_1^{*'}\mathbf{X}_1'\mathbf{y} \\
&= \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_1^{*'}\mathbf{X}_1'\mathbf{X}_1\hat{\boldsymbol{\beta}}^* \quad \text{[by (7.8)]} \\
&= (\hat{\boldsymbol{\beta}}_1'\mathbf{X}_1' + \hat{\boldsymbol{\beta}}_2'\mathbf{X}_2')(\mathbf{X}_1\hat{\boldsymbol{\beta}}_1 + \mathbf{X}_2\hat{\boldsymbol{\beta}}_2) - (\hat{\boldsymbol{\beta}}_1' + \hat{\boldsymbol{\beta}}_2'\mathbf{A}')\mathbf{X}_1'\mathbf{X}_1(\hat{\boldsymbol{\beta}}_1 + \mathbf{A}\hat{\boldsymbol{\beta}}_2).
\end{aligned}$$

Multiplying this out and substituting $(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$ for $\mathbf{A}$, we obtain (8.18). □

In (8.18), it is clear that $\text{SS}(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)$ is due to $\boldsymbol{\beta}_2$. We also see in (8.18) a direct correspondence between $\text{SS}(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)$ and the noncentrality parameter λ_1 in Theorem 8.2b (ii) or the expected mean square in Table 8.3.

Example 8.2b. The full–reduced-model test of $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ in Table 8.3 can be used to test for significance of a single $\hat{\beta}_j$. To illustrate, suppose that we wish to test $H_0 : \beta_k = 0$, where $\boldsymbol{\beta}$ is partitioned as

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{k-1} \\ \hline \beta_k \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \beta_k \end{pmatrix}.$$

Then $\mathbf{X}$ is partitioned as $\mathbf{X} = (\mathbf{X}_1, \mathbf{x}_k)$, where $\mathbf{x}_k$ is the last column of $\mathbf{X}$ and $\mathbf{X}_1$ contains all columns except $\mathbf{x}_k$. The reduced model is $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1^* + \boldsymbol{\varepsilon}^*$, and $\boldsymbol{\beta}_1^*$ is estimated as $\hat{\boldsymbol{\beta}}_1^* = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}$. In this case, $h = 1$, and the F statistic in (8.17) becomes

$$F = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \hat{\boldsymbol{\beta}}_1^{*\prime}\mathbf{X}_1'\mathbf{y}}{(\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y})/(n-k-1)}, \tag{8.20}$$

which is distributed as $F(1, n-k-1)$ if $H_0 : \beta_k = 0$ is true. □

Example 8.2c. The test in Section 8.1 for overall regression can be obtained as a full–reduced-model test. In this case, the partitioning of $\mathbf{X}$ and of $\boldsymbol{\beta}$ is $\mathbf{X} = (\mathbf{j}, \mathbf{X}_1)$ and

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \hline \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \boldsymbol{\beta}_1 \end{pmatrix}.$$

The reduced model is $\mathbf{y} = \beta_0^*\mathbf{j} + \boldsymbol{\varepsilon}^*$, for which we have

$$\hat{\beta}_0^* = \bar{y} \quad \text{and} \quad \text{SS}(\beta_0^*) = n\bar{y}^2 \tag{8.21}$$

(see Problem 8.13). Then $\text{SS}(\boldsymbol{\beta}_1|\beta_0) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{y}^2$, which is the same as (8.6). □

8.3 *F* TEST IN TERMS OF R^2

The F statistics in Sections 8.1 and 8.2 can be expressed in terms of R^2 as defined in (7.56).

Theorem 8.3. The F statistics in (8.5) and (8.17) for testing $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ and $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$, respectively, can be written in terms of R^2 as

$$F = \frac{(\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{y}^2)/k}{(\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y})/(n-k-1)} \tag{8.22}$$

$$= \frac{R^2/k}{(1-R^2)/(n-k-1)} \tag{8.23}$$

and

$$F = \frac{(\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \hat{\boldsymbol{\beta}}_1^{*\prime}\mathbf{X}_1'\mathbf{y})/h}{(\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y})/(n-k-1)} \tag{8.24}$$

$$= \frac{(R^2 - R_r^2)/h}{(1-R^2)/(n-k-1)}, \tag{8.25}$$

where R^2 for the full model is given in (7.56) as $R^2 = (\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{y}^2)/(\mathbf{y}'\mathbf{y} - n\bar{y}^2)$ and R_r^2 for the reduced model $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1^* + \boldsymbol{\varepsilon}$ in (8.8) is similarly defined as

$$R_r^2 = \frac{\hat{\boldsymbol{\beta}}_1^{*\prime}\mathbf{X}_1'\mathbf{y} - n\bar{y}^2}{\mathbf{y}'\mathbf{y} - n\bar{y}^2}. \tag{8.26}$$

PROOF. Adding and subtracting $n\bar{y}^2$ in the denominator of (8.22) gives

$$F = \frac{(\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{y}^2)/k}{[\mathbf{y}'\mathbf{y} - n\bar{y}^2 - (\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{y}^2)]/(n-k-1)}.$$

Dividing numerator and denominator by $\mathbf{y}'\mathbf{y} - n\bar{y}^2$ yields (8.23). For (8.25), see Problem 8.15. □

In (8.25), we see that the F test for $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ is equivalent to a test for significant reduction in R^2. Note also that since $F \geq 0$ in (8.25), we have $R^2 \geq R_r^2$, which is an additional confirmation of property 3 in Section 7.7, namely, that adding an x to the model increases R^2.

Example 8.3. For the dependent variable y_2 in the chemical reaction data in Table 7.4, a full model with nine x's and a reduced model with three x's were considered in Example 8.2a. The values of R^2 for the full model and reduced model are .8485 and .3771, respectively. To test the significance of the increase in R^2

from .3771 to .8485, we use (8.25)

$$
\begin{aligned}
F &= \frac{(R^2 - R_r^2)/h}{(1 - R^2)/(n - k - 1)} = \frac{(.8485 - .3771)/6}{(1 - .8485)/9} \\
&= \frac{.07857}{.01683} = 4.6671,
\end{aligned}
$$

which is the same as the value obtained for F in Example 8.2a. □

8.4 THE GENERAL LINEAR HYPOTHESIS TESTS FOR $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ AND $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{t}$

We discuss a test for $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ in Section 8.4.1 and a test for $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{t}$ in Section 8.4.2.

8.4.1 The Test for $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$

The hypothesis $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, where $\mathbf{C}$ is a known $q \times (k + 1)$ coefficient matrix of rank $q \leq k + 1$, is known as the *general linear hypothesis*. The alternative hypothesis is H_1: $\mathbf{C}\boldsymbol{\beta} \neq \mathbf{0}$. The formulation $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ includes as special cases the hypotheses in Sections 8.1 and 8.2. The hypothesis $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ in Section 8.1 can be expressed in the form $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ as follows

$$
H_0 : \mathbf{C}\boldsymbol{\beta} = (\mathbf{0}, \mathbf{I}_k)\begin{pmatrix} \beta_0 \\ \boldsymbol{\beta}_1 \end{pmatrix} = \boldsymbol{\beta}_1 = \mathbf{0} \quad \text{[by (2.36)]},
$$

where $\mathbf{0}$ is a $k \times 1$ vector. Similarly, the hypothesis $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ in Section 8.2 can be expressed in the form $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$:

$$
H_0 : \mathbf{C}\boldsymbol{\beta} = (\mathbf{O}, \mathbf{I}_h)\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} = \boldsymbol{\beta}_2 = \mathbf{0},
$$

where the matrix $\mathbf{O}$ is $h \times (k - h + 1)$ and the vector $\mathbf{0}$ is $h \times 1$.

The formulation $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ also allows for more general hypotheses such as

$$
H_0 : 2\beta_1 - \beta_2 = \beta_2 - 2\beta_3 + 3\beta_4 = \beta_1 - \beta_4 = 0,
$$

which can be expressed as follows:

$$
H_0 : \begin{pmatrix} 0 & 2 & -1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 3 \\ 0 & 1 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.
$$

As another illustration, the hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$ can be expressed in terms of three differences, $H_0 : \beta_1 - \beta_2 = \beta_2 - \beta_3 = \beta_3 - \beta_4 = 0$, or, equivalently, as $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$:

$$H_0 : \begin{pmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

In the following theorem, we give the sums of squares used in the test of $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ versus H_1: $\mathbf{C}\boldsymbol{\beta} \neq \mathbf{0}$, along with the properties of these sums of squares. We denote the sum of squares due to $\mathbf{C}\boldsymbol{\beta}$ (due to the hypothesis) as SSH.

Theorem 8.4a. If $\mathbf{y}$ is distributed $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ and $\mathbf{C}$ is $q \times (k+1)$ of rank $q \leq k+1$, then

(i) $\mathbf{C}\hat{\boldsymbol{\beta}}$ is $N_q[\mathbf{C}\boldsymbol{\beta}, \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']$.

(ii) $\text{SSH}/\sigma^2 = (\mathbf{C}\hat{\boldsymbol{\beta}})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}/\sigma^2$ is $\chi^2(q, \lambda)$, where $\lambda = (\mathbf{C}\boldsymbol{\beta})'[\mathbf{C}(\mathbf{X}'\ \mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\boldsymbol{\beta}/2\sigma^2$.

(iii) $\text{SSE}/\sigma^2 = \mathbf{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}/\sigma^2$ is $\chi^2(n-k-1)$.

(iv) SSH and SSE are independent.

PROOF

(i) By Theorem 7.6b (i), $\hat{\boldsymbol{\beta}}$ is $N_{k+1}[\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}]$. The result then follows by Theorem 4.4a (ii).

(ii) Since $\text{cov}(\mathbf{C}\hat{\boldsymbol{\beta}}) = \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'$ and $\sigma^2[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}]^{-1}\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'/\sigma^2 = \mathbf{I}$, the result follows by Theorem 5.5.

(iii) This was established in Theorem 8.1b(ii).

(iv) Since $\hat{\boldsymbol{\beta}}$ and SSE are independent [see Theorem 7.6b(iii)], $\text{SSH} = \hat{\boldsymbol{\beta}}\mathbf{C}'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']\mathbf{C}\hat{\boldsymbol{\beta}}$ and SSE are also independent (Seber 1977, pp. 17, 33–34). For a more formal proof, see Problem 8.16. □

The F test for $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ versus $H_1 : \mathbf{C}\boldsymbol{\beta} \neq \mathbf{0}$ is given in the following theorem.

Theorem 8.4b. Let $\mathbf{y}$ be $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ and define the statistic

$$F = \frac{\text{SSH}/q}{\text{SSE}/(n-k-1)} = \frac{(\mathbf{C}\hat{\boldsymbol{\beta}})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}/q}{\text{SSE}/(n-k-1)}, \tag{8.27}$$

where $\mathbf{C}$ is $q \times (k+1)$ of rank $q \leq k+1$ and $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. The distribution of F in (8.27) is as follows:

(i) If $H_0: \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is false, then

$$F \text{ is distributed as } F(q, n-k-1, \lambda),$$

where $\lambda = (\mathbf{C}\boldsymbol{\beta})'\left[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right]^{-1}\mathbf{C}\boldsymbol{\beta}/2\sigma^2$.

(ii) If $H_0: \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is true, then

$$F \text{ is distributed as } F(q, n-k-1).$$

Proof

(i) This result follows from (5.30) and Theorem 8.4a.
(ii) This result follows from (5.28) and Theorem 8.4a. □

The F test for $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ in Theorem 8.4b is usually called the *general linear hypothesis test*. The degrees of freedom q is the number of linear combinations in $\mathbf{C}\beta$. The test for $H_0: \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is carried out as follows. Reject H_0 if $F \geq F_{\alpha,q,n-k-1}$, where F is as given in (8.27) and $F_{\alpha,q,n-k-1}$ is the upper α percentage point of the (central) F distribution. Alternatively, we can reject H_0 if $p \leq \alpha$ where p is the p value for F. [The p value is the probability that $F(q, n-k-1)$ exceeds the observed F value.] Since $\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'$ is positive definite (see Problem 8.17), $\lambda > 0$ if H_0 is false, where $\lambda = (\mathbf{C}\boldsymbol{\beta})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\boldsymbol{\beta}/2\sigma^2$. Hence we reject $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ for large values of F.

In Theorems 8.4a and 8.4b, SSH could be written as $(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{0})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{0})$, which is the squared distance between $\mathbf{C}\hat{\boldsymbol{\beta}}$ and the hypothesized value of $\mathbf{C}\boldsymbol{\beta}$. The distance is standardized by the covariance matrix of $\mathbf{C}\hat{\boldsymbol{\beta}}$. Intuitively, if H_0 is true, $\mathbf{C}\hat{\boldsymbol{\beta}}$ tends to be close to $\mathbf{0}$ so that the numerator of F in (8.27) is small. On the other hand, if $\mathbf{C}\boldsymbol{\beta}$ is very different from $\mathbf{0}$, the numerator of F tends to be large.

The expected mean squares for the F test are given by

$$E\left(\frac{\text{SSH}}{q}\right) = \sigma^2 + \frac{1}{q}(\mathbf{C}\boldsymbol{\beta})'\left[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right]^{-1}\mathbf{C}\boldsymbol{\beta}, \tag{8.28}$$

$$E\left(\frac{\text{SSE}}{n-k-1}\right) = \sigma^2.$$

These expected mean squares provide additional motivation for rejecting H_0 for large values of F. If H_0 is true, both expected mean squares are equal to σ^2; if H_0 is false, $E(\text{SSH}/q) > E[\text{SSE}/(n-q-1)]$.

The F statistic in (8.27) is invariant to full-rank linear transformations on the x's or on y.

Theorem 8.4c. Let $\mathbf{z} = c\mathbf{y}$ and $\mathbf{W} = \mathbf{X}\mathbf{K}$, where $\mathbf{K}$ is nonsingular (see Corollary 1 to Theorem 7.3e for the form of $\mathbf{K}$). The F statistic in (8.27) is unchanged by these transformations.

PROOF. See Problem 8.18. □

In the first paragraph of this section, it was noted that the hypothesis $H_0: \boldsymbol{\beta}_2 = \mathbf{0}$ can be expressed in the form $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$. Since we used a full–reduced-model approach to develop the test for $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$, we expect that the general linear hypothesis test is also a full–reduced-model test. This is confirmed in the following theorem.

Theorem 8.4d. The F test in Theorem 8.4b for the general linear hypothesis $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is a full–reduced-model test.

PROOF. The reduced model under H_0 is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \text{ subject to } \mathbf{C}\boldsymbol{\beta} = \mathbf{0}. \tag{8.29}$$

Using Lagrange multipliers (Section 2.14.3), it can be shown (see Problem 8.19) that the estimator for $\boldsymbol{\beta}$ in this reduced model is

$$\hat{\boldsymbol{\beta}}_c = \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}, \tag{8.30}$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is estimated from the full model unrestricted by the hypothesis and the subscript c in $\hat{\boldsymbol{\beta}}_c$ indicates that $\boldsymbol{\beta}$ is estimated subject to the constraint $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$. In (8.29), the $\mathbf{X}$ matrix for the reduced model is unchanged from the full model, and the regression sum of squares for the reduced model is therefore $\hat{\boldsymbol{\beta}}_c'\mathbf{X}'\mathbf{y}$ (for a more formal justification of $\hat{\boldsymbol{\beta}}_c'\mathbf{X}'\mathbf{y}$, see Problem 8.20). Hence, the regression sum of squares due to the hypothesis is

$$\mathrm{SSH} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \hat{\boldsymbol{\beta}}_c'\mathbf{X}'\mathbf{y}. \tag{8.31}$$

By substituting $\hat{\boldsymbol{\beta}}_c$ [as given by (8.30)] into (8.31), we obtain

$$\mathrm{SSH} = (\mathbf{C}\hat{\boldsymbol{\beta}})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\beta}} \tag{8.32}$$

(see Problem 8.21), thus establishing that the F test in Theorem 8.4b for $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, is a full–reduced-model test. □

Example 8.4.1a. In many cases, the hypothesis can be incorporated directly into the model to obtain the reduced model. Suppose that the full model is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$$

and the hypothesis is $H_0: \beta_1 = 2\beta_2$. Then the reduced model becomes

$$\begin{aligned} y_i &= \beta_0 + 2\beta_2 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i \\ &= \beta_{c0} + \beta_{c2}(2x_{i1} + x_{i2}) + \beta_{c3} x_{i3} + \varepsilon_i, \end{aligned}$$

where β_{ci} indicates a parameter subject to the constraint $\beta_1 = 2\beta_2$. The full model and reduced model could be fit, and the difference $\mathrm{SS}(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \hat{\boldsymbol{\beta}}^{*\prime}\mathbf{X}_1'\mathbf{y}$ would be the same as SSH in (8.32). □

If $\mathbf{C}\boldsymbol{\beta} \neq \mathbf{0}$, the estimator $\hat{\boldsymbol{\beta}}_c$ in (8.30) is a biased estimator of $\boldsymbol{\beta}$, but the variances of the $\hat{\beta}_{cj}$'s in $\hat{\boldsymbol{\beta}}_c$ are reduced, as shown in the following theorem.

Theorem 8.4e. The mean vector and covariance matrix of $\hat{\boldsymbol{\beta}}_c$ in (8.30) are as follows:

(i) $E(\hat{\boldsymbol{\beta}}_c) = \boldsymbol{\beta} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\left[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right]^{-1}\mathbf{C}\boldsymbol{\beta}.$ (8.33)

(ii) $\mathrm{cov}(\hat{\boldsymbol{\beta}}_c) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} - \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\left[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right]^{-1}\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}.$ (8.34)

Proof. See Problem 8.22. □

Since the second matrix on the right side of (8.34) is positive semidefinite, the diagonal elements of $\mathrm{cov}(\hat{\boldsymbol{\beta}}_c)$ are less than those of $\mathrm{cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$; that is, $\mathrm{var}(\hat{\beta}_{cj}) \leq \mathrm{var}(\hat{\beta}_j)$ for $j = 0, 1, 2, \cdots, k$, where $\hat{\beta}_{cj}$ is the jth diagonal element of $\mathrm{cov}(\hat{\boldsymbol{\beta}}_c)$ in (8.34). This is analogous to the inequality $\mathrm{var}(\hat{\beta}_j^*) < \mathrm{var}(\hat{\beta}_j)$ in Theorem 7.9c, where $\hat{\beta}_j^*$ is from the reduced model.

Example 8.4.1b. Consider the dependent variable y_1 in the chemical reaction data in Table 7.4. For the model $y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$, we test $H_0 : 2\beta_1 = 2\beta_2 = \beta_3$ using (8.27) in Theorem 8.4b. To express H_0 in the form $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, the matrix $\mathbf{C}$ becomes

$$\mathbf{C} = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 2 & -1 \end{pmatrix},$$

and we obtain

$$\mathbf{C}\hat{\boldsymbol{\beta}} = \begin{pmatrix} -.1214 \\ -.6118 \end{pmatrix},$$

$$\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}' = \begin{pmatrix} .003366 & -.006943 \\ -.006943 & .044974 \end{pmatrix},$$

$$F = \frac{\begin{pmatrix} -.1214 \\ -.6118 \end{pmatrix}' \begin{pmatrix} .003366 & -.006943 \\ -.006943 & .044974 \end{pmatrix}^{-1} \begin{pmatrix} -.1214 \\ -.6118 \end{pmatrix} \Big/ 2}{5.3449}$$

$$= \frac{28.62301/2}{5.3449} = 2.6776,$$

which has $p = .101$. □

8.4.2 The Test for $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{t}$

The test for $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{t}$ is a straightforward extension of the test for $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$. With the additional flexibility provided by $\boldsymbol{t}$, we can test hypotheses such as $H_0 : \beta_2 = \beta_1 + 5$. We assume that the system of equations $\mathbf{C}\boldsymbol{\beta} = \mathbf{t}$ is consistent, that is, that $\text{rank}(\mathbf{C}) = \text{rank}(\mathbf{C}, \mathbf{t})$ (see Theorem 2.7). The requisite sums of squares and their properties are given in the following theorem, which is analogous to Theorem 8.4a.

Theorem 8.4f. If $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ and $\mathbf{C}$ is $q \times (k+1)$ of rank $q \leq k+1$, then

(i) $\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{t}$ is $N_q[\mathbf{C}\boldsymbol{\beta} - \mathbf{t}, \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']$.

(ii) $\text{SSH}/\sigma^2 = (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{t})'\left[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right]^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{t})/\sigma^2$ is $\chi^2(q, \lambda)$ where $\lambda = (\mathbf{C}\boldsymbol{\beta} - \mathbf{t})'\,[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}(\mathbf{C}\boldsymbol{\beta} - \mathbf{t})/2\sigma^2$.

(iii) $\text{SSE}/\sigma^2 = \mathbf{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}/\sigma^2$ is $\chi^2(n-k-1)$.

(iv) SSH and SSE are independent.

PROOF

(i) By Theorem 7.6b (i), $\hat{\boldsymbol{\beta}}$ is $N_{k+1}[\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}]$. The result follows by Corollary 1 to Theorem 4.4a.

(ii) By part (i), $\text{cov}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{t}) = \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'$. The result follows as in the proof of Theorem 8.4a (ii).

(iii) See Theorem 8.1b (ii).

(iv) Since $\hat{\boldsymbol{\beta}}$ and SSE are independent [see Theorem 7.6b (iii)], SSH and SSE are independent [see Seber (1977, pp. 17, 33–34)]. For a more formal proof, see Problem 8.23. □

An F test for $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{t}$ versus $H_1 : \mathbf{C}\boldsymbol{\beta} \neq \mathbf{t}$ is given in the following theorem, which is analogous to Theorem 8.4b.

Theorem 8.4g. Let $\mathbf{y}$ be $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ and define an F statistic as follows:

$$F = \frac{\text{SSH}/q}{\text{SSE}/(n-k-1)}$$

$$= \frac{(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{t})'\left[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right]^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{t})/q}{\text{SSE}/(n-k-1)}, \tag{8.35}$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. The distribution of F in (8.35) is as follows:

(i) If $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{t}$ is false, then

$$F \text{ is distributed as } F(q, n-k-1, \lambda),$$

where $\lambda = (\mathbf{C}\boldsymbol{\beta} - \mathbf{t})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}(\mathbf{C}\boldsymbol{\beta} - \mathbf{t})/2\sigma^2$.

(ii) If $H_0: \mathbf{C}\boldsymbol{\beta} = \mathbf{t}$ is true, then $\lambda = 0$ and

$$F \text{ is distributed as } F(q, n-k-1).$$

PROOF

(i) This result follows from (5.28) and Theorem 8.4f.

(ii) This result follows from (5.30) and Theorem 8.4f. □

The test for $H_0: \mathbf{C}\boldsymbol{\beta} = \mathbf{t}$ is carried out as follows. Reject H_0 if $F \geq F_{\alpha,q,n-k-1}$, where $F_{\alpha,q,\,n-k-1}$ is the upper α percentage point of the central F distribution. Alternatively, we can reject H_0 if $p \leq \alpha$, where p is the p value for F.

The expected mean squares for the F test are given by

$$\begin{aligned} E\left(\frac{\mathrm{SSH}}{q}\right) &= \sigma^2 + \frac{1}{q}(\mathbf{C}\boldsymbol{\beta} - \mathbf{t})'\left[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right]^{-1}(\mathbf{C}\boldsymbol{\beta} - \mathbf{t}), \\ E\left(\frac{\mathrm{SSE}}{n-k-1}\right) &= \sigma^2. \end{aligned} \tag{8.36}$$

By extension of Theorem 8.4d, the F test for $H_0: \mathbf{C}\boldsymbol{\beta} = \mathbf{t}$ in Theorem 8.4g is a full–reduced-model test (see Problem 8.24 for a partial result).

8.5 TESTS ON $\boldsymbol{\beta}_j$ AND $\mathbf{a}'\boldsymbol{\beta}$

We consider tests for a single β_j or a single linear combination $\mathbf{a}'\boldsymbol{\beta}$ in Section 8.5.1 and tests for several β_j's or several $\mathbf{a}_i'\boldsymbol{\beta}$'s in Section 8.5.2.

8.5.1 Testing One $\boldsymbol{\beta}_j$ or One $\mathbf{a}'\boldsymbol{\beta}$

Tests for an individual β_j can be obtained using either the full–reduced-model approach in Section 8.2 or the general linear hypothesis approach in Section 8.4 The test statistic for $H_0: \beta_k = 0$ using a full–reduced–model is given in (8.20) as

$$F = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \hat{\boldsymbol{\beta}}_1^{*\prime}\mathbf{X}_1'\mathbf{y}}{\mathrm{SSE}/(n-k-1)}, \tag{8.37}$$

which is distributed as $F(1, n-k-1)$ if H_0 is true. In this case, β_k is the last β, so that $\boldsymbol{\beta}$ is partitioned as $\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \beta_k \end{pmatrix}$ and $\mathbf{X}$ is partitioned as $\mathbf{X} = (\mathbf{X}_1, \mathbf{x}_k)$, where $\mathbf{x}_k$ is the last column of $\mathbf{X}$. Then $\mathbf{X}_1$ in the reduced model $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1^* + \boldsymbol{\varepsilon}$ contains all the columns of $\mathbf{X}$ except the last.

To test $H_0 : \beta_j = 0$ by means of the general linear hypothesis test of $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ (Section 8.4.1), we first consider a test of $H_0 : \mathbf{a}'\boldsymbol{\beta} = 0$ for a single linear combination, for example, $\mathbf{a}'\boldsymbol{\beta} = (0, 2, -2, 3, 1)\boldsymbol{\beta}$. Using $\mathbf{a}'$ in place of the matrix $\mathbf{C}$ in $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, we have $q = 1$, and (8.27) becomes

$$F = \frac{(\mathbf{a}'\hat{\boldsymbol{\beta}})'\left[\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}\right]^{-1}\mathbf{a}'\hat{\boldsymbol{\beta}}}{\mathrm{SSE}/(n-k-1)} = \frac{(\mathbf{a}'\hat{\boldsymbol{\beta}})^2}{s^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}, \tag{8.38}$$

where $s^2 = \mathrm{SSE}/(n-k-1)$. The F statistic in (8.38) is distributed as $F(1, n-k-1)$ if $H_0 : \mathbf{a}'\boldsymbol{\beta} = 0$ is true.

To test $H_0 : \beta_j = 0$ using (8.38), we define $\mathbf{a}' = (0, \ldots, 0, 1, 0, \ldots, 0)$, where the 1 is in the jth position. This gives

$$F = \frac{\hat{\beta}_j^2}{s^2 g_{jj}}, \tag{8.39}$$

where g_{jj} is the jth diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$. If $H_0 : \beta_j = 0$ is true, F in (8.39) is distributed as $F(1, n-k-1)$. We reject $H_0 : \beta_j = 0$ if $F \geq F_{\alpha,1,n-k-1}$ or, equivalently, if $p \leq \alpha$, where p is the p value for F.

By Theorem 8.4d (see also Problem 8.25), the F statistics in (8.37) and (8.39) are the same (for $j = k$). This confirms that (8.39) tests $H_0 : \beta_j = 0$ adjusted for the other β's.

Since the F statistic in (8.39) has 1 and $n-k-1$ degrees of freedom, we can equivalently use the t statistic

$$t_j = \frac{\hat{\beta}_j}{s\sqrt{g_{jj}}} \tag{8.40}$$

to test the effect of β_j above and beyond the other β's (see Problem 5.16). We reject $H_0 : \beta_j = 0$ if $|t_j| \geq t_{\alpha/2,n-k-1}$ or, equivalently, if $p \leq \alpha$, where p is the p value. For a two-tailed t test such as this one, the p value is twice the probability that $t(n-k-1)$ exceeds the absolute value of the observed t.

For $j = 1$, (8.40) becomes $t = \hat{\beta}_1/s\sqrt{g_{11}}$, which is not the same as $t = \hat{\beta}_1 \Big/ \left[s\Big/\sqrt{\sum_i (x_i - \bar{x})^2}\right]$ in (6.14). Unless the x's are orthogonal, $g_{11}^{-1} \neq \sum_i (x_{1i} - \bar{x}_1)^2$.

8.5.2 Testing Several β_j's or $\mathbf{a}_i'\boldsymbol{\beta}$'s

We sometimes want to carry out several separate tests rather than a single joint test of the hypotheses. For example, the test in (8.40) might be carried out separately for each $\boldsymbol{\beta}_i$, $i = 1, \ldots, k$ rather than the joint test of $H_0 : \boldsymbol{\beta}_1 = 0$ in (8.5). Similarly, we might want to carry out separate tests for several (say, d) $\mathbf{a}_i\boldsymbol{\beta}$'s using (8.38) rather than the joint test of $H_0 : \mathbf{C}\boldsymbol{\beta} = 0$ using (8.27), where

$$\mathbf{C} = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_d \end{pmatrix}.$$

In such situations there are two different α levels, the overall or *familywise* α level (α_f) and the α level for each test or *comparisonwise* α level (α_c). In some cases researchers desire to control α_c when doing several tests (Saville 1990), and so no changes are needed in the testing procedure. In other cases, the desire is to control α_f. In yet other cases, especially those involving thousands of separate tests (e.g., microarray data), it makes sense to control other quantities such as the false discovery rate (Benjamini and Hochberg 1995, Benjamini and Yekutieli 2001). This will not be discussed further here. We consider two ways to control α_f when several tests are made.

The first of these methods is the Bonferroni approach (Bonferroni 1936), which reduces α_c for each test, so that α_f is less than the desired level of α^*. As an example, suppose that we carry out the k tests of H_{0j}: $\beta_j = 0, j = 1, 2, \ldots, k$. Let E_j be the event that the jth test rejects H_{0j} when it is true, where $P(E_j) = \alpha_c$. The overall α_f can be defined as

$$\begin{aligned} \alpha_f &= P(\text{reject at least one } H_{0j} \text{ when all } H_{0j} \text{ are true}) \\ &= P(E_1 \text{ or } E_2 \ldots \text{ or } E_k). \end{aligned}$$

Expressing this more formally and applying the Bonferroni inequality, we obtain

$$\begin{aligned} \alpha_f &= P(E_1 \cup E_2 \cup \cdots \cup E_k) \\ &\leq \sum_{j=1}^{k} P(E_j) = \sum_{j=1}^{k} \alpha_c = k\alpha_c. \end{aligned} \tag{8.41}$$

We can thus ensure that α_f is less than or equal to the desired α^* by simply setting $\alpha_c = \alpha^*/k$. Since α_f in (8.41) is at most α^*, the Bonferroni procedure is a conservative approach.

To test $H_{0j} : \beta_j = 0, j = 1, 2, \ldots, k$, with $\alpha_f \leq \alpha^*$, we use (8.40)

$$t_j = \frac{\hat{\beta}_j}{s\sqrt{g_{jj}}}, \tag{8.42}$$

and reject H_{0j} if $|t_j| \geq t_{\alpha^*/2k, n-k-1}$. Bonferroni critical values $t_{\alpha^*/2k,\nu}$ are available in Bailey (1977). See also Rencher (2002, pp. 562–565). The critical values $t_{\alpha^*/2k,\nu}$ can also be found using many software packages. Alternatively, we can carry out the test by the use of p values and reject H_{0j} if $p \leq \alpha^*/k$.

More generally, to test $H_{0i}: \mathbf{a}_i'\boldsymbol{\beta} = 0$ for $i = 1, 2, \ldots, d$ with $\alpha_f \leq \alpha^*$, we use (8.38)

$$F_i = \frac{(\mathbf{a}_i'\hat{\boldsymbol{\beta}})'\left[\mathbf{a}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}_i\right]^{-1}\mathbf{a}_i'\hat{\boldsymbol{\beta}}}{s^2} \tag{8.43}$$

and reject H_{0i} if $F_i \geq F_{\alpha^*/d, 1, n-k-1}$. The critical values $F_{\alpha^*/d}$ are available in many software packages. To use p values, reject H_{0i} if $p \leq \alpha^*/d$.

The above Bonferroni procedures do not require independence of the $\hat{\beta}_j$'s; they are valid for any covariance structure on the $\hat{\beta}_j$'s. However, the logic of the Bonferroni procedure for testing H_{0i}: $\mathbf{a}_i'\boldsymbol{\beta} = 0$ for $i = 1, 2, \ldots, d$ requires that the coefficient vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_d$ be specified before seeing the data. If we wish to choose values of $\mathbf{a}_i$ after looking at the data, we must use the Scheffé procedure described below. Modifications of the Bonferroni approach have been proposed that are less conservative but still control α_f. For examples of these modified procedures, see Holm (1979), Shaffer (1986), Simes (1986), Holland and Copenhaver (1987), Hochberg (1988), Hommel (1988), Rom (1990), and Rencher (1995, Section 3.4.4). Comparisons of these procedures have been made by Holland (1991) and Broadbent (1993).

A second approach to controlling α_f due to Scheffé (1953; 1959, p. 68) yields simultaneous tests of $H_0: \mathbf{a}'\boldsymbol{\beta} = 0$ for all possible values of $\mathbf{a}$ including those chosen after looking at the data. We could also test $H_0: \mathbf{a}'\boldsymbol{\beta} = t$ for arbitrary t. For any given $\mathbf{a}$, the hypothesis $H_0: \mathbf{a}'\boldsymbol{\beta} = 0$ is tested as usual by (8.38)

$$\begin{aligned} F &= \frac{(\mathbf{a}'\hat{\boldsymbol{\beta}})'\left[\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}\right]^{-1}\mathbf{a}'\hat{\boldsymbol{\beta}}}{s^2} \\ &= \frac{(\mathbf{a}'\hat{\boldsymbol{\beta}})^2}{s^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}, \end{aligned} \tag{8.44}$$

but the test proceeds by finding a critical value large enough to hold for all possible $\mathbf{a}$. Accordingly, we now find the distribution of $\max_{\mathbf{a}} F$.

Theorem 8.5

(i) The maximum value of F in (8.44) is given by

$$\max_{\mathbf{a}} \frac{(\mathbf{a}'\hat{\boldsymbol{\beta}})^2}{s^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}} = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}}{s^2}. \tag{8.45}$$

(ii) If $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, then $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}/(k+1)s^2$ is distributed as $F(k+1, n-k-1)$. Thus

$$\max_{\mathbf{a}} \frac{(\mathbf{a}'\hat{\boldsymbol{\beta}})^2}{s^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}(k+1)}$$

is distributed as $F(k+1, n-k-1)$.

PROOF

(i) Using the quotient rule, chain rule, and Section 2.14.1, we differentiate $(\mathbf{a}'\hat{\boldsymbol{\beta}})^2/\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}$ with respect to $\mathbf{a}$ and set the result equal to $\mathbf{0}$:

$$\frac{\partial}{\partial \mathbf{a}} \frac{(\mathbf{a}'\hat{\boldsymbol{\beta}})^2}{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}} = \frac{[\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}]2(\mathbf{a}'\hat{\boldsymbol{\beta}})\hat{\boldsymbol{\beta}} - (\mathbf{a}'\hat{\boldsymbol{\beta}})^2 2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}{[\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}]^2} = \mathbf{0}.$$

Multiplying by $[\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}]^2/2\mathbf{a}'\hat{\boldsymbol{\beta}}$ and treating 1×1 matrices as scalars, we obtain

$$[\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}]\,\hat{\boldsymbol{\beta}} - \mathbf{a}'\hat{\boldsymbol{\beta}}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a} = \mathbf{0},$$

$$\mathbf{a} = \frac{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}{\mathbf{a}'\hat{\boldsymbol{\beta}}}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = c\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}},$$

where $c = \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}/\mathbf{a}'\hat{\boldsymbol{\beta}}$. Substituting $\mathbf{a} = c\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$ into (8.44) gives

$$\max_{\mathbf{a}} \frac{(\mathbf{a}'\hat{\boldsymbol{\beta}})^2}{s^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}} = \frac{(c\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}})^2}{s^2 c\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}c\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}} = \frac{c^2(\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}})^2}{s^2c^2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}} = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}}{s^2}.$$

(ii) Using $\mathbf{C} = \mathbf{I}_{k+1}$ in (8.27), we have, by Theorem 8.4b (ii), that

$$F = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}}{(k+1)s^2} \text{ is distributed as } F(k+1, n-k-1).$$

□

By Theorem 8.5(ii), we have

$$P\left[\max_{\mathbf{a}} \frac{(\mathbf{a}'\hat{\boldsymbol{\beta}})^2}{s^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}(k+1)} \geq F_{\alpha^*,k+1,n-k-1}\right] = \alpha^*,$$

$$P\left[\max_{\mathbf{a}} \frac{(\mathbf{a}'\hat{\boldsymbol{\beta}})^2}{s^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}} \geq (k+1)F_{\alpha^*,k+1,n-k-1}\right] = \alpha^*.$$

Thus, to test $H_0 : \mathbf{a}'\boldsymbol{\beta} = 0$ for any and all $\mathbf{a}$ (including values of $\mathbf{a}$ chosen after seeing the data) with $\alpha_f \leq \alpha^*$, we calculate F in (8.44) and reject H_0 if $F \geq (k+1)F_{\alpha^*,k+1,n-k-1}$.

To test for individual β_j's using using Scheffé's procedure, we set $\mathbf{a}' = (0, \ldots, 0, 1, 0, \ldots, 0)$ with a 1 in the jth position. Then F in (8.44) reduces to $F = \hat{\beta}_j^2/s^2 g_{jj}$ in (8.39), and the square root is $t_j = \hat{\beta}_j/s\sqrt{g_{jj}}$ in (8.42). By Theorem 8.5, we reject $H_0 : \mathbf{a}'\boldsymbol{\beta} = \beta_j = 0$ if $|t_j| \geq \sqrt{(k+1)F_{\alpha^*,k+1,n-k-1}}$.

For practical purposes $[k \leq (n-3)]$, we have

$$t_{\alpha^*/2k,\, n-k-1} < \sqrt{(k+1)F_{\alpha^*,\,k+1,\,n-k-1}},$$

and thus the Bonferroni tests for individual β_j's in (8.42) are usually more powerful than the Scheffé tests. On the other hand, for a large number of linear combinations $\mathbf{a}'\boldsymbol{\beta}$, the Scheffé test is better since $(k+1)F_{\alpha^*,\,k+1,\,n-k-1}$ is constant, while the critical value $F_{\alpha^*/d,1,n-k-1}$ for Bonferroni tests in (8.43) increases with the number of tests d and eventually exceeds the critical value for Scheffé tests.

It has been assumed that the tests in this section for $H_0 : \beta_j = 0$ are carried out without regard to whether the overall hypothesis $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ is rejected. However, if the test statistics $t_j = \hat{\beta}_j / s\sqrt{g_{jj}}, j = 1, 2, \ldots, k$, in (8.42) are calculated only if $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ is rejected using F in (8.5), then clearly α_f is reduced and the conservative critical values $t_{\alpha^*/2k,\,n-k-1}$ and $\sqrt{(k+1)F_{\alpha^*,k+1,n-k-1}}$ become even more conservative. Using this protected testing principle (Hocking 1996, p. 106), we can even use the critical value $t_{\alpha^*/2,\,n-k-1}$ for all k tests and α_f will still be close to α^*. [For illustrations of this familywise error rate structure, see Hummel and sligo (1971) and Rencher and Scott (1990).] A similar statement can be made for testing the overall hypothesis $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ followed by t tests or F tests of $H_0 : \mathbf{c}_i'\boldsymbol{\beta} = 0$ using the rows of $\mathbf{C}$.

Example 8.5.2. We test $H_{01} : \beta_1 = 0$ and $H_{02} : \beta_2 = 0$ for the data in Table 7.1. Using (8.42) and the results in Examples 7.3.1(a), 7.33 and 8.1, we have

$$t_1 = \frac{\hat{\beta}_1}{s\sqrt{g_{11}}} = \frac{3.0118}{\sqrt{2.8288}\sqrt{.16207}} = \frac{3.0118}{.67709} = 4.448,$$

$$t_2 = \frac{\hat{\beta}_2}{s\sqrt{g_{22}}} = \frac{-1.2855}{\sqrt{2.8288}\sqrt{.08360}} = \frac{-1.2855}{0.48629} = -2.643.$$

Using α=.05 for each test, we reject both H_{01} and H_{02} because $t_{.025,9} = 2.262$. The (two-sided) p values are .00160 and .0268, respectively. If we use $\alpha = .05/2 = .025$ for a Bonferroni test, we would not reject H_{02} since $p = .0268 > .025$. However, using the protected testing principle, we would reject H_{02} because the overall regression hypothesis $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ was rejected in Example 8.1. □

8.6 CONFIDENCE INTERVALS AND PREDICTION INTERVALS

In this section we consider a confidence region for $\boldsymbol{\beta}$, confidence intervals for β_j, $\mathbf{a}'\boldsymbol{\beta}$, $E(y)$, and σ^2, and prediction intervals for future observations. We assume throughout Section 8.6 that $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$.

8.6.1 Confidence Region for $\boldsymbol{\beta}$

If $\mathbf{C}$ is equal to $\mathbf{I}$ and $\mathbf{t}$ is equal to $\boldsymbol{\beta}$ in (8.35), q becomes $k+1$, we obtain a central F distribution, and we can make the probability statement

$$P[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/(k+1)s^2 \leq F_{\alpha,k+1,n-k-1}] = 1 - \alpha,$$

where $s^2 = \text{SSE}/(n-k-1)$. From this statement, a $100(1-\alpha)\%$ joint confidence region for $\beta_0, \beta_1, \ldots, \beta_k$ in $\boldsymbol{\beta}$ is defined to consist of all vectors $\boldsymbol{\beta}$ that satisfy

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq (k+1)s^2 F_{\alpha,k+1,n-k-1}. \tag{8.46}$$

For $k = 1$, this region can be plotted as an ellipse in two dimensions. For $k > 1$, the ellipsoidal region in (8.46) is unwieldy to interpret and report, and we therefore consider intervals for the individual β_j's.

8.6.2 Confidence Interval for β_j

If $\beta_j \neq 0$, we can subtract β_j in (8.40) so that $t_j = (\hat{\beta}_j - \beta_j)/s\sqrt{g_{jj}}$ has the central t distribution, where g_{jj} is the jth diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$. Then

$$P\left[-t_{\alpha/2,n-k-1} \leq \frac{\hat{\beta}_j - \beta_j}{s\sqrt{g_{jj}}} \leq t_{\alpha/2,n-k-1}\right] = 1 - \alpha.$$

Solving the inequality for β_j gives

$$P(\hat{\beta}_j - t_{\alpha/2,n-k-1}s\sqrt{g_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{\alpha/2,n-k-1}s\sqrt{g_{jj}}) = 1 - \alpha.$$

Before taking the sample, the probability that the random interval will contain β_j is $1 - \alpha$. *After* taking the sample, the $100(1-\alpha)\%$ confidence interval for β_j

$$\hat{\beta}_j \pm t_{\alpha/2,\, n-k-1}s\sqrt{g_{jj}} \tag{8.47}$$

is no longer random, and thus we say that we are $100(1-\alpha)\%$ *confident* that the interval contains β_j.

Note that the confidence coefficient $1 - \alpha$ holds only for a single confidence interval for one of the β_j's. For confidence intervals for all $k + 1$ of the β's that hold simultaneously with overall confidence coefficient $1 - \alpha$, see Section 8.6.7.

Example 8.6.2. We compute a 95% confidence interval for each β_j using y_2 in the chemical reaction data in Table 7.4 (see Example 8.2a). The matrix $(\mathbf{X}'\mathbf{X})^{-1}$ (see the answer to Problem 7.52) and the estimate $\hat{\boldsymbol{\beta}}$ have the following values:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 65.37550 & -0.33885 & -0.31252 & -0.02041 \\ -0.33885 & 0.00184 & 0.00127 & -0.00043 \\ -0.31252 & 0.00127 & 0.00408 & -0.00176 \\ -0.02041 & -0.00043 & -0.00176 & 0.02161 \end{pmatrix},$$

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} -26.0353 \\ 0.4046 \\ 0.2930 \\ 1.0338 \end{pmatrix}.$$

For β_1, we obtain by (8.47),

$$\begin{aligned}
&\hat{\beta}_1 \pm t_{.025,15} s \sqrt{g_{11}} \\
&.4046 \pm (2.1314)(4.0781)\sqrt{.00184} \\
&.4046 \pm .3723, \\
&(.0322,\ .7769).
\end{aligned}$$

For the other β_j's, we have

$$\begin{aligned}
\beta_0:&\quad -26.0353 \pm 70.2812 \\
&\quad (-96.3165,\ 44.2459), \\
\beta_2:&\quad .2930 \pm .5551 \\
&\quad (-.2621,\ .8481), \\
\beta_3:&\quad 1.0338 \pm 1.27777 \\
&\quad (-.2439,\ 2.3115).
\end{aligned}$$

The confidence coefficient .95 holds for only one of the four confidence intervals. For more than one interval, see Example 8.6.7. □

8.6.3 Confidence Interval for $\mathbf{a}'\boldsymbol{\beta}$

If $\mathbf{a}'\boldsymbol{\beta} \neq 0$, we can subtract $\mathbf{a}'\boldsymbol{\beta}$ from $\mathbf{a}'\hat{\boldsymbol{\beta}}$ in (8.44) to obtain

$$F = \frac{(\mathbf{a}'\hat{\boldsymbol{\beta}} - \mathbf{a}'\boldsymbol{\beta})^2}{s^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}},$$

which is distributed as $F(1, n-k-1)$. Then by Problem 5.16,

$$t = \frac{\mathbf{a}'\hat{\boldsymbol{\beta}} - \mathbf{a}'\boldsymbol{\beta}}{s\sqrt{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}} \tag{8.48}$$

is distributed as $t(n-k-1)$, and a $100(1-\alpha)\%$ confidence interval for a single value of $\mathbf{a}'\boldsymbol{\beta}$ is given by

$$\mathbf{a}'\hat{\boldsymbol{\beta}} \pm t_{\alpha/2,n-k-1} s\sqrt{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}. \tag{8.49}$$

8.6.4 Confidence Interval for $E(y)$

Let $\mathbf{x}_0 = (1, x_{01}, x_{02}, \ldots, x_{0k})'$ denote a particular choice of $\mathbf{x} = (1, x_1, x_2, \ldots, x_k)'$. Note that $\mathbf{x}_0$ need not be one of the $\mathbf{x}$'s in the sample; that is, $\mathbf{x}_0'$ need not be a row of $\mathbf{X}$. If $\mathbf{x}_0$ is very far outside the area covered by the sample however, the prediction may be poor. Let y_0 be an observation corresponding to $\mathbf{x}_0$. Then

$$y_0 = \mathbf{x}_0'\boldsymbol{\beta} + \varepsilon,$$

and [assuming that the model is correct so that $E(\varepsilon) = 0$]

$$E(y_0) = \mathbf{x}_0'\boldsymbol{\beta}. \tag{8.50}$$

We wish to find a confidence interval for $E(y_0)$, that is, for the mean of the distribution of y-values corresponding to $\boldsymbol{x}_0$.

By Corollary 1 to Theorem 7.6d, the minimum variance unbiased estimator of $E(y_0)$ is given by

$$\widehat{E(y_0)} = \mathbf{x}_0'\hat{\boldsymbol{\beta}}. \tag{8.51}$$

Since (8.50) and (8.51) are of the form $\mathbf{a}'\boldsymbol{\beta}$ and $\mathbf{a}'\hat{\boldsymbol{\beta}}$, respectively, we obtain a $100(1-\alpha)\%$ confidence interval for $E(y_0) = \mathbf{x}_0'\boldsymbol{\beta}$ from (8.49):

$$\mathbf{x}_0'\hat{\boldsymbol{\beta}} \pm t_{\alpha/2,n-k-1}s\sqrt{\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}. \tag{8.52}$$

The confidence coefficient $1-\alpha$ for the interval in (8.52) holds only for a single choice of the vector $\mathbf{x}_0$. For intervals covering several values of $\mathbf{x}_0$ or all possible values of $\mathbf{x}_0$, see Section 8.6.7.

We can express the confidence interval in (8.52) in terms of the centered model in Section 7.5, $y_i = \alpha + \boldsymbol{\beta}_1'(\mathbf{x}_{01} - \bar{\mathbf{x}}_1) + \varepsilon_i$, where $\mathbf{x}_{01} = (x_{01}, x_{02}, \ldots, x_{0k})'$ and $\bar{\mathbf{x}}_1 = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k)'$. [We use the notation $\mathbf{x}_{01}$ to distinguish this vector from $\mathbf{x}_0 = (1, x_{01}, x_{02}, \ldots, x_{0k})'$ above.] For the centered model, (8.50), (8.51), and (8.52) become

$$E(y_0) = \alpha + \boldsymbol{\beta}_1'(\mathbf{x}_{01} - \bar{\mathbf{x}}_1), \tag{8.53}$$

$$\widehat{E(y_0)} = \bar{y} + \hat{\boldsymbol{\beta}}_1'(\mathbf{x}_{01} - \bar{\mathbf{x}}_1), \tag{8.54}$$

$$\bar{y} + \hat{\boldsymbol{\beta}}_1'(\mathbf{x}_{01} - \bar{\mathbf{x}}_1) \pm t_{\alpha/2,n-k-1}s\sqrt{\frac{1}{n} + (\mathbf{x}_{01} - \bar{\mathbf{x}}_1)'(\mathbf{X}_c'\mathbf{X}_c)^{-1}(\mathbf{x}_{01} - \bar{\mathbf{x}}_1)}. \tag{8.55}$$

Note that in the form shown in (8.55), it is clear that if $\mathbf{x}_{01}$ is close to $\bar{\mathbf{x}}_1$ the interval is narrower; in fact, it is narrowest for $\mathbf{x}_{01} = \bar{\mathbf{x}}$. The width of the interval increases as the distance of $\mathbf{x}_{01}$ from $\bar{\mathbf{x}}_1$ increases.

For the special case of simple linear regression, (8.50), (8.51), and (8.55) reduce to

$$E(y_0) = \beta_0 + \beta_1 x_0, \tag{8.56}$$

$$\widehat{E(y_0)} = \hat{\beta}_0 + \hat{\beta}_1 x_0, \tag{8.57}$$

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\alpha/2,n-2}s\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}, \tag{8.58}$$

where s is given by (6.11). The width of the interval in (8.58) depends on how far x_0 is from $\bar{x}$.

Example 8.6.4. For the grades data in Example 6.2, we find a 95% confidence interval for $E(y_0)$, where $x_0 = 80$. Using (8.58), we obtain

$$\begin{gathered}\hat{\beta}_0 + \hat{\beta}_1(80) \pm t_{.025,16}s\sqrt{\frac{1}{18} + \frac{(80 - 58.056)^2}{19530.944}},\\ 80.5386 \pm 2.1199(13.8547)(.2832),\\ 80.5386 \pm 8.3183,\\ (72.2204, 88.8569).\end{gathered}$$

□

8.6.5 Prediction Interval for a Future Observation

A "confidence interval" for a future observation y_0 corresponding to $\mathbf{x}_0$ is called a *prediction interval*. We speak of a prediction interval rather than a confidence interval because y_0 is an individual observation and is thereby a random variable rather than a parameter. To be $100(1-\alpha)\%$ confident that the interval contains y_0, the prediction interval will clearly have to be wider than a confidence interval for the parameter $E(y_0)$.

Since $y_0 = \mathbf{x}_0'\beta + \varepsilon_0$, we predict y_0 by $\hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}}$, which is also the estimator of $E(y_0) = \mathbf{x}_0'\boldsymbol{\beta}$. The random variables y_0 and $\hat{y}_0$ are independent because y_0 is a future observation to be obtained independently of the n observations used to compute $\hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}}$. Hence the variance of $y_0 - \hat{y}_0$ is

$$\text{var}(y_0 - \hat{y}_0) = \text{var}(y_0 - \mathbf{x}_0'\hat{\boldsymbol{\beta}}) = \text{var}(\mathbf{x}_0'\boldsymbol{\beta} + \varepsilon_0 - \mathbf{x}_0'\hat{\boldsymbol{\beta}}).$$

Since $\mathbf{x}_0'\boldsymbol{\beta}$ is a constant, this becomes

$$\begin{aligned}\text{var}(y_0 - \hat{y}_0) &= \text{var}(\varepsilon_0) + \text{var}(\mathbf{x}_0'\hat{\boldsymbol{\beta}}) = \sigma^2 + \sigma^2\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 \\ &= \sigma^2\left[1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0\right],\end{aligned} \tag{8.59}$$

which is estimated by $s^2[1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0]$. It can be shown that $E(y_0 - \hat{y}_0) = 0$ and that s^2 is independent of both y_0 and $\hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}}$. Therefore, the t statistic

$$t = \frac{y_0 - \hat{y}_0 - 0}{s\sqrt{1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}} \tag{8.60}$$

is distributed as $t(n - k - 1)$, and

$$P = \left[-t_{\alpha/2,n-k-1} \leq \frac{y_0 - \hat{y}_0}{s\sqrt{1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}} \leq t_{\alpha/2,n-k-1}\right] = 1 - \alpha.$$

The inequality can be solved for y_0 to obtain the $100(1-\alpha)\%$ prediction interval

$$\hat{y}_0 - t_{\alpha/2,\, n-k-1}s\sqrt{1+\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0} \leq y_0 \leq \hat{y}_0 + t_{\alpha/2,\, n-k-1}s\sqrt{1+\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}$$

or, using $\hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}}$, we have

$$\mathbf{x}_0'\hat{\boldsymbol{\beta}} \pm t_{\alpha/2,n-k-1}s\sqrt{1+\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}. \tag{8.61}$$

Note that the confidence coefficient $1-\alpha$ for the prediction interval in (8.61) holds for only one value of $\mathbf{x}_0$.

In $1+\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0$, the second term, $\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0$, is typically much smaller than 1 (provided k is much smaller than n) because the variance of $\hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}}$ is much less than the variance of y_0. [To illustrate, if $\mathbf{X}'\mathbf{X}$ were diagonal and $\mathbf{x}_0$ were in the area covered by the rows of $\mathbf{X}$, then $\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0$ would be a sum with $k+1$ terms, each of the form $x_{0j}^2/\sum_{i=1}^n x_{ij}^2$, which is of the order of $1/n$.] Thus prediction intervals for y_0 are generally much wider than confidence intervals for $E(y_0) = \mathbf{x}_0'\boldsymbol{\beta}$.

In terms of the centered model in Section 7.5, the $100(1-\alpha)\%$ prediction interval in (8.61) becomes

$$\bar{y} + \hat{\boldsymbol{\beta}}_1'(\mathbf{x}_{01} - \bar{\mathbf{x}}_1) \pm t_{\alpha/2,n-k-1}s\sqrt{1+\frac{1}{n}+(\mathbf{x}_{01}-\bar{\mathbf{x}}_1)'(\mathbf{X}_c'\mathbf{X}_c)^{-1}(\mathbf{x}_{01}-\bar{\mathbf{x}}_1)}. \tag{8.62}$$

For the case of simple linear regression, (8.61) and (8.62) reduce to

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\alpha/2,\, n-2}s\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}, \tag{8.63}$$

where s is given by (6.11). In (8.63), it is clear that the second and third terms within the square root are much smaller than 1 unless x_0 is far removed from the interval bounded by the smallest and largest x's.

For a prediction interval for the mean of q future observations, see Problem 8.30.

Example 8.6.5. Using the data from Example 6.2, we find a 95% prediction interval for y_0 when $x_0 = 80$. Using (8.63), we obtain

$$\hat{\beta}_0 + \hat{\beta}_1(80) \pm t_{.025,16}s\sqrt{1+\frac{1}{18}+\frac{(80-58.056)^2}{19530.944}},$$

$$80.5386 \pm 2.1199(13.8547)(1.0393),$$

$$80.5386 \pm 30.5258,$$

$$(50.0128, 111.0644).$$

Note that the prediction interval for y_0 here is much wider than the confidence interval for $E(y_0)$ in Example 8.6.4. □

8.6.6 Confidence Interval for σ^2

By Theorem 7.6b(ii), $(n-k-1)s^2/\sigma^2$ is $\chi^2(n-k-1)$. Therefore

$$P\left[\chi^2_{1-\alpha/2,\, n-k-1} \le \frac{(n-k-1)s^2}{\sigma^2} \le \chi^2_{\alpha/2,\, n-k-1}\right] = 1-\alpha, \tag{8.64}$$

where $\chi^2_{\alpha/2,\, n-k-1}$ is the upper $\alpha/2$ percentage point of the chi-square distribution and $\chi^2_{1-\alpha/2,\, n-k-1}$ is the lower $\alpha/2$ percentage point. Solving the inequality for σ^2 yields the $100(1-\alpha)\%$ confidence interval

$$\frac{(n-k-1)s^2}{\chi^2_{\alpha/2,n-k-1}} \le \sigma^2 \le \frac{(n-k-1)s^2}{\chi^2_{1-\alpha/2,\, n-k-1}}. \tag{8.65}$$

A $100(1-\alpha)\%$ confidence interval for σ is given by

$$\sqrt{\frac{(n-k-1)s^2}{\chi^2_{\alpha/2,n-k-1}}} \le \sigma \le \sqrt{\frac{(n-k-1)s^2}{\chi^2_{1-\alpha/2,\, n-k-1}}}. \tag{8.66}$$

8.6.7 Simultaneous Intervals

By analogy to the discussion of testing several hypotheses (Section 8.5.2), when several intervals are computed, two confidence coefficients can be considered: familywise confidence $(1-\alpha_f)$ and individual confidence $(1-\alpha_c)$. Familywise confidence of $1-\alpha_f$ means that we are $100(1-\alpha_f)\%$ confident that every interval contains its respective parameter.

In some cases, our goal is simply to control $1-\alpha_c$ for each one of several confidence or prediction intervals so that no changes are needed to expressions (8.47), (8.49), (8.52), and (8.61). In other cases the desire is to control $1-\alpha_f$. To do so, both the Bonferroni and Scheffé methods can be adapted to the situation of multiple intervals. In yet other cases we may want to control other properties of multiple intervals (Benjamini and Yekutieli 2005).

The Bonferroni procedure increases the width of each individual interval so that $1-\alpha_f$ for the set of intervals is greater than or equal to the desired value $1-\alpha^*$. As an example suppose that it is desired to calculate the k confidence intervals for $\beta_1, \ldots, \beta_k$. Let E_j be the event that the jth interval includes β_j, and E_j^c be the complement of that event. Then by definition

$$\begin{aligned} 1-\alpha_f &= P(E_1 \cap E_2 \cap \ldots \cap E_k) \\ &= 1 - P(E_1^c \cup E_2^c \cup \ldots \cup E_k^c). \end{aligned}$$

Assuming that $P(E_j^c) = \alpha_c$ for $j = 1, \ldots, k$, the Bonferroni inequality now implies that

$$1 - \alpha_f \geq 1 - k\alpha_c.$$

Hence we can ensure that $1 - \alpha_f$ is greater than or equal to the desired $1 - \alpha^*$ by setting $1 - \alpha_c = 1 - \alpha^*/k$ for the individual intervals.

Using this approach, Bonferroni confidence intervals for $\beta_1, \beta_2, \ldots, \beta_k$ are given by

$$\hat{\beta}_j \pm t_{\alpha^*/2k,\, n-k-1} s\sqrt{g_{jj}}, \quad j = 1, 2, \ldots, k, \tag{8.67}$$

where g_{jj} is the jth element of $(\mathbf{X}'\mathbf{X})^{-1}$. Bonferroni t values $t_{\alpha^*/2k}$ are available in Bailey (1977) and can also be obtained in many software programs. For example, a probability calculator for the t, the F, and other distributions is available free from NCSS (download at www.ncss.com).

Similarly for d linear functions $\mathbf{a}_1'\boldsymbol{\beta}, \mathbf{a}_2'\boldsymbol{\beta}, \ldots, \mathbf{a}_d'\boldsymbol{\beta}$ (chosen before seeing the data), Bonferroni confidence intervals are given by

$$\mathbf{a}_i'\hat{\boldsymbol{\beta}} \pm t_{\alpha^*/2d,\, n-k-1} s\sqrt{\mathbf{a}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}_i}, \quad i = 1, 2, \ldots, d. \tag{8.68}$$

These intervals hold simultaneously with familywise confidence of at least $1 - \alpha^*$.

Bonferroni confidence intervals for $E(y_0) = \mathbf{x}_0'\boldsymbol{\beta}$ for a few values of $\mathbf{x}_0$, say, $\mathbf{x}_{01}, \mathbf{x}_{02}, \ldots, \mathbf{x}_{0d}$ are given by

$$\mathbf{x}_{0i}'\hat{\boldsymbol{\beta}} \pm t_{\alpha^*/2d,\, n-k-1} s\sqrt{\mathbf{x}_{0i}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{0i}}, \quad i = 1, 2, \ldots, d. \tag{8.69}$$

[Note that $\mathbf{x}_{01}$ here differs from $\mathbf{x}_{01}$ in (8.53)–(8.55).]

For simultaneous prediction of d new observations $y_{01}, y_{02}, \ldots, y_{0d}$ at d values of $\mathbf{x}_0$, say, $\mathbf{x}_{01}, \mathbf{x}_{02}, \ldots, \mathbf{x}_{0d}$, we can use the Bonferroni prediction intervals

$$\mathbf{x}_{0i}'\hat{\boldsymbol{\beta}} \pm t_{\alpha^*/2d, n-k-1} s\sqrt{1 + \mathbf{x}_{0i}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{0i}} \quad i = 1, 2, \ldots, d \tag{8.70}$$

[see (8.61) and (8.69)].

Simultaneous Scheffé confidence intervals for all possible linear functions $\mathbf{a}'\boldsymbol{\beta}$ (including those chosen after seeing the data) can be based on the distribution of $\max_{\mathbf{a}} F$ [Theorem 8.5(ii)]. Thus a conservative confidence interval for any and all $\mathbf{a}'\boldsymbol{\beta}$ is

$$\mathbf{a}'\hat{\boldsymbol{\beta}} \pm s\sqrt{(k+1)F_{\alpha^*,\, k+1,\, n-k-1}\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}. \tag{8.71}$$

The (potentially infinite number of) intervals in (8.71) have an overall confidence coefficient of at least $1 - \alpha^*$. For a few linear functions, the intervals in (8.68) will be narrower, but for a large number of linear functions, the intervals in (8.71) will be narrower. A comparison of $t_{\alpha^*/2d, n-k-1}$ and $\sqrt{(k+1)F_{\alpha^*,\, k+1,\, n-k-1}}$ will show which is preferred in a given case.

For confidence limits for $E(y_0) = \mathbf{x}_0'\boldsymbol{\beta}$ for all possible values of $\mathbf{x}_0$, we use (8.71):

$$\mathbf{x}_0'\hat{\boldsymbol{\beta}} \pm s\sqrt{(k+1)F_{\alpha^*,k+1,n-k-1}\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}. \tag{8.72}$$

These intervals hold simultaneously with a confidence coefficient of $1 - \alpha^*$. Thus, (8.72) becomes a confidence region that can be applied to the entire regression surface for all values of $\mathbf{x}_0$. The intervals in (8.71) and (8.72) are due to Scheffé (1953; 1959, p. 68) and Working and Hotelling (1929).

Scheffé-type prediction intervals for $y_{01}, y_{02}, \ldots, y_{0d}$ are given by

$$\mathbf{x}_{0i}'\hat{\boldsymbol{\beta}} \pm s\sqrt{dF_{\alpha^*,d,h-k-1}[1 + \mathbf{x}_{0i}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{0i}]} \quad i = 1, 2, \ldots, d \tag{8.73}$$

(see Problem 8.32). These d prediction intervals hold simultaneously with overall confidence coefficient at least $1 - \alpha^*$, but note that $dF_{\alpha^*,d,n-k-1}$ is not constant. It depends on the number of predictions.

Example 8.6.7. We compute 95% Bonferroni confidence limits for β_1, β_2, and β_3, using y_2 in the chemical reaction data in Table 7.4; see Example 8.6.2 for $(\mathbf{X}'\mathbf{X})^{-1}$ and $\hat{\boldsymbol{\beta}}$. By (8.67), we have

$$\begin{gathered}\hat{\beta}_1 \pm t_{.025/3,15}s\sqrt{g_{11}} \\ .4056 \pm (2.6937)(4.0781)\sqrt{.00184} \\ .4056 \pm .4706 \\ (-.0660, .8751),\end{gathered}$$

$$\begin{aligned}\beta_2\!: &\quad .2930 \pm .7016 \\ &\quad (-.4086, .9946)\ , \\ \beta_3\!: &\quad 1.0338 \pm 1.6147 \\ &\quad (-.5809, 2.6485).\end{aligned}$$

These three intervals hold simultaneously with confidence coefficient at least .95. □

8.7 LIKELIHOOD RATIO TESTS

The tests in Sections 8.1, 8.2, and 8.4 were derived using informal methods based on finding sums of squares that have chi-square distributions and are independent. These same tests can be obtained more formally by the likelihood ratio approach. Likelihood ratio tests have some good properties and sometimes have optimal properties.

We describe the likelihood ratio method in the simple context of testing $H_0 : \boldsymbol{\beta} = \mathbf{0}$ versus $H_1 : \boldsymbol{\beta} \neq \mathbf{0}$. The likelihood function $L(\boldsymbol{\beta}, \sigma^2)$ was defined in Section 7.6.2 as the joint density of the y's. For a random sample $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ with density $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, the likelihood function is given by (7.50) as

$$L(\boldsymbol{\beta},\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})/2\sigma^2}. \tag{8.74}$$

The likelihood ratio method compares the maximum value of $L(\boldsymbol{\beta}, \sigma^2)$ restricted by $H_0 : \boldsymbol{\beta} = \mathbf{0}$ to the maximum value of $L(\boldsymbol{\beta}, \sigma^2)$ under H_1: $\boldsymbol{\beta}_1 \neq \mathbf{0}$, which is essentially unrestricted. We denote the maximum value of $L(\boldsymbol{\beta},\sigma^2)$ restricted by $\boldsymbol{\beta} = \mathbf{0}$ as $\max_{H_0} L(\boldsymbol{\beta}, \sigma^2)$ and the unrestricted maximum as $\max_{H_1} L(\boldsymbol{\beta}, \sigma^2)$. If $\boldsymbol{\beta}$ is equal (or close) to $\mathbf{0}$, then $\max_{H_0} L(\boldsymbol{\beta}, \sigma^2)$ should be close to $\max_{H_1} L(\boldsymbol{\beta}, \sigma^2)$. If $\max_{H_0} L(\boldsymbol{\beta}, \sigma^2)$ is not close to $\max_{H_1} L(\boldsymbol{\beta}, \sigma^2)$, we would conclude that $y = (y_1, y_2, \ldots, y_n)'$ apparently did not come from $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ with $\boldsymbol{\beta} = \mathbf{0}$.

In this illustration, we can find $\max_{H_0} L(\boldsymbol{\beta}, \sigma^2)$ by setting $\boldsymbol{\beta} = \mathbf{0}$ and then estimating σ^2 as the value that maximizes $L(\mathbf{0}, \sigma^2)$. Under $H_1 : \boldsymbol{\beta} \neq \mathbf{0}$, both $\boldsymbol{\beta}$ and σ^2 are estimated without restriction as the values that maximize $L(\boldsymbol{\beta}, \sigma^2)$. [In designating the unrestricted maximum as $\max_{H_1} L(\boldsymbol{\beta}, \sigma^2)$, we are ignoring the restriction in H_1 that $\boldsymbol{\beta} \neq \mathbf{0}$.]

It is customary to describe the likelihood ratio method in terms of maximizing L subject to ω, the set of all values of $\boldsymbol{\beta}$ and σ^2 satisfying H_0, and subject to Ω, the set of all values of β and σ^2 without restrictions (other than natural restrictions such as $\sigma^2 > 0$). However, to simplify notation in cases such as this in which H_1 includes all values of $\boldsymbol{\beta}$ except $\mathbf{0}$, we refer to maximizing L under H_0 and H_1.

We compare the restricted maximum under H_0 with the unrestricted maximum under H_1 by the *likelihood ratio*

$$\begin{aligned} \mathrm{LR} &= \frac{\max_{H_0} L(\boldsymbol{\beta}, \sigma^2)}{\max_{H_1} L(\boldsymbol{\beta}, \sigma^2)} \\ &= \frac{\max L(\mathbf{0}, \sigma^2)}{\max L(\boldsymbol{\beta}, \sigma^2)}. \end{aligned} \tag{8.75}$$

It is clear that $0 \leq \mathrm{LR} \leq 1$, because the maximum of L restricted to $\boldsymbol{\beta} = \mathbf{0}$ cannot exceed the unrestricted maximum. Smaller values of **LR** would favor H_1, and larger values would favor H_0. We thus reject H_0 if $\mathrm{LR} \leq c$, where c is chosen so that $P(\mathrm{LR} \leq c) = \alpha$ if H_0 is true.

Wald (1943) showed that, under H_0

$$-2 \ln \mathrm{LR} \text{ is approximately } \chi^2(\nu)$$

for large n, where ν is the number of parameters estimated under H_1 minus the number estimated under H_0. In the case of $H_0 : \boldsymbol{\beta} = \mathbf{0}$ versus $H_1 : \boldsymbol{\beta} \neq \mathbf{0}$, we have $\nu = k + 2 - 1 = k + 1$ because $\boldsymbol{\beta}$ and σ^2 are estimated under H_1 while only σ^2 is estimated under H_0. In some cases, the χ^2 approximation is not needed because LR turns out to be a function of a familiar test statistic, such as t or F, whose exact distribution is available.

We now obtain the likelihood ratio test for $H_0 : \boldsymbol{\beta} = \mathbf{0}$. The resulting likelihood ratio is a function of the F statistic obtained in Problem 8.6 by partitioning the total sum of squares.

Theorem 8.7a. If $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, the likelihood ratio test for $H_0 : \boldsymbol{\beta} = \mathbf{0}$ can be based on

$$F = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}/(k+1)}{(\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y})/(n-k-1)}.$$

We reject H_0 if $F > F_{\alpha, k+1, n-k-1}$.

Proof. To find $\max_{H_1} L(\boldsymbol{\beta}, \sigma^2) = \max L(\boldsymbol{\beta}, \sigma^2)$, we use the maximum likelihood estimators $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ and $\hat{\sigma}^2 = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})/n$ from Theorem 7.6a. Substituting these in (8.74), we obtain

$$\begin{aligned}
\max_{H_1} L(\boldsymbol{\beta}, \sigma^2) &= \max L(\boldsymbol{\beta}, \sigma^2) = L(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) \\
&= \frac{1}{(2\pi\hat{\sigma}^2)^{n/2}} e^{-(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})/2\hat{\sigma}^2} \\
&= \frac{n^{n/2}e^{-n/2}}{(2\pi)^{n/2}\left[(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})\right]^{n/2}}.
\end{aligned} \tag{8.76}$$

To find $\max_{H_0} L(\boldsymbol{\beta}, \sigma^2) = \max L(\mathbf{0}, \sigma^2)$, we solve $\partial \ln L(\mathbf{0}, \sigma^2)/\partial\sigma^2 = 0$ to obtain

$$\hat{\sigma}_0^2 = \frac{\mathbf{y}'\mathbf{y}}{n}. \tag{8.77}$$

Then

$$\begin{aligned}
\max_{H_0} L(\boldsymbol{\beta}, \sigma^2) &= \max L(\mathbf{0}, \sigma^2) = L(\mathbf{0}, \hat{\sigma}_0^2) \\
&= \frac{1}{(2\pi\hat{\sigma}_0^2)^{n/2}} e^{-\mathbf{y}'\mathbf{y}/2\hat{\sigma}_0^2} \\
&= \frac{n^{n/2}e^{-n/2}}{(2\pi)^{n/2}(\mathbf{y}'\mathbf{y})^{n/2}}.
\end{aligned} \tag{8.78}$$

Substituting (8.76) and (8.78) into (8.75), we obtain

$$\begin{aligned} \mathrm{LR} &= \frac{\max_{H_0} L(\boldsymbol{\beta}, \sigma^2)}{\max_{H_1} L(\boldsymbol{\beta}, \sigma^2)} = \left[\frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{\mathbf{y}'\mathbf{y}} \right]^{n/2} \\ &= \left[\frac{1}{1 + (k+1)F/(n-k-1)} \right]^{n/2}, \end{aligned} \tag{8.79}$$

where

$$F = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}/(k+1)}{(\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y})/(n-k-1)}.$$

Thus, rejecting $H_0 : \boldsymbol{\beta} = \mathbf{0}$ for a small value of LR is equivalent to rejecting H_0 for a large value of F. □

We now show that the F test in Theorem 8.4b for the general linear hypothesis $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is a likelihood ratio test.

Theorem 8.7b. If $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, then the F test for $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ in Theorem 8.4b is equivalent to the likelihood ratio test.

Proof. Under $H_1 : \mathbf{C}\boldsymbol{\beta} \neq \mathbf{0}$, which is essentially unrestricted, $\max_{H_1} L(\boldsymbol{\beta}, \sigma^2)$ is given by (8.76). To find $\max_{H_0} L(\boldsymbol{\beta}, \sigma^2) = \max L(\boldsymbol{\beta}, \sigma^2)$ subject to $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, we use the method of Lagrange multipliers (Section 2.14.3) and work with $L(\boldsymbol{\beta}, \sigma^2)$ to simplify the differentiation:

$$\begin{aligned} v &= \ln L(\boldsymbol{\beta}, \sigma^2) + \boldsymbol{\lambda}'(\mathbf{C}\boldsymbol{\beta} - \mathbf{0}) \\ &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma^2} + \boldsymbol{\lambda}'\mathbf{C}\boldsymbol{\beta}. \end{aligned}$$

Expanding $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ and differentiating with respect to $\boldsymbol{\beta}, \boldsymbol{\lambda}$, and σ^2, we obtain

$$\frac{\partial v}{\partial \boldsymbol{\beta}} = (2\mathbf{X}'\mathbf{y} - 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta})/2\sigma^2 + \mathbf{C}'\boldsymbol{\lambda} = \mathbf{0}, \tag{8.80}$$

$$\frac{\partial v}{\partial \boldsymbol{\lambda}} = \mathbf{C}\boldsymbol{\beta} = \mathbf{0}, \tag{8.81}$$

$$\frac{\partial v}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0. \tag{8.82}$$

Eliminating $\boldsymbol{\lambda}$ and solving for $\boldsymbol{\beta}$ and σ^2 gives

$$\hat{\boldsymbol{\beta}}_0 = \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}, \tag{8.83}$$

$$\hat{\sigma}_0^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_0)'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_0) \tag{8.84}$$

$$= \hat{\sigma}^2 + \frac{1}{n}(\mathbf{C}\hat{\boldsymbol{\beta}})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\beta}} \tag{8.85}$$

(Problems 8.35 and 8.36), where $\hat{\sigma}^2 = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})/n$ and $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ are the maximum likelihood estimates from Theorem 7.6a. Thus

$$\begin{aligned}
\max_{H_0} L(\boldsymbol{\beta}, \sigma^2) &= L(\hat{\boldsymbol{\beta}}_0, \hat{\sigma}_0^2) \\
&= \frac{1}{(2\pi)^{n/2}(\hat{\sigma}_0^2)^{n/2}} e^{-(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}_0)'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}_0)/2\hat{\sigma}_0^2} \\
&= \frac{n^{n/2}e^{-n/2}}{(2\pi)^{n/2}\left\{\mathrm{SSE} + (\mathbf{C}\hat{\boldsymbol{\beta}})'\left[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right]^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}\right\}^{n/2}},
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{LR} &= \frac{\max_{H_0} L(\boldsymbol{\beta}, \sigma^2)}{\max_{H_1} L(\boldsymbol{\beta}, \sigma^2)} \\
&= \left[\frac{\mathrm{SSE}}{\mathrm{SSE} + (\mathbf{C}\hat{\boldsymbol{\beta}})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}}\right]^{n/2} \\
&= \left[\frac{1}{1 + \mathrm{SSH}/\mathrm{SSE}}\right]^{n/2} = \left[\frac{1}{1 + qF/(n-k-1)}\right]^{n/2},
\end{aligned}$$

where $\mathrm{SSH} = (\mathbf{C}\hat{\boldsymbol{\beta}})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}$, $\mathrm{SSE} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$, and F is given in (8.27). □

PROBLEMS

8.1 Show that $\mathrm{SSR} = \hat{\boldsymbol{\beta}}_1'\mathbf{X}_c'\mathbf{X}_c\hat{\boldsymbol{\beta}}_1$ in (8.1) becomes $\mathbf{y}'\mathbf{X}_c(\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_c'\mathbf{y}$ as in (8.2).

8.2 (**a**) Show that $\mathbf{H}_c[\mathbf{I} - (1/n)\mathbf{J}] = \mathbf{H}_c$, as in (8.3) in Theorem 8.1a(i), where $\mathbf{H}_c = \mathbf{X}_c(\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_c'$.

(**b**) Prove Theorem 8.1a(ii).

(**c**) Prove Theorem 8.1a(iii).

(**d**) Prove Theorem 8.1a(iv).

8.3 Show that $\lambda_1 = \boldsymbol{\beta}_1'\mathbf{X}_c\mathbf{X}_c\boldsymbol{\beta}_1/2\sigma^2$ as in Theorem 8.1b(i).

8.4 Prove Theorem 8.1b(ii).

8.5 Show that $E(\text{SSR}/k) = \sigma^2 + (1/k)\boldsymbol{\beta}_1'\mathbf{X}_c'\mathbf{X}_c\boldsymbol{\beta}_1$, as in the expected mean square column of Table 8.1. Employ the following two approaches:

(**a**) Use Theorem 5.2a.

(**b**) Use the noncentrality parameter in (5.19).

8.6 Develop a test for $H_0 : \boldsymbol{\beta} = \mathbf{0}$ in the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$. (It was noted at the beginning of Section 8.1 that this hypothesis is of little practical interest because it includes $\beta_0 = 0$.) Use the partitioning $\mathbf{y}'\mathbf{y} = (\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}) + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$, and proceed as follows:

(**a**) Show that $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ and $\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = \mathbf{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}$.

(**b**) Let $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Show that $\mathbf{H}$ and $\mathbf{I} - \mathbf{H}$ are idempotent of rank $k + 1$ and $n - k - 1$, respectively.

(**c**) Show that $\mathbf{y}'\mathbf{H}\mathbf{y}/\sigma^2$ is $\chi^2(k + 1, \lambda_1)$, where $\lambda_1 = \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}/2\sigma^2$, and that $\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}/\sigma^2$ is $\chi^2(n - k - 1)$.

(**d**) Show that $\mathbf{y}'\mathbf{H}\mathbf{y}$ and $\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}$ are independent.

(**e**) Show that

$$\frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}}{(k+1)s^2} = \frac{\mathbf{y}'\mathbf{H}\mathbf{y}/(k+1)}{\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}/(n-k-1)}$$

is distributed as $F(k + 1, n - k - 1, \lambda_1)$.

8.7 Show that $\mathbf{H}\mathbf{H}_1 = \mathbf{H}_1$ and $\mathbf{H}_1\mathbf{H} = \mathbf{H}_1$, as in (8.15), where $\mathbf{H}$ and $\mathbf{H}_1$ are as defined in (8.11) and (8.12).

8.8 Show that conditions (a) and (b) of Corollary 1 to Theorem 5.6c are satisfied for the sum of quadratic forms in (8.12), as noted in the proof of Theorem 8.2b.

8.9 Show that $\lambda_1 = \boldsymbol{\beta}_2'[\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2]\boldsymbol{\beta}_2/2\sigma^2$ as in Theorem 8.2b(ii).

8.10 Show that $\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$ is positive definite, as noted below Theorem 8.2b.

8.11 Show that $E[\text{SS}(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)/h] = \sigma^2 + \boldsymbol{\beta}_2'[\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2]\boldsymbol{\beta}_2/h$ as in Table 8.3.

8.12 Find the expected mean square corresponding to the numerator of the F statistic in (8.20) in Example 8.2b.

8.13 Show that $\hat{\beta}_0^* = \bar{y}$ and $\text{SS}(\beta_0^*) = n\bar{y}^2$, as in (8.21) in Example 8.2c.

8.14 In the proof of Theorem 8.2d, show that $(\hat{\boldsymbol{\beta}}_1'\mathbf{X}_1' + \hat{\boldsymbol{\beta}}_2'\mathbf{X}_2')(\mathbf{X}_1\hat{\boldsymbol{\beta}}_1 + \mathbf{X}_2\hat{\boldsymbol{\beta}}_2) - (\hat{\boldsymbol{\beta}}_1' + \hat{\boldsymbol{\beta}}_2'\mathbf{A}')\mathbf{X}_1'\mathbf{X}_1(\hat{\boldsymbol{\beta}}_1 + \mathbf{A}\hat{\boldsymbol{\beta}}_2) = \hat{\boldsymbol{\beta}}_2'[\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2]\hat{\boldsymbol{\beta}}_2$.

8.15 Express the test for $H_0: \boldsymbol{\beta}_2 = \mathbf{0}$ in terms of R^2, as in (8.25) in Theorem 8.3.

8.16 Prove Theorem 8.4a(iv).

8.17 Show that $\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'$ is positive definite, as noted following Theorem 8.4b.

8.18 Prove Theorem 8.4c.

8.19 Show that in the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ subject to $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ in (8.29), the estimator of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}}_c = \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}$ as in (8.30), where $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Use a Lagrange multiplier $\boldsymbol{\lambda}$ and minimize $u = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\lambda}'(\mathbf{C}\boldsymbol{\beta} - \mathbf{0})$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ as follows:

(**a**) Differentiate u with respect to $\boldsymbol{\lambda}$ and set the result equal to $\mathbf{0}$ to obtain $\mathbf{C}\hat{\boldsymbol{\beta}}_c = \mathbf{0}$.

(**b**) Differentiate u with respect to $\boldsymbol{\beta}$ and set the result equal to $\mathbf{0}$ to obtain

$$\hat{\boldsymbol{\beta}}_c = \hat{\boldsymbol{\beta}} - \tfrac{1}{2}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\boldsymbol{\lambda}, \qquad (1)$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

(**c**) Multiply (1) in part (b) by $\mathbf{C}$, use $\mathbf{C}\hat{\boldsymbol{\beta}}_c = \mathbf{0}$ from part (a), solve for $\boldsymbol{\lambda}$, and substitute back into (1).

8.20 Show that $\hat{\boldsymbol{\beta}}_c'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}_c = \hat{\boldsymbol{\beta}}_c'\mathbf{X}'\mathbf{y}$, thus demonstrating directly that the sum of squares due to the reduced model is $\hat{\boldsymbol{\beta}}_c'\mathbf{X}'\mathbf{y}$ and that (8.31) holds.

8.21 Show that for the general linear hypothesis $H_0: \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ in Theorem 8.4d, we have $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \hat{\boldsymbol{\beta}}_c'\mathbf{X}'\mathbf{y} = (\mathbf{C}\hat{\boldsymbol{\beta}})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}$ as in (8.32), where $\hat{\boldsymbol{\beta}}_c$ is as given in (8.30).

8.22 Prove Theorem 8.4e.

8.23 Prove Theorem 8.4f(iv) by expressing SSH and SSE as quadratic forms in the same normally distributed random vector.

8.24 Show that the estimator for $\boldsymbol{\beta}$ in the reduced model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ subject to $\mathbf{C}\boldsymbol{\beta} = \mathbf{t}$ is given by $\hat{\boldsymbol{\beta}}_c = \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{t})$, where $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

8.25 Show that $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \hat{\boldsymbol{\beta}}_1^{*\prime}\mathbf{X}_1'\mathbf{y}$ in (8.37) is equal to $\hat{\beta}_k^2/g_{kk}$ in (8.39) (for $j = k$), as noted below (8.39).

8.26 Obtain the confidence interval for $\mathbf{a}'\boldsymbol{\beta}$ in (8.49) from the t statistic in (8.48).

8.27 Show that the confidence interval for $\mathbf{x}_0'\boldsymbol{\beta}$ in (8.52) is the same as that for the centered model in (8.55).

8.28 Show that the confidence interval for $\beta_0 + \beta_1 x_0$ in (8.58) follows from (8.55).

8.29 Show that $t = (y_0 - \hat{y}_0)/s\sqrt{1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}$ in (8.60) is distributed as $t(n - k - 1)$.

8.30 (**a**) Given that $\bar{y}_0 = \sum_i^q = y_{0i}/q$ is the mean of q future observations at $\mathbf{x}_0$, show that a $100(1 - \alpha)\%$ prediction interval for $\bar{y}_0$ is given by $\mathbf{x}_0'\hat{\boldsymbol{\beta}} \pm t_{\alpha/2,\, n-k-1} s\sqrt{1/q + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}$.

(**b**) Show that for simple linear regression, the prediction interval for $\bar{y}_0$ in part (a) reduces to $\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\alpha/2,\, n-2} s\sqrt{1/q + 1/n + (x_0 - \bar{x})^2/\sum_{i=1}^n (x_i - \bar{x})^2}$.

8.31 Obtain the confidence interval for σ^2 in (8.65) from the probability statement in (8.64).

8.32 Show that the Scheffé prediction intervals for d future observations are given by (8.73).

8.33 Verify (8.76)–(8.79) in the proof of Theorem 8.7a.

8.34 Verify (8.80), $\partial v/\partial \boldsymbol{\beta} = (2\mathbf{X}'\mathbf{y} - 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta})/2\sigma^2 + \mathbf{C}'\boldsymbol{\lambda}$.

8.35 Show that the solution to (8.80)–(8.82) is given by $\hat{\boldsymbol{\beta}}_0$ and $\hat{\sigma}_0^2$ in (8.83) and (8.84).

8.36 Show that $(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_0)'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_0) = n\hat{\sigma}^2 + (\mathbf{C}\hat{\boldsymbol{\beta}})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}$ as in (8.85).

8.37 Use the gas vapor data in Table 7.3.

(**a**) Test the overall regression hypothesis $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ using (8.5) [or (8.22)] and (8.23).

(**b**) Test $H_0 : \beta_1 = \beta_3 = 0$, that is, that x_1 and x_3 do not significantly contribute above and beyond x_2 and x_4.

(**c**) Test $H_0 : \beta_j = 0$ for $j = 1, 2, 3, 4$ using t_j in (8.40). Use $t_{.05/2}$ for each test and also use a Bonferroni approach based on $t_{.05/8}$ (or compare the p value to $.05/4$).

(**d**) Using general linear hypothesis tests, test $H_0 : \beta_1 = \beta_2 = 12\beta_3 = 12\beta_4$, $H_{01} : \beta_1 = \beta_2, H_{02} : \beta_2 = 12\beta_3, H_{03} : \beta_3 = \beta_4$, and $H_{04} : \beta_1 = \beta_2$ and $\beta_3 = \beta_4$.

(**e**) Find confidence intervals for $\beta_1, \beta_2, \beta_3$ and β_4 using both (8.47) and (8.67).

8.38 Use the land rent data in Table 7.5.

(**a**) Test the overall regression hypothesis $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ using (8.5) [or (8.22)] and (8.23).

(**b**) Test $H_0 : \beta_j = 0$ for $j = 1, 2, 3$ using t_j in (8.40). Use $t_{.05/2}$ for each test and also use a Bonferroni approach based on $t_{.05/6}$ (or compare the p value to $.05/3$).

(**c**) Find confidence intervals for $\beta_1, \beta_2, \beta_3$ using both (8.47) and (8.67).

(**d**) Using (8.52), find a 95% confidence interval for $E(y_0) = \mathbf{x}_0'\boldsymbol{\beta}$, where $\mathbf{x}_0' = (1{,}15{,}30{,}.5)$.

(e) Using (8.61), find a 95% prediction interval for $y_0 = \mathbf{x}_0'\boldsymbol{\beta} + \varepsilon$, where $\mathbf{x}_0' = (1,15,30,.5)$.

8.39 Use y_2 in the chemical reaction data in Table 7.4.

(a) Using (8.52), find a 95% confidence interval for $E(y_0) = \mathbf{x}_0'\boldsymbol{\beta}$, where $\mathbf{x}_0' = (1, 165, 32,5)$.

(b) Using (8.61), find a 95% prediction interval for $y_0 = \mathbf{x}_0'\boldsymbol{\beta} + \varepsilon$, where $\mathbf{x}_0' = (1,165,32,5)$.

(c) Test $H_0{:}2\beta_1 = 2\beta_2 = \beta_3$ using (8.27). (This was done for y_1 in Example 8.4.b.)

8.40 Use y_1 in the chemical reaction data in Table 7.4. The full model with second-order terms and the reduced model with only linear terms were fit in Problem 7.52.

(a) Test $H_0 : \beta_4 = \beta_5 = \cdots = \beta_9 = 0$, that is, that the second-order terms are not useful in predicting y_1. (This was done for y_2 in Example 8.2a.)

(b) Test the significance of the increase in R^2 from the reduced model to the full model. (This was done for y^2 in Example 8.3. See Problem 7.52 for values of R^2.)

(c) Find a 95% confidence interval for each of β_0, β_1, β_2, β_3 using (8.47).

(d) Find Bonferroni confidence intervals for β_1, β_2, β_3 using (8.67).

(e) Using (8.52), find a 95% confidence interval for $E(y_0) = \mathbf{x}_0'\boldsymbol{\beta}$, where $\mathbf{x}_0' = (1,165,32,5)$.

(f) Using (8.61), find a 95%, prediction interval for $y_0 = \mathbf{x}_0'\boldsymbol{\beta} + \varepsilon$, where $\mathbf{x}_0' = (1,165,32,5)$.