Ch7. Multiple Regression: Tests of Hypothesis and Confidence Intervals

In this chapter we consider hypothesis tests and confidence intervals for the parameters $\beta_0, \cdots, \beta_k$ in $\beta$ in the model $y = X\beta + \epsilon$. We will assume throughout the chapter that $y$ is $N_n(X\beta, \sigma^2 I)$, where $X$ is $n \times (k + 1)$ of rank $k + 1 < n$, and the $x$’s are fixed constants.

1 Test of Overall Regression

We begin with a test of the overall regression hypothesis that none of the $x$’s predict $y$. This hypothesis can be expressed as $H_0 : \beta_1 = 0$, where $\beta_1 = (\beta_1, \cdots, \beta_k)'$. Note that we wish to test $H_0 : \beta_1 = 0$, not $H_0 : \beta = 0$, where $\beta = (\beta_0, \beta_1')'$. Since $\beta_0$ is usually not zero, rejection of $H_0 : \beta = 0$ might be due solely to $\beta_0$, and we would not learn if the $x$’s predict $y$. 
For this test, we use the centered model

\[ y = (j, X_c) \begin{pmatrix} \alpha \\ \beta_1 \end{pmatrix} + \epsilon, \]  

where \( X_c = (I - \frac{1}{n}J)X_1 \) is the centered matrix and \( X_1 \) contains all the columns of \( X \) except the first. Note the partition identity

\[ y'(I - \frac{1}{n}J)y = SSR + SSE \]

\[ = y'X_c(X'_cX_c)^{-1}X'_cy + y'(I - \frac{1}{n}J)y - y'X_c(X'_cX_c)^{-1}X'_cy \]

\[ = y'Ay + y'(I - \frac{1}{n}J - A)y, \]

where \( A = X_c(X'_cX_c)^{-1}X'_c. \)

In the following theorem we establish some properties of the matrices of the quadratic forms in (2).
Theorem 1.1. The matrices \( I - \frac{1}{n} J, A = X_c(X'_cX_c)^{-1}X'_c, \) and \( I - \frac{1}{n} J - A \) have the following properties:

(i) \( A(I - \frac{1}{n} J) = A. \)

(ii) \( A \) is idempotent of rank \( k. \)

(iii) \( I - \frac{1}{n} J - A \) is idempotent of rank \( n - k - 1. \)

(iv) \( A(I - \frac{1}{n} J - A) = O. \)

PROOF: Part (i) follows from \( X'_cj = 0. \) The other parts are left as exercises.

Theorem 1.2. If \( y \) is \( N_n(X\beta, \sigma^2 I) \), then \( SSR/\sigma^2 = \hat{\beta}'_1X'_cX_c\hat{\beta}_1/\sigma^2 \) and \( SSE/\sigma^2 = [\sum_{i=1}^n(y_i - \bar{y})^2 - \hat{\beta}'_1X'_cX_c\hat{\beta}_1]/\sigma^2 \) have the following distributions:

(i) \( SSR/\sigma^2 \) is \( \chi^2(k, \lambda_1) \), where \( \lambda_1 = \hat{\beta}'_1X'_cX_c\hat{\beta}_1/[2\sigma^2]. \)

(ii) \( SSE/\sigma^2 \) is \( \chi^2(n - k - 1). \)

PROOF: The results follows from (2), theorem 1.1 (ii) and (iii).
The following theorem shows that $SSR$ and $SSE$ are independent.

**Theorem 1.3** If $y$ is $N_n(X\beta, \sigma^2 I)$, then $SSR$ and $SSE$ are independent.

**Proof:** It follows from equation (2) and theorem 1.1(iv).

**Theorem 1.4** If $y$ is $N_n(X\beta, \sigma^2 I)$, the distribution of

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

is as follows:

(i) If $H_0 : \beta_1 = 0$ is false, then

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

where $\lambda_1 = \beta_1' X_c' X_c \beta_1 / 2\sigma^2$.

(ii) If $H_0 : \beta_1 = 0$ is true, then $\lambda_1 = 0$ and

$F$ is distributed as $F(k, n-k-1)$. 
Table 1: Analysis of Variance (ANOVA) for $F$-test of $H_0 : \beta_1 = 0$.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>d.f.</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>Expected Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Due to $\beta_1$</td>
<td>$k$</td>
<td>SSR</td>
<td>$SSR/k$</td>
<td>$\sigma^2 + \frac{1}{k} \beta_1' X_c' X_c \beta_1$</td>
</tr>
<tr>
<td>Error</td>
<td>$n-k-1$</td>
<td>SSE</td>
<td>$SSE/(n-k-1)$</td>
<td>$\sigma^2$</td>
</tr>
<tr>
<td>Total</td>
<td>$n-1$</td>
<td>SST</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Proof:** Exercise.

Note that $\lambda_1 = 0$ if and only if $\beta_1 = 0$, since $X_c' X_c$ is positive definite.

The test for $H_0 : \beta_1 = 0$ is carried out as follows: reject $H_0$ if $F \geq F_{\alpha,k,n-k-1}$, where $F \geq F_{\alpha,k,n-k-1}$ is the upper $\alpha$ percentile of the central $F$-distribution. In table 1, we summarize the results leading to the $F$-test.

If $H_0 : \beta_1 = 0$ is true, both of the expected mean squares in table 1 are equal to $\sigma^2$, and we expect $F$ to be near 1. If $\beta_1 \neq 0$, then $E(SSR/k) \geq \sigma^2$ since $X_c' X_c$ is positive definite, and we expect $F$ to exceed 1. We therefore reject $H_0$ for
Example 1.1  For the data in Table 7.1, test $H_0 : \beta_1 = 0$. In this case,

\[
\begin{align*}
\beta_1 &= (\beta_1, \beta_2)', \\
X'y &= (90, 482, 872)', \\
\hat{\beta} &= (5.3754, 3.0118, -1.2855)', \\
y'y &= \sum_{i=1}^{12} y_i^2 = 840, \\
\hat{\beta}'X'y &= 814.5410, \\
n\bar{y}^2 &= 675.
\end{align*}
\]

Thus, we have

\[
\begin{align*}
SSR &= \hat{\beta}'X'y - n\bar{y}^2 = 139.5410, \\
SSE &= y'y - \hat{\beta}'X'y = 25.4590, \\
SST &= y'y - n\bar{y}^2 = 165.
\end{align*}
\]
Table 2: ANOVA for overall regression test for the data in Table 7.1

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Due to $\beta_1$</td>
<td>2</td>
<td>139.5410</td>
<td>69.7705</td>
<td>24.665</td>
</tr>
<tr>
<td>Error</td>
<td>9</td>
<td>25.4590</td>
<td>2.8288</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>165.0000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The $F$-test is given in Table 2. Since $24.665 > F_{0.05, 2, 9} = 4.26$, we reject $H_0$ and conclude that at least one of $\beta_1$ or $\beta_2$ is not zero.

2 Test on a subset of the $\beta$’s

In some cases, we wish to test the hypothesis that a subset of the $x$’s is not useful in predicting $y$. Without loss of generality, we assume that the $\beta$’s to be tested have been arranged last in $\beta$, with a corresponding arrangement of the columns of $X$, i.e., the
The model becomes

\[ y = X\beta + \epsilon = (X_1, X_2) \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \epsilon \]

\[ = X_1\beta_1 + X_2\beta_2 + \epsilon, \]

where \( \beta_2 \) contains the \( \beta \)'s to be tested. The intercept \( \beta_0 \) would ordinarily included in \( \beta_1 \).

To test \( H_0 : \beta_2 = 0 \) versus \( H_1 : \beta_2 \neq 0 \), we use a full-reduced-model approach. The full model is given by (3), the reduced model becomes

\[ y = X_1\beta_1^* + \epsilon^*. \]

Typically, \( \beta_1^* \) and \( \epsilon^* \) will be different from \( \beta_1 \) and \( \epsilon \) in the full model (unless \( X_1 \) and \( X_2 \) are orthogonal). For \( y'y \) we have the following partitioning:

\[ y'y = (y'y - \hat{\beta}'X'y) + (\hat{\beta}'X'y - \hat{\beta}_1'*X_1y) + \hat{\beta}_1'*X_1y \]

\[ = SSE + SS(\beta_2|\beta_1) + SS(\beta_1^*) \]
where $\hat{\beta} = (X'X)^{-1}X'y$ and $\hat{\beta}_1^* = (X'X_1)^{-1}X'_1y$. Note that

$$SS(\beta_2|\beta_1) = (\hat{\beta}'X'y - n\bar{y}^2) - (\hat{\beta}_1^*'X_1y - n\bar{y}^2) = SSR(\text{full}) - SSR(\text{reduced}).$$

Substituting $\hat{\beta}$ and $\hat{\beta}_1^*$ into equation (4), we have

$$y'y = y'[I - X(X'X)^{-1}X']y + y'[X(X'X)^{-1}X' - X_1(X'_1X_1)^{-1}X'_1]y + y'X_1(X'_1X_1)^{-1}X'_1y$$

$$= y'(I - A_1)y + y'(A_1 - A_2)y + y'A_2y,$$

where $A_1 = X(X'X)^{-1}X'$ and $A_2 = X_1(X'_1X_1)^{-1}X'_1$.  

**Theorem 2.1** The matrix $A_1 - A_2 = X(X'X)^{-1}X' - X_1(X'_1X_1)^{-1}X_1$ is idempotent with rank $h$, where $h$ is the number of elements in $\beta_2$.

**Proof:** Multiplying $X$ by $A_1$, we have

$$A_1X = X(X'X)^{-1}X'X = X,$$
or

\[ X = X (X' X)^{-1} X' X. \]  \hspace{1cm} (6)

Partitioning \( X \) on the left side of (6) and the last \( X \) on the right side, we have

\[ X_1 = X (X' X)^{-1} X' X X_1, \]
\[ X_2 = X (X' X)^{-1} X' X X_2. \]  \hspace{1cm} (7)

Using (7) and its transpose, we have

\[ A_1 A_2 = A_2 \quad \text{and} \quad A_2 A_1 = A_2. \]

Note \( A_1 \) and \( A_2 \) are both idempotent. Thus

\[ (A_1 - A_2)^2 = A_1 - A_1 A_2 - A_2 A_1 + A_2^2 \]
\[ = A_1 - A_2 - A_2 + A_2 = A_1 - A_2. \]

Hence, \( A_1 - A_2 \) is idempotent.

\[ rank(A_1 - A_2) = tr(A_1 - A_2) = tr(A_1) - tr(A_2) \]
\[ = tr(I_{k+1}) - tr(I_{k-h+1}) = k + 1 - (k - h + 1) = h. \]
The next theorem gives the distribution of \( y'(I - A_1)y \) and \( y'(A_1 - A_2)y \) and shows that they are independent.

**Theorem 2.2** If \( y \) is \( N_n(X\beta, \sigma^2 I) \), then

(i) \( y'(I - A_1)y/\sigma^2 \) is \( \chi^2(n - k - 1) \).

(ii) \( y'(A_1 - A_2)y/\sigma^2 \) is \( \chi^2(h, \lambda_1) \), where

\[
\lambda_1 = \beta' \left[ X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2 \right] \beta_2 / 2\sigma^2.
\]

(iii) \( y'(I - A_1)y \) and \( y'(A_1 - A_2)y \) are independent.

**Proof:** (i) and (ii) are left for exercises. (iii) Note that \( A_2 = A_1A_2 \), hence

\[
(I - A_1)(A_1 - A_2) = (I - A_1)(A_1 - A_1A_2) = O.
\]

Since \( X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2 \) is positive definite (see problem 8.10 of Rencher and Schaalje (2008)), \( \lambda_1 = 0 \) if and only if \( \beta = 0 \).
Theorem 2.3  Let $\mathbf{y}$ be $N_n(\mathbf{X}\beta, \sigma^2\mathbf{I})$ and define an $F$-statistic as follows:

$$F = \frac{\mathbf{y}'(\mathbf{A}_1 - \mathbf{A}_2)\mathbf{y}/h}{\mathbf{y}'(\mathbf{I} - \mathbf{A})\mathbf{y}/(n - k - 1)} = \frac{SS(\beta_2|\beta_1)/h}{SSE/(n - k - 1)} \frac{[SSR(\text{full}) - SSR(\text{reduced})]/h}{SSE/(n - k - 1)}.$$

The distribution of $F$ is as follows:

(i) If $H_0 : \beta_2 = 0$ is false, then

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

where $\lambda_1 = \beta_2'[\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2]\beta_2/2\sigma^2$.

(ii) If $H_0 : \beta_2 = 0$ is true, then $\lambda_1 = 0$ and

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

PROOF: Exercise.
Table 3: Analysis of Variance for $F$-test of $H_0 : \beta_2 = 0$.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>d.f.</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Due to $\beta_2$ adjusted for $\beta_1$</td>
<td>$h$</td>
<td>$SS(\beta_2</td>
<td>\beta_1)$</td>
<td>$SS(\beta_2</td>
</tr>
<tr>
<td>Error</td>
<td>$n - k - 1$</td>
<td>$SSE$</td>
<td>$SSE/(n - k - 1)$</td>
<td>$SSE/(n - k - 1)$</td>
</tr>
<tr>
<td>Total</td>
<td>$n - 1$</td>
<td>$SST$</td>
<td>$SST$</td>
<td></td>
</tr>
</tbody>
</table>

The test for $H_0 : \beta_2 = 0$ is carried as follows: reject $H_0$ if $F \geq F_{\alpha,h,n-k-1}$, where $F_{\alpha,h,n-k-1}$ is the upper percentile of the central $F$-distribution. Note

$$E[SS(\beta_2|\beta_1)/h] = \sigma^2 + \frac{1}{h}\beta_2'[X_2'X_2 - X_2'(X_1'X_1)^{-1}X_1'X_2]\beta_2,$$

and

$$E[SSE/(n - k - 1)] = \sigma^2,$$

hence, if $H_0$ is false, $F$ will have a large value. The test is summarized in table 3.
3 The general linear hypothesis tests for $H_0 : C\beta = 0$ and $H_0 : C\beta = t$

3.1 The Test for $H_0 : C\beta = 0$

General linear hypothesis: the hypothesis $H_0 : C\beta = 0$, where $C$ is a $q \times (k+1)$ coefficient matrix of rank $q \leq k + 1$, is known as the general linear hypothesis. The alternative hypothesis is $H_1 : C\beta \neq 0$. Note that the formulation $H_0 : C\beta = 0$ includes as special cases the hypotheses in last two sections.

Theorem 3.1 If $y$ is distributed as $N_n(X\beta, \sigma^2 I)$ and $C$ is $q \times (k+1)$ of rank $q \leq k + 1$, then

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

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

$$\lambda = (C\beta)'[C(X'X)^{-1}C']^{-1}C\beta/2\sigma^2;$$
(iii) \( \frac{SSE}{\sigma^2} = y'(I - X(X'X)^{-1}X')y/\sigma^2 \) is \( \chi^2(n - k - 1) \).

(iv) SSH and SSE are independent.

**Proof:** (ii) Since \( \text{cov}(C\hat{\beta}) = \sigma^2 C(X'X)^{-1}C' \) and \( \sigma^2[C(X'X)^{-1}C']^{-1}C(X'X)^{-1}C'/\sigma^2 \) \( I \), which is idempotent.

(iv) Since \( SSH = (C\hat{\beta})'[C(X'X)^{-1}C']^{-1}C\hat{\beta} = y'X \cdots X'y \), and \( X'(I - X(X'X)^{-1}X') = O \), SSE and SSH are independent.

**Theorem 3.2** Let \( y \) be \( N_n(X\beta, \sigma^2 I) \) and define the statistic

\[
F = \frac{SSH/q}{SSE/(n - k - 1)} = \frac{(C\hat{\beta})'[C(X'X)^{-1}C']^{-1}C\hat{\beta}/q}{SSE/(n - k - 1)}
\]

where \( C \) is \( q \times (k + 1) \) of ran \( mk \ q \leq k + 1 \) and \( \hat{\beta} = (X'X)^{-1}X'y \). The distribution of \( F \) is as follows:

(i) If \( H_0 : C\beta = 0 \) is false, then
\( F \) is distributed as \( F(q, n - k - 1, \lambda) \),

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

(ii) If \( H_0 : C\beta = 0 \) is true, then

\( F \) is distributed as \( F(q, n - k - 1) \).

**Proof**: Exercise.

This test is usually called the *general linear hypothesis test*. The degrees of freedom \( q \) is the number of linear combinations in \( C\beta \). Reject \( H_0 \) if \( F \geq F_{\alpha,q,n-k-1} \), where \( F \geq F_{\alpha,q,n-k-1} \) is the upper percentile of the central \( F \)-distribution. Since \( C(X'X)^{-1}C' \) is positive definite, \( \lambda > 0 \) if \( H_0 \) is false. Hence, we reject \( H_0 \) for large values of \( F \). The expected mean squares for the \( F \)-test are given by

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

\[
E\left( \frac{SSE}{n - k - 1} \right) = \sigma^2.
\]
Theorem 3.3  The $F$-test in theorem 3.2 for the general linear hypothesis $H_0 : C\beta = 0$ is a full-and-reduced-model test.

PROOF: The reduced model under $H_0$ is

$$y = X\beta + \epsilon \text{ subject to } C\beta = 0.$$  

Using Lagrange multipliers, it can be shown (problem 8.19 of Rencher and Schaalje (2008)) that the estimator for $\beta$ in this reduced model is

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

where $\hat{\beta}$ is estimated from the full model unrestricted by the hypothesis. In the reduced model, the $X$ matrix is unchanged from the full model, and the regression sum of squares for the reduced model is therefore $\hat{\beta}_c'X'y$. (since $C\hat{\beta}_c = 0$.) Hence, the regression sum of squares due to the hypothesis is

$$SSH = \hat{\beta}'X'y - \hat{\beta}_c'X'y.$$  

By substituting $\hat{\beta}_c$ into SSH, we obtain

$$SSH = (C\hat{\beta})'[C(X'X)^{-1}C']^{-1}C\hat{\beta},$$
thus establishing that the $F$-test in theorem 3.2 for $H_0: C\beta = 0$ is a full-and-reduced-model test.

3.2 The Test for $H_0: C\beta = t$

We assume that the system $C\beta = t$ is consistent, that is, $\text{rank}(C) = \text{rank}(C,t)$.

**Theorem 3.4** If $y$ is distributed as $N_n(X\beta, \sigma^2 I)$ and $C$ is $q \times (k+1)$ of rank $q \leq k+1$, then

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

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

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

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

(iv) SSH and SSE are independent.
**Proof:** (iv) A simple argument is that: since \( \hat{\beta} \) and SSE are independent, SSH and SSE are independent (Seber 1977, pp.17, 33-34). A formal proof is as follows.

First, note that

\[
C(X'X)^{-1}X'y - t = C(X'X)^{-1}X'[y - XC'(CC')^{-1}t],
\]

so that

\[
SSH = [y - XC'(C'C)^{-1}t]'A[y - XC'(C'C)^{-1}t],
\]

where \( A = X(X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}C(X'X)^{-1}X' \).

\[
SSE = [y - XC'(C'C)^{-1}t]'B[y - XC'(C'C)^{-1}t],
\]

where \( B = I - X(X'X)^{-1}X' \). Since \( AB = O \), SSH and SSE are independent.

**Theorem 3.5** Let \( y \) be \( N_n(X\beta, \sigma^2 I) \) and define the statistic

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

\[
= \frac{(C\hat{\beta} - t)'[C(X'X)^{-1}C']^{-1}(C\hat{\beta} - t)/q}{SSE/(n - k - 1)}
\]

(8)
where \( C \) is \( q \times (k + 1) \) of ranmk \( q \leq k + 1 \) and \( \hat{\beta} = (X'X)^{-1}X'y \). The distribution of \( F \) is as follows:

(i) If \( H_0 : C\beta = t \) is false, then

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

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

(ii) If \( H_0 : C\beta = t \) is true, then \( \lambda = 0 \) and

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

### 3.3 Tests on \( \beta_j \) and \( a'\beta \)

A test for an individual \( \beta_j \) can be obtained using either the full-and-reduced-model approach in section 2 or the general linear hypothesis approach in section 3.

The test statistic for \( H_0 : \beta_k = 0 \) using a full and reduced model is then

\[
F = \frac{\hat{\beta}'X'y - \hat{\beta}^*X_1y}{SSE/(n - k - 1)},
\]
which is distributed as $F(1, n - k - 1)$ if $H_0$ is true.

To test $H_0 : \mathbf{a}'\mathbf{\beta} = 0$ for a single linear combination, for example, $\mathbf{a}' = (0, 1, -1, 1)$, we use $\mathbf{a}'$ in place of the matrix $\mathbf{C}$ in $H_0 : \mathbf{C}\mathbf{\beta} = 0$. Then $q = 1$, and the test statistic becomes

$$F = \frac{(\mathbf{a}'\hat{\mathbf{\beta}})'[\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}]^{-1}(\mathbf{a}'\hat{\mathbf{\beta}})}{SSE/(n - k - 1)} = \frac{(\mathbf{a}'\hat{\mathbf{\beta}})^2}{s^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}, \quad (10)$$

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

One special case of $\mathbf{a}' = (0, \cdots, 0, 1, 0, \cdots, 0)$, where 1 is in the $j$th position. This gives

$$F = \frac{\hat{\beta}_j^2}{s^2g_{jj}}, \quad (11)$$

where $g_{jj}$ is the $j$th diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$. If $H_0 : \beta_j = 0$ is true, $F$ 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 $F$. (The $p$-value is the probability that $F(1, n - k - 1)$ exceeds the observed value of $F$.)
Since the $F$-statistic has 1 and $n - k - 1$ degrees of freedom, we can equivalently use the $t$-statistic

$$t_j = \frac{\hat{\beta}_j}{s\sqrt{g_{jj}}} \quad (12)$$

to test the effect of $\beta_j$.

4 Confidence Intervals and Prediction Intervals

4.1 Confidence Region for $\beta$

If $C = I$ in (8), $q$ becomes $k + 1$, and $t = \beta$, we obtain a central $F$-distribution and make the probability statement

$$P[(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta)/(k + 1)s^2 \leq F_{\alpha,k+1,n-k-1}] = 1 - \alpha,$$
where \( s^2 = SSH/(n - k - 1) \). From this statement, a \( 100(1 - \alpha)\% \) joint confidence region for \( \beta_0, \beta_1, \ldots, \beta_k \) in \( \beta \) is given by all vectors \( \beta \) that satisfy

\[
(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) \leq (k + 1)s^2F_{\alpha,k+1,n-k-1}.
\]

For \( k = 1 \), this region can be plotted as an ellipse in two dimensions. For \( k > 1 \), the elliptical region is unwidely to interpret.

### 4.2 Confidence Interval for \( \beta_j \)

If \( \beta_j \neq 0 \), we can subtract \( \beta_j \) in (12) so that

\[
t_j = \frac{\hat{\beta}_j - \beta_j}{s\sqrt{g_{jj}}}
\]

has the central \( t \)-distribution, where \( g_{jj} \) is the \( j \)th diagonal element of \( (X'X)^{-1} \). Then

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

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

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

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

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

**Example 4.1**  *Example 8.6.2 (Rencher and Schaalje(2008), pp.210).*

### 4.3 Confidence Interval for \( a' \beta \)

If \( a' \beta \neq 0 \), we can subtract \( a' \beta \) from \( a' \hat{\beta} \) in (11) to obtain

\[ F = \frac{(a' \beta - a' \beta)^2}{s^2 a'(X'X)^{-1}a}, \]
which is distributed as $F(1, n - k - 1)$. Then

$$t = \frac{a'\beta - a'\hat{\beta}}{s \sqrt{a'(X'X)^{-1}a}},$$

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

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

### 4.4 Confidence Interval for $E(y)$

Let $x_0 = (1, x_{01}, x_{02}, \cdots, x_{0k})'$ denote a particular choice of $x$. Note that $x_0$ need not be one of the $x$’s in the sample. Let $y_0$ be an observation corresponding to $x_0$. Then

$$y_0 = X_0'\beta + \epsilon,$$

and

$$E(y_0) = x_0'\beta.$$  \hspace{1cm} (13)
The minimum variance unbiased estimator of $E(y_0)$ is given by

$$\widehat{E}(y_0) = x'_0 \hat{\beta}. \quad (14)$$

Since (13) and (14) are of the form $a' \beta$ and $a' \hat{\beta}$, respectively, we obtain a $100 \times (1 - \alpha)\%$ confidence interval for $E(y_0)$:

$$x'_0 \hat{\beta} \pm t_{\alpha/2, n-k-1} s \sqrt{x'_0 (X'X)^{-1} x_0}.$$ 

The confidence interval can also be expressed in terms of the centered model,

$$E(y_0) = \alpha + \beta'_1 (x_{01} - \bar{x}_1),$$

$$\widehat{E}(y_0) = \bar{y} + \hat{\beta}'_1 (x_{01} - \bar{x}_1),$$

$$\bar{y} + \hat{\beta}'_1 (x_{01} - \bar{x}_1) \pm t_{\alpha/2, n-k-1} s \sqrt{\frac{1}{n} + (x_{01} - \bar{x}_1)'(X'_c X_c)^{-1} (x_{01} - \bar{x}_1)}.$$
4.5 Prediction interval for a future observation

For a future observation, we have

\[ \text{var}(y_0 - \hat{y}_0) = \text{var}(y_0) + \text{var}(x'_0\hat{\beta}) = \sigma^2 + \sigma^2 x'_0 (X'X)^{-1} x_0 \]

\[ = \sigma^2 [1 + x'_0 (X'X)^{-1} x_0], \]

which is estimated by \( s^2 [1 + x'_0 (X'X)^{-1} x_0] \). Note that in the derivation, we used the fact that \( y_0 \) is independent of \( \hat{y}_0 \). Since \( s^2 \) is independent of both \( y_0 \) and \( \hat{y}_0 = x'_0\hat{\beta} \), we have that

\[ t = \frac{y_0 - \hat{y}_0}{s \sqrt{1 + x'_0 (X'X)^{-1} x_0}} \]

is distributed as \( t(n - k - 1) \), and the 100(1 - \( \alpha \))\% prediction interval is

\[ x'_0\hat{\beta} \pm t_{\alpha/2, n-k-1} s \sqrt{1 + x'_0 (X'X)^{-1} x_0}. \]

In terms of the centered model, the 100(1 - \( \alpha \))\% prediction interval becomes

\[ \bar{y} + \hat{\beta}'_1 (x_{01} - \bar{x}_1) \pm t_{\alpha/2, n-k-1} s \sqrt{1 + \frac{1}{n} + (x_{01} - \bar{x}_1)'(X'_cX_c)^{-1}(x_{01} - \bar{x}_1)}. \]
4.6 Confidence interval for $\sigma^2$

Since $(n - k - 1)s^2/\sigma^2$ is $\chi^2(n - k - 1)$, the 100$(1 - \alpha)$% confidence interval for $\sigma^2$ is

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

A 100$(1 - \alpha)$% confidence interval for $\sigma$ is

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

5 Likelihood Ratio Tests

Suppose that $\mathbf{x} = (x_1, \cdots, x_n)$ has density or frequency function $p(\mathbf{x}, \theta)$ and we wish to test $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$. The test statistic we want to consider is the likelihood ratio given by

$$LR = \frac{\sup\{p(\mathbf{x}, \theta) : \theta \in \Theta_0\}}{\sup\{p(\mathbf{x}, \theta) : \theta \in \Theta_1\}}.$$
Tests that reject $H_0$ for small value of $LR$ are called likelihood ratio tests.

In the cases we shall consider, $p(x, \theta)$ is a continuous function of $\theta$ and $\Theta_0$ is of smaller dimension than $\Theta = \Theta_0 \cap \Theta_1$ so that the likelihood ratio equals the test statistic

$$\lambda(x) = \frac{\sup\{p(x, \theta) : \theta \in \Theta_0\}}{\sup\{p(x, \theta) : \theta \in \Theta\}},$$

whose computation is often simple. It follows that (Wilks, 1938) for $n \to \infty$

$$-2 \log \lambda(x) \to \chi^2_d,$$

where $d = dim(\Theta) - dim(\Theta_0)$. In some cases, the $\chi^2$ approximation is not needed because $\lambda(x)$ turns out to be a function of a familiar test statistic, such as $t$ or $F$, whose exact distribution is available.

**Theorem 5.1** If $y$ is $N_n(X\beta, \sigma^2 I)$, the likelihood ratio test for $H_0 : \beta = 0$ can be test on

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

We reject $H_0$ if $F > F_{\alpha, k+1, n-k-1}$. 
PROOF: To find $\sup L(\beta, \sigma^2)$, we use the MLEs $\hat{\beta} = (X'X)^{-1}X'y$ and $\hat{\sigma}^2 = (y - X\hat{\beta})'(y - X\hat{\beta})/n$. Substituting we have

$$\sup L(\beta, \sigma^2) = \frac{n^{n/2}e^{-n/2}}{(2\pi)^{n/2}[(y - X\hat{\beta})'(y - X\hat{\beta})]^{n/2}}.$$ 

To find $\sup L(0, \sigma^2)$, we solve $\partial L(0, \sigma^2)/\partial \sigma^2 = 0$ to obtain

$$\hat{\sigma}_0^2 = by'y/n.$$ 

Then

$$\sup_{H_0} L(\beta, \sigma^2) = L(0, \hat{\sigma}_0^2) = \frac{n^{n/2}e^{-n/2}}{(2\pi)^{n/2}(y'y)^{n/2}}.$$ 

Thus, we have

$$\lambda(x) = \left[\frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{y'y}\right]^{n/2} = \left[\frac{1}{1 + (k + 1)F/(n - k - 1)}\right]^{n/2},$$ 

where

$$F = \frac{\hat{\beta}'X'y/(k + 1)}{(y'y - \hat{\beta}'X'y)/(n - k - 1)}.$$
Thus, rejecting $h_0 : \beta = 0$ for a small value of $\lambda(x)$ is equivalent to rejecting $H_0$ for a large value of $F$.

The four steps of deriving likelihood ratio tests:

1. Calculate the MLE $\hat{\theta}$ of $\theta$.

2. Calculate the MLE $\hat{\theta}_0$ where $\theta$ may vary only over $\Theta_0$.

3. Form $\lambda(x) = p(x, \hat{\theta}_0)/p(x, \hat{\theta})$.

4. Find a function $h$ which is strictly decreasing on the range of $\lambda$ such that $h(\lambda(x))$ has a simple form and a tabled distribution under $h_0$. Since $h(\lambda(x))$ is equivalent to $\lambda(x)$ we specify the size $\alpha$ likelihood ratio test through the test statistic $h(\lambda(x))$ and its $(1 - \alpha)$th quantile obtained from the table.