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 $\boldsymbol{\beta}$ in the model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. We will assume throughout the chapter that $\boldsymbol{y}$ is $N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I})$, where $\boldsymbol{X}$ 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 $H_0: \boldsymbol{\beta}_1 = 0$, where $\boldsymbol{\beta}_1 = (\beta_1, \cdots, \beta_k)'$. Note that we wish to test $H_0: \boldsymbol{\beta}_1 = 0$, not $H_0: \boldsymbol{\beta} = 0$, where $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1')'$. Since β_0 is usually not zero, rejection of $H_0: \boldsymbol{\beta} = 0$ might be due solely to β_0, and we would not learn if the x's predict y.

For this test, we use the centered model

$$\boldsymbol{y} = (\boldsymbol{j}, \boldsymbol{X}_c) \begin{pmatrix} \alpha \\ \boldsymbol{\beta}_1 \end{pmatrix} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{X}_c = (\boldsymbol{I} - \frac{1}{n}\boldsymbol{J})\boldsymbol{X}_1$ is the centered matrix and $\boldsymbol{X}_1$ contains all the columns of $\boldsymbol{X}$ except the first. Note the partition identity

$$\begin{aligned} \boldsymbol{y}'(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J})\boldsymbol{y} &= SSR + SSE \\ &= \boldsymbol{y}'\boldsymbol{X}_c(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{X}_c'\boldsymbol{y} + \boldsymbol{y}'(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J})\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}_c(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{X}_c'\boldsymbol{y} \\ &= \boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} + \boldsymbol{y}'(I - \frac{1}{n}\boldsymbol{J} - \boldsymbol{A})\boldsymbol{y}, \end{aligned} \tag{2}$$

where $\boldsymbol{A} = \boldsymbol{X}_c(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{X}_c'$.

In the following theorem we establish some properties of the matrices of the quadratic forms in (2).

Theorem 1.1 *The matrices* $\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}$, $\boldsymbol{A} = \boldsymbol{X}_c(\boldsymbol{X}'_c\boldsymbol{X}_c)^{-1}\boldsymbol{X}'_c$, *and* $\boldsymbol{I} - \frac{1}{n}\boldsymbol{J} - \boldsymbol{A}$ *have the following properties:*

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

(ii) $\boldsymbol{A}$ *is idempotent of rank* k.

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

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

PROOF: Part (i) follows from $\boldsymbol{X}'_c\boldsymbol{j} = 0$. The other parts are left as exercises.

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

(i) SSR/σ^2 *is* $\chi^2(k, \lambda_1)$, *where* $\lambda_1 = \hat{\boldsymbol{\beta}}'_1\boldsymbol{X}'_c\boldsymbol{X}_c\hat{\boldsymbol{\beta}}_1/[2\sigma^2]$.

(ii) SSE/σ^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 $\boldsymbol{y}$ is $N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I})$, then SSR and SSE are independent.*

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

Theorem 1.4 *If $\boldsymbol{y}$ is $N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I})$, 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 $H_0: \boldsymbol{\beta}_1 = 0$ is false, then

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

where $\lambda_1 = \boldsymbol{\beta}_1'\boldsymbol{X}_c'\boldsymbol{X}_c\boldsymbol{\beta}_1/2\sigma^2$.

(ii) If $H_0: \boldsymbol{\beta}_1 = 0$ is true, then $\lambda_1 = 0$ and

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

Table 1: Analysis of Variance (ANOVA) for F-test of $H_0 : \boldsymbol{\beta}_1 = 0$.

Source of Variation	d.f.	Sum of Squares	Mean Square	Expected Mean Squa
Due to $\boldsymbol{\beta}_1$	k	SSR	SSR/k	$\sigma^2 + \frac{1}{k}\boldsymbol{\beta}_1'\boldsymbol{X}_c'\boldsymbol{X}_c\boldsymbol{\beta}$
Error	$n-k-1$	SSE	$SSE/(n-k-1)$	σ^2
Total	$n-1$	SST		

PROOF: Exercise.

Note that $\lambda_1 = 0$ if and only if $\boldsymbol{\beta}_1 = 0$, since $\boldsymbol{X}_c'\boldsymbol{X}_c$ is positive definite.

The test for $H_0 : \boldsymbol{\beta}_1 = 0$ 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 α percentile of the centrail F-distribution. In table 1, we summarize the results leading to the F-test.

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

Example 1.1 *For the data in Table 7.1, test* $H_0: \boldsymbol{\beta}_1 = 0$*. In this case,*

$$\begin{aligned}
\boldsymbol{\beta}_1 &= (\beta_1, \beta_2)', \\
\boldsymbol{X}'\boldsymbol{y} &= (90, 482, 872)', \\
\hat{\boldsymbol{\beta}} &= (5.3754, 3.0118, -1.2855)', \\
\boldsymbol{y}'\boldsymbol{y} &= \sum_{i=1}^{12} y_i^2 = 840, \\
\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} &= 814.5410, \\
n\bar{y}^2 &= 675.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
SSR &= \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - n\bar{y}^2 = 139.5410, \\
SSE &= \boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} = 25.4590, \\
SST &= \boldsymbol{y}'\boldsymbol{y} - n\bar{y}^2 = 165.
\end{aligned}$$

Table 2: ANOVA for overall regression test for the data in Table 7.1

Source	d.f.	Sum of Squares	Mean Square	F
Due to $\boldsymbol{\beta}_1$	2	139.5410	69.7705	24.665
Error	9	25.4590	2.8288	
Total	11	165.0000		

The F-test is given in Table 2. Since $24.665 > F_{.05,2,9} = 4.26$, we reject H_0 and conclude that at least one of β_1 or β_2 is not zero.

2 Test on a subset of the β'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 β's to be tested have been arranged last in $\boldsymbol{\beta}$, with a corresponding arrangement of the columns of $\boldsymbol{X}$, i.e., the model becomes

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = (\boldsymbol{X}_1, \boldsymbol{X}_2) \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \boldsymbol{\epsilon} \\ = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}, \tag{3}$$

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

To test $H_0 : \boldsymbol{\beta}_2 = 0$ versus $H_1 : \boldsymbol{\beta}_2 \neq 0$, we use a full-reduced-model approach. The full model is given by (3), the reduced model becomes

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1^* + \boldsymbol{\epsilon}^*.$$

Typically, β_1^* and $\boldsymbol{\epsilon}^*$ will be different from $\boldsymbol{\beta}_1$ and $\boldsymbol{\epsilon}$ in the full model (unless $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are orthogonal). For $\boldsymbol{y}'\boldsymbol{y}$ we have the following partitioning:

$$\boldsymbol{y}'\boldsymbol{y} = (\boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}) + (\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}_1^{*'}\boldsymbol{X}_1\boldsymbol{y}) + \hat{\boldsymbol{\beta}}_1^{*'}\boldsymbol{X}_1\boldsymbol{y} \\ = SSE + SS(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) + SS(\boldsymbol{\beta}_1^*) \tag{4}$$

where $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ and $\hat{\boldsymbol{\beta}}_1^* = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{y}$. Note that

$$SS(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = (\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - n\bar{y}^2) - (\hat{\boldsymbol{\beta}}_1^{*'}\boldsymbol{X}_1\boldsymbol{y} - n\bar{y}^2) = SSR(full) - SSR(reduced).$$

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

$$\begin{aligned} \boldsymbol{y}'\boldsymbol{y} &= \boldsymbol{y}'[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y} + \boldsymbol{y}'[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' - \boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1']\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{X}_1(\boldsymbol{X}_1' \\ &= \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{A}_1)\boldsymbol{y} + \boldsymbol{y}'(\boldsymbol{A}_1 - \boldsymbol{A}_2)\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{A}_2\boldsymbol{y}, \end{aligned} \tag{5}$$

where $\boldsymbol{A}_1 = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and $\boldsymbol{A}_2 = \boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'$.

Theorem 2.1 *The matrix* $\boldsymbol{A}_1 - \boldsymbol{A}_2 = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' - \boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1$ *is idempotent with rank* h*, where* h *is the number of elements in* $\boldsymbol{\beta}_2$*.*

PROOF: Multiplying $\boldsymbol{X}$ by $\boldsymbol{A}_1$, we have

$$\boldsymbol{A}_1\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{X},$$

or

$$\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}. \tag{6}$$

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

$$\begin{aligned} \boldsymbol{X}_1 &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{X}_1, \\ \boldsymbol{X}_2 &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{X}_2. \end{aligned} \tag{7}$$

Using (7) and its transpose, we have

$$\boldsymbol{A}_1\boldsymbol{A}_2 = \boldsymbol{A}_2 \quad and \quad \boldsymbol{A}_2\boldsymbol{A}_1 = \boldsymbol{A}_2.$$

Note $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are both idempotent. Thus

$$\begin{aligned} (\boldsymbol{A}_1 - \boldsymbol{A}_2)^2 &= \boldsymbol{A}_1 - \boldsymbol{A}_1\boldsymbol{A}_2 - \boldsymbol{A}_2\boldsymbol{A}_1 + \boldsymbol{A}_2^2 \\ &= \boldsymbol{A}_1 - \boldsymbol{A}_2 - \boldsymbol{A}_2 + \boldsymbol{A}_2 = \boldsymbol{A}_1 - \boldsymbol{A}_2. \end{aligned}$$

Hence, $\boldsymbol{A}_1 - \boldsymbol{A}_2$ is idempotent.

$$\begin{aligned} rank(\boldsymbol{A}_1 - \boldsymbol{A}_2) &= tr(\boldsymbol{A}_1 - \boldsymbol{A}_2) = tr(\boldsymbol{A}_1) - tr(\boldsymbol{A}_2) \\ &= tr(I_{k+1}) - tr(I_{k-h+1}) = k + 1 - (k - h + 1) = h. \end{aligned}$$

The next theorem gives the distribution of $\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{A}_1)\boldsymbol{y}$ and $\boldsymbol{y}'(\boldsymbol{A}_1-\boldsymbol{A}_2)\boldsymbol{y}$ and shows that they are independent.

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

(i) $\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{A}_1)\boldsymbol{y}/\sigma^2$ *is* $\chi^2(n-k-1)$.

(ii) $\boldsymbol{y}'(\boldsymbol{A}_1-\boldsymbol{A}_2)\boldsymbol{y}/\sigma^2$ *is* $\chi^2(h,\lambda_1)$, *where*

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

(iii) $\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{A}_1)\boldsymbol{y}$ *and* $\boldsymbol{y}'(\boldsymbol{A}_1-\boldsymbol{A}_2)\boldsymbol{y}$ *are independent.*

PROOF: (i) and (ii) are left for exercises. (iii) Note that $\boldsymbol{A}_2=\boldsymbol{A}_1\boldsymbol{A}_2$, hence

$$(\boldsymbol{I}-\boldsymbol{A}_1)(\boldsymbol{A}_1-\boldsymbol{A}_2)=(\boldsymbol{I}-\boldsymbol{A}_1)(\boldsymbol{A}_1-\boldsymbol{A}_1\boldsymbol{A}_2)=\boldsymbol{O}.$$

Since $\boldsymbol{X}_2'\boldsymbol{X}_2-\boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2$ is positive definite (see problem 8.10 of Rencher and Schaalje (2008)), $\lambda_1=0$ if and only if $\boldsymbol{\beta}=0$.

Theorem 2.3 *Let $\boldsymbol{y}$ be $N_n(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$ and define an F-statistic as follows:*

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

The distribution of F is as follows:

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

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

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

(ii) If $H_0: \boldsymbol{\beta}_2 = 0$ 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 $H_0: \ \boldsymbol{\beta}_2 = 0$.

Source of Variation	d.f.	Sum of Squares	Mean Square	F-statistic
Due to $\boldsymbol{\beta}_2$ adjusted for $\boldsymbol{\beta}_1$	h	$SS(\boldsymbol{\beta}_2 \vert \boldsymbol{\beta}_1)$	$SS(\boldsymbol{\beta}_2 \vert \boldsymbol{\beta}_1)/h$	$\frac{SS(\boldsymbol{\beta}_2 \vert \boldsymbol{\beta}_1)/h}{SSE/(n-k-1)}$
Error	$n-k-1$	SSE	$SSE/(n-k-1)$	
Total	$n-1$	SST		

The test for $H_0: \ \boldsymbol{\beta}_2 = 0$ 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

$$E[SS(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)/h] = \sigma^2 + \frac{1}{h}\boldsymbol{\beta}_2'[\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2]\boldsymbol{\beta}_2,$$

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 : $\boldsymbol{C\beta} = 0$ and $H_0 : \boldsymbol{C\beta} = \boldsymbol{t}$

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

General linear hypothesis: the hypothesis $H_0 : \boldsymbol{C\beta} = 0$, where $\boldsymbol{C}$ 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 $H_1 : \boldsymbol{C\beta} \neq 0$. Note that the formulation $H_0 : \boldsymbol{C\beta} = 0$ includes as special cases the hypotheses in last two sections.

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

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

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

$$\lambda = (\boldsymbol{C\beta})'[\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}']^{-1}\boldsymbol{C\beta}/2\sigma^2;$$

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

(iv) SSH and SSE are independent.

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

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

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

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

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

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

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

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

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

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

PROOF: Exercise.

This test is usually called the *general linear hypothesis test*. The degrees of freedom q is the number of linear combinations in $\boldsymbol{C\beta}$. 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 $\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'$ 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

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

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

Theorem 3.3 *The F-test in theorem 3.2 for the general linear hypothesis $H_0 : \boldsymbol{C}\boldsymbol{\beta} = 0$ is a full-and-reduced-model test.*

PROOF: The reduced model under H_0 is

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

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

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

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

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

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

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

thus establishing that the F-test in theorem 3.2 for $H_0 : \boldsymbol{C\beta} = 0$ is a full-and-reduced-model test.

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

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

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

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

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

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

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

(iv) SSH and SSE are independent.

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

First, note that

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

so that

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

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

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

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

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

$$\begin{aligned} F &= \frac{SSH/q}{SSE/(n-k-1)} \\ &= \frac{(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{t})'[\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}']^{-1}(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{t})/q}{SSE/(n-k-1)} \end{aligned} \quad (8)$$

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

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

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

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

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

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

3.3 Tests on β_j and $\boldsymbol{a}'\boldsymbol{\beta}$

A test for an individual β_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

$$F = \frac{\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}^{*'}\boldsymbol{X}_1\boldsymbol{y}}{SSE/(n-k-1)}, \tag{9}$$

which is distributed as $F(1, n-k-1)$ if H_0 is true.

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

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

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

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

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

where g_{jj} is the jth diagonal element of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$. 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-statustic has 1 and $n-k-1$ degrees of freedom, we can equivalently use the t-statistic

$$t_j = \frac{\hat{\beta}_j}{s\sqrt{g_{jj}}} \tag{12}$$

to test the effect of β_j.

4 Confidence Intervals and Prediction Intervals

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

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

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

where $s^2 = SSH/(n-k-1)$. From this statement, a $100(1-\alpha)\%$ joint confidence region for $\beta_0, \beta_1, \cdots, \beta_k$ in $\boldsymbol{\beta}$ is given by all vectors $\boldsymbol{\beta}$ that satisfy

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

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 β_j

If $\beta_j \neq 0$, we can subtract β_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 jth diagonal element of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$. 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 β_j is $1 - \alpha$. After taking the sample, the $100(1-\alpha)\%$ confidence interval for β_j,

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

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

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

4.3 Confidence Interval for $\boldsymbol{a}'\boldsymbol{\beta}$

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

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

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

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

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

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

4.4 Confidence Interval for $E(\boldsymbol{y})$

Let $\boldsymbol{x}_0 = (1, x_{01}, x_{02}, \cdots, x_{0k})'$ denote a particular choice of $\boldsymbol{x}$. Note that $\boldsymbol{x}_0$ need not be one of the $\boldsymbol{x}$'s in the sample. Let y_0 be an observation corresponding to $\boldsymbol{x}_0$. Then

$$y_0 = \boldsymbol{X}_0'\boldsymbol{\beta} + \epsilon,$$

and

$$E(y_0) = \boldsymbol{x}_0'\boldsymbol{\beta}. \tag{13}$$

The minimum variance unbiased estimator of $E(y_0)$ is given by

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

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

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

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

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

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

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

4.5 Prediction interval for a future observation

For a future observation, we have

$$var(y_0 - \hat{y_0}) = var(y_0) + var(\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}) = \sigma^2 + \sigma^2 \boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0$$
$$= \sigma^2[1 + \boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0],$$

which is estimated by $s^2[1+\boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0]$. 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 $\hat{y_0} = \boldsymbol{x}_0'\hat{\boldsymbol{\beta}}$, we have that

$$t = \frac{y_0 - \hat{y_0}}{s\sqrt{1 + \boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0}}$$

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

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

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

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

4.6 Confidence interval for σ^2

Since $(n-k-1)s^2/\sigma^2$ is $\chi^2(n-k-1)$, the $100(1-\alpha)\%$ confidence interval for σ^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 σ 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 $\boldsymbol{x} = (x_1, \cdots, x_n)$ has density or frequency function $p(\boldsymbol{x}, \theta)$ 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

$$LR = \frac{\sup\{p(\boldsymbol{x}, \theta) : \theta \in \Theta_0\}}{\sup\{p(\boldsymbol{x}, \theta) : \theta \in \Theta_1\}}.$$

Tests that reject H_0 for small value of LR are called *likelihood ratio tests*.

In the cases we shall consider, $p(\boldsymbol{x}, \theta)$ is a continuous function of θ and Θ_0 is of smaller dimension than $\Theta = \Theta_0 \cap \Theta_1$ so that the likelihood ratio equals the test statistic

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

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

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

where $d = dim(\Theta) - dim(\Theta_0)$. In some cases, the χ^2 approximation is not needed because $\lambda(\boldsymbol{x})$ 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 $\boldsymbol{y}$ is $N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I})$, the likelihood ratio test for $H_0 : \boldsymbol{\beta} = 0$ can be test on*

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

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

PROOF: To find $\sup L(\boldsymbol{\beta}, \sigma^2)$, we use the MLEs $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ and $\hat{\sigma}^2 = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})/n$. Substituting we have

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

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

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

Then

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

Thus, we have

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

where

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

Thus, rejecting $h_0 : \boldsymbol{\beta} = 0$ for a small value of $\lambda(\boldsymbol{x})$ 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 θ.

(2) Calculate the MLE $\hat{\theta}_0$ where θ may vary only over Θ_0.

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

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