CHAPTER 10

Inferences About Normal Models

10.1 The Distributions of Certain Quadratic Forms

A homogeneous polynomial of degree 2 in n variables is called a *quadratic* form in those variables. If both the variables and the coefficients are real, the form is called a *real quadratic* form. Only real quadratic forms will be considered in this book. To illustrate, the form $X_1^2 + X_1X_2 + X_2^2$ is a quadratic form in the two variables X_1 and X_2; the form $X_1^2 + X_2^2 + X_3^2 - 2X_1X_2$ is a quadratic form in the three variables X_1, X_2, and X_3; but the form $(X_1 - 1)^2 + (X_2 - 2)^2 = X_1^2 + X_2^2 - 2X_1 - 4X_2 + 5$ is not a quadratic form in X_1 and X_2, although it is a quadratic form in the variables $X_1 - 1$ and $X_2 - 2$.

Let $\bar{X}$ and S^2 denote, respectively, the mean and the variance of a random sample $X_1, X_2, \ldots, X_n$ from an arbitrary distribution. Thus

$$nS^2 = \sum_1^n (X_i - \bar{X})^2 = \sum_1^n \left(X_i - \frac{X_1 + X_2 + \cdots + X_n}{n}\right)^2$$

$$= \frac{n-1}{n}(X_1^2 + X_2^2 + \cdots + X_n^2)$$

$$-\frac{2}{n}(X_1X_2 + \cdots + X_1X_n + \cdots + X_{n-1}X_n)$$

is a quadratic form in the n variables $X_1, X_2, \ldots, X_n$. If the sample arises from a distribution that is $N(\mu, \sigma^2)$, we know that the random variable nS^2/σ^2 is $\chi^2(n-1)$ regardless of the value of μ. This fact proved useful in our search for a confidence interval for σ^2 when μ is unknown.

It has been seen that tests of certain statistical hypotheses require a statistic that is a quadratic form. For instance, Example 2, Section 9.2, made use of the statistic $\sum_1^n X_i^2$, which is a quadratic form in the variables $X_1, X_2, \ldots, X_n$. Later in this chapter, tests of other statistical hypotheses will be investigated, and it will be seen that functions of statistics that are quadratic forms will be needed to carry out the tests in an expeditious manner. But first we shall make a study of the distribution of certain quadratic forms in normal and independent random variables.

The following theorem will be proved in Section 10.9.

Theorem 1. *Let $Q = Q_1 + Q_2 + \cdots + Q_{k-1} + Q_k$, where $Q, Q_1, \ldots, Q_k$ are $k+1$ random variables that are real quadratic forms in n independent random variables which are normally distributed with the means $\mu_1, \mu_2, \ldots, \mu_n$ and the same variance σ^2. Let Q/σ^2, $Q_1/\sigma^2, \ldots, Q_{k-1}/\sigma^2$ have chi-square distributions with degrees of freedom $r, r_1, \ldots, r_{k-1}$, respectively. Let Q_k be nonnegative. Then:*

(a) *$Q_1, \ldots, Q_k$ are independent, and hence*
(b) *Q_k/σ^2 has a chi-square distribution with $r - (r_1 + \cdots + r_{k-1}) = r_k$ degrees of freedom.*

Three examples illustrative of the theorem will follow. Each of these examples will deal with a distribution problem that is based on the remarks made in the subsequent paragraph.

Let the random variable X have a distribution that is $N(\mu, \sigma^2)$. Let a and b denote positive integers greater than 1 and let $n = ab$. Consider a random sample of size $n = ab$ from this normal distribution. The observations of the random sample will be denoted by the symbols

$$\begin{array}{cccccc} X_{11}, & X_{12}, & \ldots, & X_{1j}, & \ldots, & X_{1b} \\ X_{21}, & X_{22}, & \ldots, & X_{2j}, & \ldots, & X_{2b} \\ \vdots & & & & & \\ X_{i1}, & X_{i2}, & \ldots, & X_{ij}, & \ldots, & X_{ib} \\ \vdots & & & & & \\ X_{a1}, & X_{a2}, & \ldots, & X_{aj}, & \ldots, & X_{ab}. \end{array}$$

In this notation, the first subscript indicates the row, and the second subscript indicates the column in which the observation appears. Thus X_{ij} is in row i and column j, $i = 1, 2, \ldots, a$ and $j = 1, 2, \ldots, b$. By assumption these $n = ab$ random variables are independent, and each has the same normal distribution with mean μ and variance σ^2. Thus, if we wish, we may consider each row as being a random sample of size b from the given distribution; and we may consider each column as being a random sample of size a from the given distribution. We now define $a + b + 1$ statistics. They are

$$\bar{X}_{..} = \frac{X_{11} + \cdots + X_{1b} + \cdots + X_{a1} + \cdots + X_{ab}}{ab} = \frac{\sum_{i=1}^{a} \sum_{j=1}^{b} X_{ij}}{ab},$$

$$\bar{X}_{i.} = \frac{X_{i1} + X_{i2} + \cdots + X_{ib}}{b} = \frac{\sum_{j=1}^{b} X_{ij}}{b}, \qquad i = 1, 2, \ldots, a,$$

and

$$\bar{X}_{.j} = \frac{X_{1j} + X_{2j} + \cdots + X_{aj}}{a} = \frac{\sum_{i=1}^{a} X_{ij}}{a}, \qquad j = 1, 2, \ldots, b.$$

The statistic $\bar{X}_{..}$ is the mean of the random sample of size $n = ab$; the statistics $\bar{X}_{1.}, \bar{X}_{2.}, \ldots, \bar{X}_{a.}$ are, respectively, the means of the rows; and the statistics $\bar{X}_{.1}, \bar{X}_{.2}, \ldots, \bar{X}_{.b}$ are, respectively, the means of the columns. Three examples illustrative of the theorem will follow.

Example 1. Consider the variance S^2 of the random sample of size $n = ab$. We have the algebraic identity

$$
\begin{aligned}
abS^2 &= \sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij} - \bar{X}_{..})^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}[(X_{ij} - \bar{X}_{i.}) + (\bar{X}_{i.} - \bar{X}_{..})]^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij} - \bar{X}_{i.})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{X}_{i.} - \bar{X}_{..})^2 \\
&\quad + 2\sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij} - \bar{X}_{i.})(\bar{X}_{i.} - \bar{X}_{..}).
\end{aligned}
$$

The last term of the right-hand member of this identity may be written

$$2\sum_{i=1}^{a}\left[(\bar{X}_{i.} - \bar{X}_{..})\sum_{j=1}^{b}(X_{ij} - \bar{X}_{i.})\right] = 2\sum_{i=1}^{a}[(\bar{X}_{i.} - \bar{X}_{..})(b\bar{X}_{i.} - b\bar{X}_{i.})] = 0,$$

and the term

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{X}_{i.} - \bar{X}_{..})^2$$

may be written

$$b\sum_{i=1}^{a}(\bar{X}_{i.} - \bar{X}_{..})^2.$$

Thus

$$abS^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij} - \bar{X}_{i.})^2 + b\sum_{i=1}^{a}(\bar{X}_{i.} - \bar{X}_{..})^2,$$

or, for brevity,

$$Q = Q_1 + Q_2.$$

Clearly, Q, Q_1, and Q_2 are quadratic forms in the $n = ab$ variables X_{ij}. We shall use the theorem with $k = 2$ to show that Q_1 and Q_2 are independent. Since S^2 is the variance of a random sample of size $n = ab$ from the given normal distribution, then abS^2/σ^2 has a chi-square distribution with $ab - 1$ degrees of freedom. Now

$$\frac{Q_1}{\sigma^2} = \sum_{i=1}^{a}\left[\frac{\sum_{j=1}^{b}(X_{ij} - \bar{X}_{i.})^2}{\sigma^2}\right].$$

For each fixed value of i, $\sum_{j=1}^{b}(X_{ij} - \bar{X}_{i.})^2/b$ is the variance of a random sample of size b from the given normal distribution, and, accordingly, $\sum_{j=1}^{b} (X_{ij} - \bar{X}_{i.})^2/\sigma^2$ has a chi-square distribution with $b - 1$ degrees of freedom. Because the X_{ij} are independent, Q_1/σ^2 is the sum of a independent random variables, each having a chi-square distribution with $b - 1$ degrees of freedom. Hence Q_1/σ^2 has a chi-square distribution with $a(b - 1)$ degrees of freedom. Now $Q_2 = b \sum_{i=1}^{a} (\bar{X}_{i.} - \bar{X}_{..})^2 \ge 0$. In accordance with the theorem, Q_1 and Q_2 are independent, and Q_2/σ^2 has a chi-square distribution with $ab - 1 - a(b - 1) = a - 1$ degrees of freedom.

Example 2. In abS^2 replace $X_{ij} - \bar{X}_{..}$ by $(X_{ij} - \bar{X}_{.j}) + (\bar{X}_{.j} - \bar{X}_{..})$ to obtain

$$abS^2 = \sum_{j=1}^{b} \sum_{i=1}^{a} [(X_{ij} - \bar{X}_{.j}) + (\bar{X}_{.j} - \bar{X}_{..})]^2,$$

or

$$abS^2 = \sum_{j=1}^{b} \sum_{i=1}^{a} (X_{ij} - \bar{X}_{.j})^2 + a \sum_{j=1}^{b} (\bar{X}_{.j} - \bar{X}_{..})^2,$$

or, for brevity,

$$Q = Q_3 + Q_4.$$

It is easy to show (Exercise 10.1) that Q_3/σ^2 has a chi-square distribution with $b(a - 1)$ degrees of freedom. Since $Q_4 = a \sum_{j=1}^{b} (\bar{X}_{.j} - \bar{X}_{..})^2 \ge 0$, the theorem enables us to assert that Q_3 and Q_4 are independent and that Q_4/σ^2 has a chi-square distribution with $ab - 1 - b(a - 1) = b - 1$ degrees of freedom.

Example 3. In abS^2 replace $X_{ij} - \bar{X}_{..}$ by $(\bar{X}_{i.} - \bar{X}_{..}) + (\bar{X}_{.j} - \bar{X}_{..}) + (X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})$ to obtain (Exercise 10.2)

$$abS^2 = b \sum_{i=1}^{a} (\bar{X}_{i.} - \bar{X}_{..})^2 + a \sum_{j=1}^{b} (\bar{X}_{.j} - \bar{X}_{..})^2 + \sum_{j=1}^{b} \sum_{i=1}^{a} (X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2,$$

or, for brevity,

$$Q = Q_2 + Q_4 + Q_5,$$

where Q_2 and Q_4 are as defined in Examples 1 and 2. From Examples 1 and 2, Q/σ^2, Q_2/σ^2, and Q_4/σ^2 have chi-square distributions with $ab - 1$, $a - 1$, and $b - 1$ degrees of freedom, respectively. Since $Q_5 \ge 0$, the theorem asserts that Q_2, Q_4, and Q_5 are independent and that Q_5/σ^2 has a chi-square distribution with $ab - 1 - (a - 1) - (b - 1) = (a - 1)(b - 1)$ degrees of freedom.

Once these quadratic form statistics have been shown to be independent, a multiplicity of F-statistics can be defined. For instance,

$$\frac{Q_4/[\sigma^2(b-1)]}{Q_3/[\sigma^2 b(a-1)]} = \frac{Q_4/(b-1)}{Q_3/[b(a-1)]}$$

has an F-distribution with $b - 1$ and $b(a - 1)$ degrees of freedom; and

$$\frac{Q_4/[\sigma^2(b-1)]}{Q_5/[\sigma^2(a-1)(b-1)]} = \frac{Q_4/(b-1)}{Q_5/(a-1)(b-1)}$$

has an F-distribution with $b - 1$ and $(a - 1)(b - 1)$ degrees of freedom. In the subsequent sections it will be seen that some likelihood ratio tests of certain statistical hypotheses can be based on these F-statistics.

EXERCISES

10.1. In Example 2 verify that $Q = Q_3 + Q_4$ and that Q_3/σ^2 has a chi-square distribution with $b(a - 1)$ degrees of freedom.

10.2. In Example 3 verify that $Q = Q_2 + Q_4 + Q_5$.

10.3. Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution $N(\mu, \sigma^2)$. Show that

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=2}^{n} (X_i - \bar{X}')^2 + \frac{n-1}{n}(X_1 - \bar{X}')^2,$$

where $\bar{X} = \sum_{i=1}^{n} X_i/n$ and $\bar{X}' = \sum_{i=2}^{n} X_i/(n-1)$.

Hint: Replace $X_i - \bar{X}$ by $(X_i - \bar{X}') - (X_1 - \bar{X}')/n$. Show that $\sum_{i=2}^{n} (X_i - \bar{X}')^2/\sigma^2$ has a chi-square distribution with $n - 2$ degrees of freedom. Prove that the two terms in the right-hand member are independent. What then is the distribution of

$$\frac{[(n-1)/n](X_1 - \bar{X}')^2}{\sigma^2}?$$

10.4. Let X_{ijk}, $i = 1, \ldots, a$; $j = 1, \ldots, b$; $k = 1, \ldots, c$, be a random sample of size $n = abc$ from a normal distribution $N(\mu, \sigma^2)$. Let $\bar{X}_{\ldots} = \sum_{k=1}^{c} \sum_{j=1}^{b} \sum_{i=1}^{a} X_{ijk}/n$ and $\bar{X}_{i..} = \sum_{k=1}^{c} \sum_{j=1}^{b} X_{ijk}/bc$. Show that

$$\sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} (X_{ijk} - \bar{X}_{\ldots})^2 = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} (X_{ijk} - \bar{X}_{i..})^2 + bc \sum_{i=1}^{a} (\bar{X}_{i..} - \bar{X}_{\ldots})^2.$$

Show that $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(X_{ijk}-\bar{X}_{i..})^2/\sigma^2$ has a chi-square distribution with $a(bc-1)$ degrees of freedom. Prove that the two terms in the right-hand member are independent. What, then, is the distribution of $bc\sum_{i=1}^{a}(\bar{X}_{i..}-\bar{X}_{...})^2/\sigma^2$? Furthermore, let $\bar{X}_{.j.}=\sum_{k=1}^{c}\sum_{i=1}^{a}X_{ijk}/ac$ and $\bar{X}_{ij.}=\sum_{k=1}^{c}X_{ijk}/c$. Show that

$$\begin{aligned}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}&(X_{ijk}-\bar{X}_{...})^2\\ &=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(X_{ijk}-\bar{X}_{ij.})^2\\ &\qquad+bc\sum_{i=1}^{a}(\bar{X}_{i..}-\bar{X}_{...})^2+ac\sum_{j=1}^{b}(\bar{X}_{.j.}-\bar{X}_{...})^2\\ &\qquad+c\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{X}_{ij.}-\bar{X}_{i..}-\bar{X}_{.j.}+\bar{X}_{...})^2.\end{aligned}$$

Show that the four terms in the right-hand member, when divided by σ^2, are independent chi-square variables with $ab(c-1)$, $a-1$, $b-1$, and $(a-1)(b-1)$ degrees of freedom, respectively.

10.5. Let X_1, X_2, X_3, X_4 be a random sample of size $n=4$ from the normal distribution $N(0, 1)$. Show that $\sum_{i=1}^{4}(X_i-\bar{X})^2$ equals

$$\frac{(X_1-X_2)^2}{2}+\frac{[X_3-(X_1+X_2)/2]^2}{3/2}+\frac{[X_4-(X_1+X_2+X_3)/3]^2}{4/3}$$

and argue that these three terms are independent, each with a chi-square distribution with 1 degree of freedom.

10.2 A Test of the Equality of Several Means

Consider b independent random variables that have normal distributions with unknown means $\mu_1, \mu_2, \ldots, \mu_b$, respectively, and unknown but common variance σ^2. Let $X_{1j}, X_{2j}, \ldots, X_{aj}$ represent a random sample of size a from the normal distribution with mean μ_j and variance σ^2, $j=1, 2, \ldots, b$. It is desired to test the composite hypothesis $H_0: \mu_1=\mu_2=\cdots=\mu_b=\mu$, μ unspecified, against all possible alternative hypotheses H_1. A likelihood ratio test will be used. Here the total parameter space is

$$\Omega=\{(\mu_1, \mu_2, \ldots, \mu_b, \sigma^2): \quad -\infty<\mu_j<\infty, \quad 0<\sigma^2<\infty\}$$

and

$$\omega = \{(\mu_1, \mu_2, \ldots, \mu_b, \sigma^2): \quad -\infty < \mu_1 = \mu_2 = \cdots$$
$$= \mu_b = \mu < \infty, \quad 0 < \sigma^2 < \infty\}.$$

The likelihood functions, denoted by $L(\omega)$ and $L(\Omega)$ are, respectively,

$$L(\omega) = \left(\frac{1}{2\pi\sigma^2}\right)^{ab/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij} - \mu)^2\right]$$

and

$$L(\Omega) = \left(\frac{1}{2\pi\sigma^2}\right)^{ab/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij} - \mu_j)^2\right].$$

Now

$$\frac{\partial \ln L(\omega)}{\partial \mu} = \frac{\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij} - \mu)}{\sigma^2}$$

and

$$\frac{\partial \ln L(\omega)}{\partial(\sigma^2)} = -\frac{ab}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij} - \mu)^2.$$

If we equate these partial derivatives to zero, the solutions for μ and σ^2 are, respectively, in ω,

$$\frac{\sum_{j=1}^{b}\sum_{i=1}^{a} x_{ij}}{ab} = \bar{x}_{..},$$

$$\frac{\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij} - \bar{x}_{..})^2}{ab} = v, \tag{1}$$

and these values maximize $L(\omega)$. Furthermore,

$$\frac{\partial \ln L(\Omega)}{\partial \mu_j} = \frac{\sum_{i=1}^{a}(x_{ij} - \mu_j)}{\sigma^2}, \qquad j = 1, 2, \ldots, b,$$

and

$$\frac{\partial \ln L(\Omega)}{\partial(\sigma^2)} = -\frac{ab}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{j=1}^{b}\sum_{i=1}^{a}(x_{ij} - \mu_j)^2.$$

If we equate these partial derivatives to zero, the solutions for $\mu_1, \mu_2, \ldots, \mu_b$, and σ^2 are, respectively, in Ω,

$$\frac{\sum_{i=1}^{a} x_{ij}}{a} = \bar{x}_{.j}, \qquad j = 1, 2, \ldots, b,$$

$$\frac{\sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \bar{x}_{.j})^2}{ab} = w, \tag{2}$$

and these values maximize $L(\Omega)$. These maxima are, respectively,

$$L(\hat{\omega}) = \left[\frac{ab}{2\pi \sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \bar{x}_{..})^2} \right]^{ab/2} \exp \left[-\frac{ab \sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \bar{x}_{..})^2}{2 \sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \bar{x}_{..})^2} \right]$$

$$= \left[\frac{ab}{2\pi \sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \bar{x}_{..})^2} \right]^{ab/2} e^{-ab/2}$$

and

$$L(\hat{\Omega}) = \left[\frac{ab}{2\pi \sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \bar{x}_{.j})^2} \right]^{ab/2} e^{-ab/2}.$$

Finally,

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \left[\frac{\sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \bar{x}_{.j})^2}{\sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \bar{x}_{..})^2} \right]^{ab/2}.$$

In the notation of Section 10.1, the statistics defined by the functions $\bar{x}_{..}$ and v given by Equations (1) of this section are

$$\bar{X}_{..} = \sum_{j=1}^{b} \sum_{i=1}^{a} \frac{X_{ij}}{ab} \qquad \text{and} \qquad S^2 = \sum_{j=1}^{b} \sum_{i=1}^{a} \frac{(X_{ij} - \bar{X}_{..})^2}{ab} = \frac{Q}{ab};$$

while the statistics defined by the functions $\bar{x}_{.1}, \bar{x}_{.2}, \ldots, \bar{x}_{.b}$ and w given by Equations (2) in this section are, respectively, $\bar{X}_{.j} = \sum_{i=1}^{a} X_{ij}/a$,

$j = 1, 2, \ldots, b$, and $Q_3/ab = \sum_{j=1}^{b} \sum_{i=1}^{a} (X_{ij} - \bar{X}_{.j})^2/ab$. Thus, in the notation of Section 10.1, $\lambda^{2/ab}$ defines the statistic Q_3/Q.

We reject the hypothesis H_0 if $\lambda \le \lambda_0$. To find λ_0 so that we have a desired significance level α, we must assume that the hypothesis H_0 is true. If the hypothesis H_0 is true, the random variables X_{ij} constitute a random sample of size $n = ab$ from a distribution that is normal with mean μ and variance σ^2. This being the case, it was shown in Example 2, Section 10.1, that $Q = Q_3 + Q_4$, where $Q_4 = a \sum_{j=1}^{b} (\bar{X}_{.j} - \bar{X}_{..})^2$; that Q_3 and Q_4 are independent; and that Q_3/σ^2 and Q_4/σ^2 have chi-square distributions with $b(a-1)$ and $b-1$ degrees of freedom, respectively. Thus the statistic defined by $\lambda^{2/ab}$ may be written

$$\frac{Q_3}{Q_3 + Q_4} = \frac{1}{1 + Q_4/Q_3}.$$

The significance level of the test of H_0 is

$$\alpha = \Pr\left[\frac{1}{1 + Q_4/Q_3} \le \lambda_0^{2/ab}; H_0\right]$$

$$= \Pr\left[\frac{Q_4/(b-1)}{Q_3/[b(a-1)]} \ge c; H_0\right],$$

where

$$c = \frac{b(a-1)}{b-1}(\lambda_0^{-2/ab} - 1).$$

But

$$F = \frac{Q_4/[\sigma^2(b-1)]}{Q_3/[\sigma^2 b(a-1)]} = \frac{Q_4/(b-1)}{Q_3/[b(a-1)]}$$

has an F-distribution with $b-1$ and $b(a-1)$ degrees of freedom. Hence the test of the composite hypothesis $H_0: \mu_1 = \mu_2 = \cdots = \mu_b = \mu$, μ unspecified, against all possible alternatives may be based on an F-statistic. The constant c is so selected as to yield the desired value of α.

Remark. It should be pointed out that a test of the equality of the b means μ_j, $j = 1, 2, \ldots, b$, does not require that we take a random sample of size a from each of the b normal distributions. That is, the samples may be of different sizes, say $a_1, a_2, \ldots, a_b$. A consideration of this procedure is left to Exercise 10.6.

Suppose now that we wish to compute the power of the test of H_0 against H_1 when H_0 is false, that is, when we do not have $\mu_1 = \mu_2 = \cdots = \mu_b = \mu$. It will be seen in Section 10.3 that when H_1 is true, no longer is Q_4/σ^2 a random variable that is $\chi^2(b-1)$. Thus we cannot use an F-statistic to compute the power of the test when H_1 is true. This problem is discussed in Section 10.3.

An observation should be made in connection with maximizing a likelihood function with respect to certain parameters. Sometimes it is easier to avoid the use of the calculus. For example, $L(\Omega)$ of this section can be maximized with respect to μ_j, for every fixed positive σ^2, by minimizing

$$z = \sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \mu_j)^2$$

with respect to μ_j, $j = 1, 2, \ldots, b$. Now z can be written as

$$z = \sum_{j=1}^{b} \sum_{i=1}^{a} [(x_{ij} - \bar{x}_{.j}) + (\bar{x}_{.j} - \mu_j)]^2$$

$$= \sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij} - \bar{x}_{.j})^2 + a \sum_{j=1}^{b} (\bar{x}_{.j} - \mu_j)^2.$$

Since each term in the right-hand member of the preceding equation is nonnegative, clearly z is a minimum, with respect to μ_j, if we take $\mu_j = \bar{x}_{.j}$, $j = 1, 2, \ldots, b$.

EXERCISES

10.6. Let $X_{1j}, X_{2j}, \ldots, X_{a_j j}$ represent independent random samples of sizes a_j from normal distributions with means μ_j and variances σ^2, $j = 1, 2, \ldots, b$. Show that

$$\sum_{j=1}^{b} \sum_{i=1}^{a_j} (X_{ij} - \bar{X}_{..})^2 = \sum_{j=1}^{b} \sum_{i=1}^{a_j} (X_{ij} - \bar{X}_{.j})^2 + \sum_{j=1}^{b} a_j(\bar{X}_{.j} - \bar{X}_{..})^2,$$

or $Q' = Q_3' + Q_4'$. Here $\bar{X}_{..} = \sum_{j=1}^{b} \sum_{i=1}^{a_j} X_{ij} / \sum_{j=1}^{b} a_j$ and $\bar{X}_{.j} = \sum_{i=1}^{a_j} X_{ij}/a_j$. If $\mu_1 = \mu_2 = \cdots = \mu_b$, show that Q'/σ^2 and Q_3'/σ^2 have chi-square distributions. Prove that Q_3' and Q_4' are independent, and hence Q_4'/σ^2 also has a chi-square distribution. If the likelihood ratio λ is used to test $H_0: \mu_1 = \mu_2 = \cdots = \mu_b = \mu$, μ unspecified and σ^2 unknown, against all possible alternatives, show that $\lambda \le \lambda_0$ is equivalent to the computed $F \ge c$, where

$$F = \frac{\left(\sum_{j=1}^{b} a_j - b\right)Q_4'}{(b-1)Q_3'}.$$

What is the distribution of F when H_0 is true?

10.7. Consider the T-statistic that was derived through a likelihood ratio for testing the equality of the means of two normal distributions having common variance in Example 2 in Section 9.3. Show that T^2 is exactly the F-statistic of Exercise 10.6 with $a_1 = n$, $a_2 = m$, and $b = 2$. Of course, $X_1, \ldots, X_n, \bar{X}$ are replaced with $X_{11}, \ldots, X_{1n}, \bar{X}_{1.}$ and $Y_1, \ldots, Y_m, \bar{Y}$ by $X_{21}, \ldots, X_{2m}, \bar{X}_{2.}$.

10.8. In Exercise 10.6, show that the linear functions $X_{ij} - \bar{X}_{.j}$ and $\bar{X}_{.j} - \bar{X}_{..}$ are uncorrelated.

Hint: Recall the definitions of $\bar{X}_{.j}$ and $\bar{X}_{..}$ and, without loss of generality, we can let $E(X_{ij}) = 0$ for all i, j.

10.9. The following are observations associated with independent random samples from three normal distributions having equal variances and respective means μ_1, μ_2, μ_3.

I	II	III
0.5	2.1	3.0
1.3	3.3	5.1
−1.0	0.0	1.9
1.8	2.3	2.4
	2.5	4.2
		4.1

Compute the F-statistic that is used to test $H_0: \mu_1 = \mu_2 = \mu_3$.

10.10. Using the notation of this section, assume that the means satisfy the condition that $\mu = \mu_1 + (b-1)d = \mu_2 - d = \mu_3 - d = \cdots = \mu_b - d$. That is, the last $b - 1$ means are equal but differ from the first mean μ_1, provided that $d \ne 0$. Let independent random samples of size a be taken from the b normal distributions with common unknown variance σ^2.

(a) Show that the maximum likelihood estimators of μ and d are $\hat{\mu} = \bar{X}_{..}$ and

$$\hat{d} = \frac{\sum_{j=2}^{b} \bar{X}_{.j}/(b-1) - \bar{X}_{.1}}{b}.$$

(b) Using Exercise 10.3, find Q_6 and $Q_7 = c\hat{d}^2$ so that, when $d = 0$, Q_7/σ^2 is $\chi^2(1)$ and

$$\sum_{i=1}^{a} \sum_{j=1}^{b} (X_{ij} - \bar{X}_{..})^2 = Q_3 + Q_6 + Q_7.$$

(c) Argue that the three terms in the right-hand member of part (b), once divided by σ^2, are independent random variables with chi-square distributions, provided that $d = 0$.

(d) The ratio $Q_7/(Q_3 + Q_6)$ times what constant has an F-distribution, provided that $d = 0$? Note that this F is really the square of the two-sample T used to test the equality of the mean of the first distribution and the common mean of the other distributions, in which the last $b - 1$ samples are combined into one.

10.3 Noncentral χ^2 and Noncentral F

Let $X_1, X_2, \ldots, X_n$ denote independent random variables that are $N(\mu_i, \sigma^2)$, $i = 1, 2, \ldots, n$, and let $Y = \sum_1^n X_i^2/\sigma^2$. If each μ_i is zero, we know that Y is $\chi^2(n)$. We shall now investigate the distribution of Y when each μ_i is not zero. The m.g.f. of Y is given by

$$M(t) = E\left[\exp\left(t \sum_{i=1}^{n} \frac{X_i^2}{\sigma^2}\right)\right]$$

$$= \prod_{i=1}^{n} E\left[\exp\left(t \frac{X_i^2}{\sigma^2}\right)\right].$$

Consider

$$E\left[\exp\left(\frac{tX_i^2}{\sigma^2}\right)\right] = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{tx_i^2}{\sigma^2} - \frac{(x_i - \mu_i)^2}{2\sigma^2}\right] dx_i.$$

The integral exists if $t < \frac{1}{2}$. To evaluate the integral, note that

$$\frac{tx_i^2}{\sigma^2} - \frac{(x_i - \mu_i)^2}{2\sigma^2} = -\frac{x_i^2(1 - 2t)}{2\sigma^2} + \frac{2\mu_i x_i}{2\sigma^2} - \frac{\mu_i^2}{2\sigma^2}$$

$$= \frac{t\mu_i^2}{\sigma^2(1 - 2t)} - \frac{1 - 2t}{2\sigma^2}\left(x_i - \frac{\mu_i}{1 - 2t}\right)^2.$$

Accordingly, with $t < \frac{1}{2}$, we have

$$E\left[\exp\left(\frac{tX_i^2}{\sigma^2}\right)\right] = \exp\left[\frac{t\mu_i^2}{\sigma^2(1-2t)}\right] \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \times \exp\left[-\frac{1-2t}{2\sigma^2}\left(x_i - \frac{\mu_i}{1-2t}\right)^2\right] dx_i.$$

If we multiply the integrand by $\sqrt{1-2t}$, $t < \frac{1}{2}$, we have the integral of a normal p.d.f. with mean $\mu_i/(1-2t)$ and variance $\sigma^2/(1-2t)$. Thus

$$E\left[\exp\left(\frac{tX_i^2}{\sigma^2}\right)\right] = \frac{1}{\sqrt{1-2t}} \exp\left[\frac{t\mu_i^2}{\sigma^2(1-2t)}\right],$$

and the m.g.f. of $Y = \sum_1^n X_i^2/\sigma^2$ is given by

$$M(t) = \frac{1}{(1-2t)^{n/2}} \exp\left[\frac{t\sum_1^n \mu_i^2}{\sigma^2(1-2t)}\right], \qquad t < \frac{1}{2}.$$

A random variable that has an m.g.f. of the functional form

$$M(t) = \frac{1}{(1-2t)^{r/2}} e^{t\theta/(1-2t)},$$

where $t < \frac{1}{2}$, $0 < \theta$, and r is a positive integer, is said to have a *noncentral chi-square distribution* with r degrees of freedom and noncentrality parameter θ. If one sets the noncentrality parameter $\theta = 0$, one has $M(t) = (1-2t)^{-r/2}$, which is the m.g.f. of a random variable that is $\chi^2(r)$. Such a random variable can appropriately be called a *central chi-square variable*. We shall use the symbol $\chi^2(r, \theta)$ to denote a noncentral chi-square distribution that has the parameters r and θ; and we shall say that a random variable is $\chi^2(r, \theta)$ when that random variable has this kind of distribution. The symbol $\chi^2(r, 0)$ is equivalent to $\chi^2(r)$. Thus our random variable $Y = \sum_1^n X_i^2/\sigma^2$ of this section is $\chi^2\left(n, \sum_1^n \mu_i^2/\sigma^2\right)$. If each μ_i is equal to zero, then Y is $\chi^2(n, 0)$ or, more simply, Y is $\chi^2(n)$.

The noncentral chi-square variables in which we have interest are certain quadratic forms, in normally distributed variables, divided by a variance σ^2. In our example it is worth noting that the noncentrality parameter of $\sum_1^n X_i^2/\sigma^2$, which is $\sum_1^n \mu_i^2/\sigma^2$, may be computed by replacing each X_i in the quadratic form by its mean μ_i, $i = 1, 2, \ldots, n$. This is no fortuitous circumstance; any quadratic form $Q = Q(X_1, \ldots, X_n)$ in normally distributed variables, which is such that Q/σ^2 is $\chi^2(r, \theta)$, has $\theta = Q(\mu_1, \mu_2, \ldots, \mu_n)/\sigma^2$; and if Q/σ^2 is a chi-square variable (central or noncentral) for certain real values of $\mu_1, \mu_2, \ldots, \mu_n$, it is chi-square (central or noncentral) for *all* real values of these means.

It should be pointed out that Theorem 1, Section 10.1, is valid whether the random variables are central or noncentral chi-square variables.

We next discuss a noncentral F-variable. If U and V are independent and are, respectively, $\chi^2(r_1)$ and $\chi^2(r_2)$, the random variable F has been defined by $F = r_2U/r_1V$. Now suppose, in particular, that U is $\chi^2(r_1, \theta)$, V is $\chi^2(r_2)$, and that U and V are independent. The random variable r_2U/r_1V is called a *noncentral F-variable* with r_1 and r_2 degrees of freedom and with noncentrality parameter θ. Note that the noncentrality parameter of F is precisely the noncentrality parameter of the random variable U, which is $\chi^2(r_1, \theta)$.

Tables of noncentral chi-square and noncentral F are available in the literature. However, like those of noncentral t, they are too bulky to be put in this book.

EXERCISES

10.11. Let Y_i, $i = 1, 2, \ldots, n$, denote independent random variables that are, respectively, $\chi^2(r_i, \theta_i)$, $i = 1, 2, \ldots, n$. Prove that $Z = \sum_1^n Y_i$ is $\chi^2\left(\sum_1^n r_i, \sum_1^n \theta_i\right)$.

10.12. Compute the mean and the variance of a random variable that is $\chi^2(r, \theta)$.

10.13. Compute the mean of a random variable that has a noncentral F-distribution with degrees of freedom r_1 and $r_2 > 2$ and noncentrality parameter θ.

10.14. Show that the square of a noncentral T random variable is a noncentral F random variable.

10.15. Let X_1 and X_2 be two independent random variables. Let X_1 and $Y = X_1 + X_2$ be $\chi^2(r_1, \theta_1)$ and $\chi^2(r, \theta)$, respectively. Here $r_1 < r$ and $\theta_1 \le \theta$. Show that X_2 is $\chi^2(r - r_1, \theta - \theta_1)$.

10.16. In Exercise 10.6, if $\mu_1, \mu_2, \ldots, \mu_b$ are not equal, what are the distributions of Q_3'/σ^2, Q_4'/σ^2, and F?

10.4 Multiple Comparisons

Consider b independent random variables that have normal distributions with unknown means $\mu_1, \mu_2, \ldots, \mu_b$, respectively, and with unknown but common variance σ^2. Let $k_1, k_2, \ldots, k_b$ represent b known real constants that are not all zero. We want to find a confidence interval for $\sum_1^b k_j\mu_j$, a linear function of the means $\mu_1, \mu_2, \ldots, \mu_b$. To do this, we take a random sample $X_{1j}, X_{2j}, \ldots, X_{aj}$ of size a from the distribution $N(\mu_j, \sigma^2)$, $j = 1, 2, \ldots, b$. If we denote $\sum_{i=1}^a X_{ij}/a$ by $\bar{X}_{.j}$, then we know that $\bar{X}_{.j}$ is $N(\mu_j, \sigma^2/a)$, that $\sum_{i=1}^a (X_{ij} - \bar{X}_{.j})^2/\sigma^2$ is $\chi^2(a-1)$, and that the two random variables are independent. Since the independent random samples are taken from the b distributions, the $2b$ random variables $\bar{X}_{.j}$, $\sum_{i=1}^a (X_{ij} - \bar{X}_{.j})^2/\sigma^2$, $j = 1, 2, \ldots, b$, are independent. Moreover, $\bar{X}_{.1}, \bar{X}_{.2}, \ldots, \bar{X}_{.b}$ and

$$\sum_{j=1}^b \sum_{i=1}^a \frac{(X_{ij} - \bar{X}_{.j})^2}{\sigma^2}$$

are independent and the latter is $\chi^2[b(a-1)]$. Let $Z = \sum_1^b k_j \bar{X}_{.j}$. Then Z is normal with mean $\sum_1^b k_j \mu_j$ and variance $\left(\sum_1^b k_j^2\right)\sigma^2/a$, and Z is independent of

$$V = \frac{1}{b(a-1)} \sum_{j=1}^b \sum_{i=1}^a (X_{ij} - \bar{X}_{.j})^2.$$

Hence the random variable

$$T = \frac{\dfrac{\sum_1^b k_j\bar{X}_{.j} - \sum_1^b k_j\mu_j}{\sqrt{\left(\sum_1^b k_i^2\right)\sigma^2/a}}}{\sqrt{V/\sigma^2}} = \frac{\sum_1^b k_j\bar{X}_{.j} - \sum_1^b k_j\mu_j}{\sqrt{\left(\sum_1^b k_j^2\right)V/a}}$$

has a t-distribution with $b(a-1)$ degrees of freedom. A positive number c can be found in Table IV in Appendix B, for certain values of α, $0 < \alpha < 1$, such that $\Pr(-c \le T \le c) = 1 - \alpha$. It follows that the probability is $1 - \alpha$ that

$$\sum_1^b k_j\bar{X}_{.j} - c\sqrt{\left(\sum_1^b k_j^2\right)\frac{V}{a}} \le \sum_1^b k_j\mu_j \le \sum_1^b k_j\bar{X}_{.j} + c\sqrt{\left(\sum_1^b k_j^2\right)\frac{V}{a}}.$$

The experimental values of $\bar{X}_{.j}$, $j = 1, 2, \ldots, b$, and V will provide a $100(1-\alpha)$ percent confidence interval for $\sum_1^b k_j\mu_j$.

It should be observed that the confidence interval for $\sum_1^b k_j\mu_j$ depends upon the particular choice of $k_1, k_2, \ldots, k_b$. It is conceivable that we may be interested in more than one linear function of $\mu_1, \mu_2, \ldots, \mu_b$, such as $\mu_2 - \mu_1$, $\mu_3 - (\mu_1 + \mu_2)/2$, or $\mu_1 + \cdots + \mu_b$. We can, of course, find for each $\sum_1^b k_j\mu_j$ a random interval that has a preassigned probability of including that particular $\sum_1^b k_j\mu_j$. But how can we compute the probability that *simultaneously* these random intervals include their respective linear functions of $\mu_1, \mu_2, \ldots, \mu_b$? The following procedure of multiple comparisons, due to Scheffé, is one solution to this problem.

The random variable

$$\frac{\sum_{j=1}^b (\bar{X}_{.j} - \mu_j)^2}{\sigma^2/a}$$

is $\chi^2(b)$ and, because it is a function of $\bar{X}_{.1}, \ldots, \bar{X}_{.b}$ alone, it is independent of the random variable

$$V = \frac{1}{b(a-1)}\sum_{j=1}^b \sum_{i=1}^a (X_{ij} - \bar{X}_{.j})^2.$$

Hence the random variable

$$F = \frac{a \sum_{j=1}^{b} (\bar{X}_{.j} - \mu_j)^2/b}{V}$$

has an F-distribution with b and $b(a-1)$ degrees of freedom. From Table V in Appendix B, for certain values of α, we can find a constant d such that $\Pr(F \le d) = 1 - \alpha$ or

$$\Pr\left[\sum_{j=1}^{b} (\bar{X}_{.j} - \mu_j)^2 \le bd\frac{V}{a}\right] = 1 - \alpha.$$

Note that $\sum_{j=1}^{b} (\bar{X}_{.j} - \mu_j)^2$ is the square of the distance, in b-dimensional space, from the point $(\mu_1, \mu_2, \ldots, \mu_b)$ to the random point $(\bar{X}_{.1}, \bar{X}_{.2}, \ldots, \bar{X}_{.b})$. Consider a space of dimension b and let $(t_1, t_2, \ldots, t_b)$ denote the coordinates of a point in that space. An equation of a hyperplane that passes through the point $(\mu_1, \mu_2, \ldots, \mu_b)$ is given by

$$k_1(t_1 - \mu_1) + k_2(t_2 - \mu_2) + \cdots + k_b(t_b - \mu_b) = 0, \tag{1}$$

where not all the real numbers k_j, $j = 1, 2, \ldots, b$, are equal to zero. The square of the distance from this hyperplane to the point $(t_1 = \bar{X}_{.1}, t_2 = \bar{X}_{.2}, \ldots, t_b = \bar{X}_{.b})$ is

$$\frac{[k_1(\bar{X}_{.1} - \mu_1) + k_2(\bar{X}_{.2} - \mu_2) + \cdots + k_b(\bar{X}_{.b} - \mu_b)]^2}{k_1^2 + k_2^2 + \cdots + k_b^2}. \tag{2}$$

From the geometry of the situation it follows that $\sum_1^b (\bar{X}_{.j} - \mu_j)^2$ is equal to the maximum of expression (2) with respect to $k_1, k_2, \ldots, k_b$. Thus the inequality $\sum_1^b (\bar{X}_{.j} - \mu_j)^2 \le (bd)(V/a)$ holds if and only if

$$\frac{\left[\sum_{j=1}^{b} k_j(\bar{X}_{.j} - \mu_j)\right]^2}{\sum_{j=1}^{b} k_j^2} \le bd\frac{V}{a}, \tag{3}$$

for every real $k_1, k_2, \ldots, k_b$, not all zero. Accordingly, these two equivalent events have the same probability, $1 - \alpha$. However, inequality (3) may be written in the form

$$\left|\sum_1^b k_j\bar{X}_{.j} - \sum_1^b k_j\mu_j\right| \le \sqrt{bd\left(\sum_1^b k_j^2\right)\frac{V}{a}}.$$

Thus the probability is $1-\alpha$ that simultaneously, for *all* real $k_1, k_2, \ldots, k_b$, not all zero,

$$\sum_1^b k_j\bar{X}_{.j} - \sqrt{bd\left(\sum_1^b k_j^2\right)\frac{V}{a}} \le \sum_1^b k_j\mu_j \le \sum_1^b k_j\bar{X}_{.j} + \sqrt{bd\left(\sum_1^b k_j^2\right)\frac{V}{a}}. \quad (4)$$

Denote by A the event where inequality (4) is true for all real $k_1, \ldots, k_b$, and denote by B the event where that inequality is true for a finite number of b-tuples $(k_1, \ldots, k_b)$. If the event A occurs, certainly the event B occurs. Hence $P(A) \le P(B)$. In the applications, one is often interested only in a finite number of linear functions $\sum_1^b k_j\mu_j$. Once the experimental values are available, we obtain from (4) a confidence interval for each of these linear functions. Since $P(B) \ge P(A) = 1-\alpha$, we have a confidence coefficient of at least $100(1-\alpha)$ percent that the linear functions are in these respective confidence intervals.

Remarks. If the sample sizes, say $a_1, a_2, \ldots, a_b$, are unequal, inequality (4) becomes

$$\sum_1^b k_j\bar{X}_{.j} - \sqrt{bd\sum_1^b \frac{k_j^2}{a_j}V} \le \sum_1^b k_j\mu_j \le \sum_1^b k_j\bar{X}_{.j} + \sqrt{bd\sum_1^b \frac{k_j^2}{a_j}V}, \quad (4')$$

where

$$\bar{X}_{.j} = \frac{\sum_{i=1}^{a_j} X_{ij}}{a_j}, \qquad V = \frac{\sum_{j=1}^{b}\sum_{i=1}^{a_j}(X_{ij}-\bar{X}_{.j})^2}{\sum_1^b (a_j-1)},$$

and d is selected from Table V with b and $\sum_1^b (a_j-1)$ degrees of freedom. Inequality (4′) reduces to inequality (4) when $a_1 = a_2 = \cdots = a_b$.

Moreover, if we restrict our attention to linear functions of the form $\sum_1^b k_j\mu_j$ with $\sum_1^b k_j = 0$ (such linear functions are called *contrasts*), the radical in inequality (4′) is replaced by

$$\sqrt{d(b-1)\sum_1^b \frac{k_j^2}{a_j}V},$$

where d is now found in Table V with $b-1$ and $\sum_1^b (a_j-1)$ degrees of freedom.

In these multiple comparisons, one often finds that the length of a confidence interval is much greater than the length of a $100(1-\alpha)$ percent confidence interval for a particular linear function $\sum_1^b k_j\mu_j$. But this is to be expected because in one case the probability $1-\alpha$ applies to just one event, and in the other it applies to the simultaneous occurrence of many events. One reasonable way to reduce the length of these intervals is to take a larger value of α, say 0.25, instead of 0.05. After all, it is still a very strong statement to say that the probability is 0.75 that *all* these events occur.

EXERCISES

10.17. If $A_1, A_2, \ldots, A_k$ are events, prove, by induction, Boole's inequality $P(A_1 \cup A_2 \cup \cdots \cup A_k) \le \sum_1^k P(A_i)$. Then show that

$$P(A_1^* \cap A_2^* \cap \cdots \cap A_k^*) \ge 1 - \sum_1^k P(A_i).$$

10.18. In the notation of this section, let $(k_{i1}, k_{i2}, \ldots, k_{ib})$, $i = 1, 2, \ldots, m$, represent a finite number of b-tuples. The problem is to find simultaneous confidence intervals for $\sum_{j=1}^b k_{ij}\mu_j$, $i = 1, 2, \ldots, m$, by a method different from that of Scheffé. Define the random variable T_i by

$$\left(\sum_{j=1}^b k_{ij}\bar{X}_{.j} - \sum_{j=1}^b k_{ij}\mu_j\right)\Bigg/\sqrt{\left(\sum_{j=1}^b k_{ij}^2\right)V/a}\,, \qquad i = 1, 2, \ldots, m.$$

(a) Let the event A_i^* be given by $-c_i \le T_i \le c_i$, $i = 1, 2, \ldots, m$. Find the random variables U_i and W_i such that $U_i \le \sum_{j=1}^b k_{ij}\mu_j \le W_i$ is equivalent to A_i^*.

(b) Select c_i such that $P(A_i^*) = 1 - \alpha/m$; that is, $P(A_i) = \alpha/m$. Use the results of Exercise 10.17 to determine a lower bound on the probability that simultaneously the random intervals $(U_1, W_1), \ldots, (U_m, W_m)$ include $\sum_{j=1}^b k_{1j}\mu_j, \ldots, \sum_{j=1}^b k_{mj}\mu_j$, respectively.

(c) Let $a = 3$, $b = 6$, and $\alpha = 0.05$. Consider the linear functions $\mu_1 - \mu_2$, $\mu_2 - \mu_3$, $\mu_3 - \mu_4$, $\mu_4 - (\mu_5 + \mu_6)/2$, and $(\mu_1 + \mu_2 + \cdots + \mu_6)/6$. Here $m = 5$. Show that the lengths of the confidence intervals given by the results of part (b) are shorter than the corresponding ones given by the method of Scheffé, as described in the text. If m becomes sufficiently large, however, this is not the case.

10.5 The Analysis of Variance

The problem considered in Section 10.2 is an example of a method of statistical inference called the *analysis of variance*. This method derives its name from the fact that the quadratic form abS^2, which is a total sum of squares, is resolved into several component parts. In this section other problems in the analysis of variance will be investigated.

Let X_{ij}, $i = 1, 2, \ldots, a$ and $j = 1, 2, \ldots, b$, denote $n = ab$ random variables that are independent and have normal distributions with common variance σ^2. The means of these normal distributions are $\mu_{ij} = \mu + \alpha_i + \beta_j$, where $\sum_1^a \alpha_i = 0$ and $\sum_1^b \beta_j = 0$. For example, take $a = 2$, $b = 3$, $\mu = 5$, $\alpha_1 = 1$, $\alpha_2 = -1$, $\beta_1 = 1$, $\beta_2 = 0$, and $\beta_3 = -1$. Then the $ab =$ six random variables have means

$$\mu_{11} = 7, \qquad \mu_{12} = 6, \qquad \mu_{13} = 5,$$

$$\mu_{21} = 5, \qquad \mu_{22} = 4, \qquad \mu_{23} = 3.$$

Had we taken $\beta_1 = \beta_2 = \beta_3 = 0$, the six random variables would have had means

$$\mu_{11} = 6, \qquad \mu_{12} = 6, \qquad \mu_{13} = 6,$$

$$\mu_{21} = 4, \qquad \mu_{22} = 4, \qquad \mu_{23} = 4.$$

Thus, if we wish to test the composite hypothesis that

$$\begin{aligned}
&\mu_{11} = \mu_{12} = \cdots = \mu_{1b},\\
&\mu_{21} = \mu_{22} = \cdots = \mu_{2b},\\
&\vdots\\
&\mu_{a1} = \mu_{a2} = \cdots = \mu_{ab},
\end{aligned}$$

we could say that we are testing the composite hypothesis that $\beta_1 = \beta_2 = \cdots = \beta_b$ (and hence each $\beta_j = 0$, since their sum is zero). On the other hand, the composite hypothesis

$$\begin{aligned}
&\mu_{11} = \mu_{21} = \cdots = \mu_{a1},\\
&\mu_{12} = \mu_{22} = \cdots = \mu_{a2},\\
&\vdots\\
&\mu_{1b} = \mu_{2b} = \cdots = \mu_{ab},
\end{aligned}$$

is the same as the composite hypothesis that $\alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$.

Remarks. The model just described, and others similar to it, are widely used in statistical applications. Consider a situation in which it is desirable to investigate the effects of two factors that influence an outcome. Thus the variety of a grain and the type of fertilizer used influence the yield; or the teacher and the size of a class may influence the score on a standard test. Let X_{ij} denote the yield from the use of variety i of a grain and type j of fertilizer. A test of the hypothesis that $\beta_1 = \beta_2 = \cdots = \beta_b = 0$ would then be a test of the hypothesis that the mean yield of each variety of grain is the same regardless of the type of fertilizer used.

There is no loss of generality in assuming that $\sum_1^a \alpha_i = \sum_1^b \beta_j = 0$. To see this, let $\mu_{ij} = \mu' + \alpha_i' + \beta_j'$. Write $\bar{\alpha}' = \Sigma\, \alpha_i'/a$ and $\bar{\beta}' = \Sigma\, \beta_j'/b$. We have $\mu_{ij} = (\mu' + \bar{\alpha}' + \bar{\beta}') + (\alpha_i' - \bar{\alpha}') + (\beta_j' - \bar{\beta}') = \mu + \alpha_i + \beta_j$, where $\Sigma\, \alpha_i = \Sigma\, \beta_j = 0$.

To construct a test of the composite hypothesis $H_0: \beta_1 = \beta_2 = \cdots = \beta_b = 0$ against all alternative hypotheses, we could obtain the corresponding likelihood ratio. However, to gain more insight into such a test, let us reconsider the likelihood ratio test of Section 10.2, namely that of the equality of the means of b distributions. There the important quadratic forms are Q, Q_3, and Q_4, which are related through the equation $Q = Q_4 + Q_3$. That is,

$$abS^2 = \sum_{j=1}^{b} \sum_{i=1}^{a} (\bar{X}_{.j} - \bar{X}_{..})^2 + \sum_{i=1}^{a} \sum_{j=1}^{b} (X_{ij} - \bar{X}_{.j})^2;$$

so we see that the total sum of squares, abS^2, is decomposed into a sum of squares, Q_4, *among* column means and a sum of squares, Q_3, *within* columns. The latter sum of squares, divided by $n = ab$, is the m.l.e. of σ^2, provided that the parameters are in Ω; and we denote it by $\widehat{\sigma^2_\Omega}$. Of course, S^2 is the m.l.e. of σ^2 under ω, here denoted by $\widehat{\sigma^2_\omega}$. So the likelihood ratio $\lambda = (\widehat{\sigma^2_\Omega}/\widehat{\sigma^2_\omega})^{ab/2}$ is a monotone function of the statistic

$$F = \frac{Q_4/(b-1)}{Q_3/[b(a-1)]}$$

upon which the test of the equality of means is based.

To help find a test for $H_0: \beta_1 = \beta_2 = \cdots = \beta_b = 0$, where $\mu_{ij} = \mu + \alpha_i + \beta_j$, return to the decomposition of Example 3, Section 10.1, namely $Q = Q_2 + Q_4 + Q_5$. That is,

$$abS^2 = \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{X}_{i.} - \bar{X}_{..})^2 + \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{X}_{.j} - \bar{X}_{..})^2$$
$$+ \sum_{i=1}^{a} \sum_{j=1}^{b} (X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2;$$

thus the total sum of squares, abS^2, is decomposed into that among *rows* (Q_2), that among *columns* (Q_4), and that *remaining* (Q_5). It is interesting to observe that $\widehat{\sigma_\Omega^2} = Q_5/ab$ is the m.l.e. of σ^2 under Ω and

$$\widehat{\sigma_\omega^2} = \frac{(Q_4 + Q_5)}{ab} = \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{(X_{ij} - \bar{X}_{i.})^2}{ab}$$

is that estimator under ω. A useful monotone function of the likelihood ratio $\lambda = (\widehat{\sigma_\Omega^2}/\widehat{\sigma_\omega^2})^{ab/2}$ is

$$F = \frac{Q_4/(b-1)}{Q_5/[(a-1)(b-1)]},$$

which has, under H_0, an F-distribution with $b - 1$ and $(a - 1)(b - 1)$ degrees of freedom. The hypothesis H_0 is rejected if $F \geq c$, where $\alpha = \Pr(F \geq c; H_0)$.

If we are to compute the power function of the test, we need the distribution of F when H_0 is not true. From Section 10.3 we know, when H_1 is true, that Q_4/σ^2 and Q_5/σ^2 are independent (central or noncentral) chi-square variables. We shall compute the noncentrality parameters of Q_4/σ^2 and Q_5/σ^2 when H_1 is true. We have $E(X_{ij}) = \mu + \alpha_i + \beta_j$, $E(\bar{X}_{i.}) = \mu + \alpha_i$, $E(\bar{X}_{.j}) = \mu + \beta_j$ and $E(\bar{X}_{..}) = \mu$. Accordingly, the noncentrality parameter of Q_4/σ^2 is

$$\frac{a \sum_{j=1}^{b} (\mu + \beta_j - \mu)^2}{\sigma^2} = \frac{a \sum_{j=1}^{b} \beta_j^2}{\sigma^2}.$$

and that of Q_5/σ^2 is

$$\frac{\sum_{j=1}^{b} \sum_{i=1}^{a} (\mu + \alpha_i + \beta_j - \mu - \alpha_i - \mu - \beta_j + \mu)^2}{\sigma^2} = 0.$$

Thus, if the hypothesis H_0 is not true, F has a noncentral F-distribution with $b - 1$ and $(a - 1)(b - 1)$ degrees of freedom and noncentrality parameter $a \sum_{j=1}^{b} \beta_j^2/\sigma^2$. The desired probabilities can then be found in tables of the noncentral F-distribution.

A similar argument can be used to construct the F needed to test the equality of row means; that is, this F is essentially the ratio of the sum of squares among rows and Q_5. In particular, this F is defined by

$$F = \frac{Q_2/(a-1)}{Q_5/[(a-1)(b-1)]}$$

and, under $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$, has an F-distribution with $a - 1$ and $(a-1)(b-1)$ degrees of freedom.

The analysis-of-variance problem that has just been discussed is usually referred to as a *two-way classification with one observation per cell.* Each combination of i and j determines a cell; thus there is a total of ab cells in this model. Let us now investigate another two-way classification problem, but in this case we take $c > 1$ independent observations per cell.

Let X_{ijk}, $i = 1, 2, \ldots, a$, $j = 1, 2, \ldots, b$, and $k = 1, 2, \ldots, c$, denote $n = abc$ random variables which are independent and which have normal distributions with common, but unknown, variance σ^2. The mean of each X_{ijk}, $k = 1, 2, \ldots, c$, is $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$, where $\sum_{i=1}^{a} \alpha_i = 0$, $\sum_{j=1}^{b} \beta_j = 0$, $\sum_{i=1}^{a} \gamma_{ij} = 0$, and $\sum_{j=1}^{b} \gamma_{ij} = 0$. For example, take $a = 2$, $b = 3$, $\mu = 5$, $\alpha_1 = 1$, $\alpha_2 = -1$, $\beta_1 = 1$, $\beta_2 = 0$, $\beta_3 = -1$, $\gamma_{11} = 1$, $\gamma_{12} = 1$, $\gamma_{13} = -2$, $\gamma_{21} = -1$, $\gamma_{22} = -1$, and $\gamma_{23} = 2$. Then the means are

$$\mu_{11} = 8, \qquad \mu_{12} = 7, \qquad \mu_{13} = 3,$$
$$\mu_{21} = 4, \qquad \mu_{22} = 3, \qquad \mu_{23} = 5.$$

Note that, if each $\gamma_{ij} = 0$, then

$$\mu_{11} = 7, \qquad \mu_{12} = 6, \qquad \mu_{13} = 5,$$
$$\mu_{21} = 5, \qquad \mu_{22} = 4, \qquad \mu_{23} = 3.$$

That is, if $\gamma_{ij} = 0$, each of the means in the first row is 2 greater than the corresponding mean in the second row. In general, if each $\gamma_{ij} = 0$, the means of row i_1 differ from the corresponding means of row i_2 by a constant. This constant may be different for different choices of i_1 and i_2. A similar statement can be made about the means of columns j_1 and j_2. The parameter γ_{ij} is called the *interaction* associated with cell (i, j). That is, the interaction between the ith level of one classification and the jth level of the other classification is γ_{ij}. One interesting hypothesis to test is that each interaction is equal to zero. This will now be investigated.

From Exercise 10.4 of Section 10.1 we have that

$$\begin{aligned}\sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} (X_{ijk} - \bar{X}_{\ldots})^2 = bc \sum_{i=1}^{a} (\bar{X}_{i..} - \bar{X}_{\ldots})^2 &+ ac \sum_{j=1}^{b} (\bar{X}_{.j.} - \bar{X}_{\ldots})^2 \\ &+ c \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{\ldots})^2 \\ &+ \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} (X_{ijk} - \bar{X}_{ij.})^2;\end{aligned}$$

that is, the total sum of squares is decomposed into that due to *row* differences, that due to *column* differences, that due to *interaction*, and that *within cells*. The test of

$$H_0: \gamma_{ij} = 0, \qquad i = 1, 2, \ldots, a, \quad j = 1, 2, \ldots, b,$$

against all possible alternatives is based upon an F with $(a-1)(b-1)$ and $ab(c-1)$ degrees of freedom,

$$F = \frac{\left[c \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{...})^2\right] \Big/ [(a-1)(b-1)]}{\left[\sum\sum\sum (X_{ijk} - \bar{X}_{ij.})^2\right] \Big/ [ab(c-1)]}.$$

The reader should verify that the noncentrality parameter of this F-distribution is equal to $c \sum_{j=1}^{b} \sum_{i=1}^{a} \gamma_{ij}^2/\sigma^2$. Thus F is central when $H_0: \gamma_{ij} = 0$, $i = 1, 2, \ldots, a$, $j = 1, 2, \ldots, b$, is true.

EXERCISES

10.19. Show that

$$\sum_{j=1}^{b} \sum_{i=1}^{a} (X_{ij} - \bar{X}_{i.})^2 = \sum_{j=1}^{b} \sum_{i=1}^{a} (X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2 + a \sum_{j=1}^{b} (\bar{X}_{.j} - \bar{X}_{..})^2.$$

10.20. If at least one $\gamma_{ij} \neq 0$, show that the F, which is used to test that each interaction is equal to zero, has noncentrality parameter equal to $c \sum_{j=1}^{b} \sum_{i=1}^{a} \gamma_{ij}^2/\sigma^2$.

10.21. Using the background of the two-way classification with one observation per cell, show that the maximum likelihood estimators of α_i, β_j, and μ are $\hat{\alpha}_i = \bar{X}_{i.} - \bar{X}_{..}$, $\hat{\beta}_j = \bar{X}_{.j} - \bar{X}_{..}$, and $\hat{\mu} = \bar{X}_{..}$, respectively. Show that these are unbiased estimators of their respective parameters and compute var $(\hat{\alpha}_i)$, var $(\hat{\beta}_j)$, and var $(\hat{\mu})$.

10.22. Prove, using the assumptions of this section, that the linear functions $X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..}$ and $\bar{X}_{.j} - \bar{X}_{..}$ are uncorrelated.

10.23. Given the following observations associated with a two-way classification with $a = 3$ and $b = 4$, compute the F-statistics used to test the equality of the column means ($\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$) and the equality of the row means ($\alpha_1 = \alpha_2 = \alpha_3 = 0$), respectively.

Row/Column	1	2	3	4
1	3.1	4.2	2.7	4.9
2	2.7	2.9	1.8	3.0
3	4.0	4.6	3.0	3.9

10.24. With the background of the two-way classification with $c > 1$ observations per cell, show that the maximum likelihood estimators of the parameters are $\hat{\alpha}_i = \bar{X}_{i..} - \bar{X}_{...}$, $\hat{\beta}_j = \bar{X}_{.j.} - \bar{X}_{...}$, $\hat{\gamma}_{ij} = \bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{...}$, and $\hat{\mu} = \bar{X}_{...}$. Show that these are unbiased estimators of the respective parameters. Compute the variance of each estimator.

10.25. Given the following observations in a two-way classification with $a = 3$, $b = 4$, and $c = 2$, compute the F-statistics used to test that all interactions are equal to zero ($\gamma_{ij} = 0$), all column means are equal ($\beta_j = 0$), and all row means are equal ($\alpha_i = 0$), respectively.

Row/Column	1	2	3	4
1	3.1	4.2	2.7	4.9
	2.9	4.9	3.2	4.5
2	2.7	2.9	1.8	3.0
	2.9	2.3	2.4	3.7
3	4.0	4.6	3.0	3.9
	4.4	5.0	2.5	4.2

10.6 A Regression Problem

There is often interest in the relation between two variables, for example, a student's scholastic aptitude test score in mathematics and this same student's grade in calculus. Frequently, one of these variables, say x, is known in advance of the other, and hence there is interest in predicting a future random variable Y. Since Y is a random variable, we cannot predict its future observed value $Y = y$ with certainty. Thus let us first concentrate on the problem of estimating the mean of Y, that is, $E(Y)$. Now $E(Y)$ is usually a function of x; for example, in our illustration with the calculus grade, say Y, we would expect $E(Y)$ to increase with increasing mathematics aptitude score x. Sometimes $E(Y) = \mu(x)$ is assumed to be of a given form, such as linear or quadratic or exponential; that is, $\mu(x)$ could be assumed to be equal to $\alpha + \beta x$ or $\alpha + \beta x + \gamma x^2$ or $\alpha e^{\beta x}$. To estimate $E(Y) = \mu(x)$, or equivalently the parameters α, β, and γ, we observe the random variable Y for each of n possibly different values of x, say $x_1, x_2, \dots, x_n$, which are not all equal. Once the n independent experiments have been performed, we have n pairs of known numbers $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. These pairs are then used to estimate the mean $E(Y)$. Problems like this are often classified under *regression* because $E(Y) = \mu(x)$ is frequently called a regression curve.

Remark. A model for the mean like $\alpha + \beta x + \gamma x^2$, is called a *linear model* because it is linear in the parameters, α, β, and γ. Thus $\alpha e^{\beta x}$ is not a linear model because it is not linear in α and β. Note that, in Sections 10.1 to 10.4, all the means were linear in the parameters and hence linear models.

Let us begin with the case in which $E(Y) = \mu(x)$ is a linear function. The n points are $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$; so the first problem is that of fitting a straight line to the set of points (see Figure 10.1). In addition to assuming that the mean of Y is a linear function, we assume that, $Y_1, Y_2, \dots, Y_n$ are independent normal variables with respective means $\alpha + \beta(x_i - \bar{x})$, $i = 1, 2, \dots, n$, and unknown variance σ^2, where $\bar{x} = \Sigma\, x_i/n$. Their joint p.d.f. is therefore the

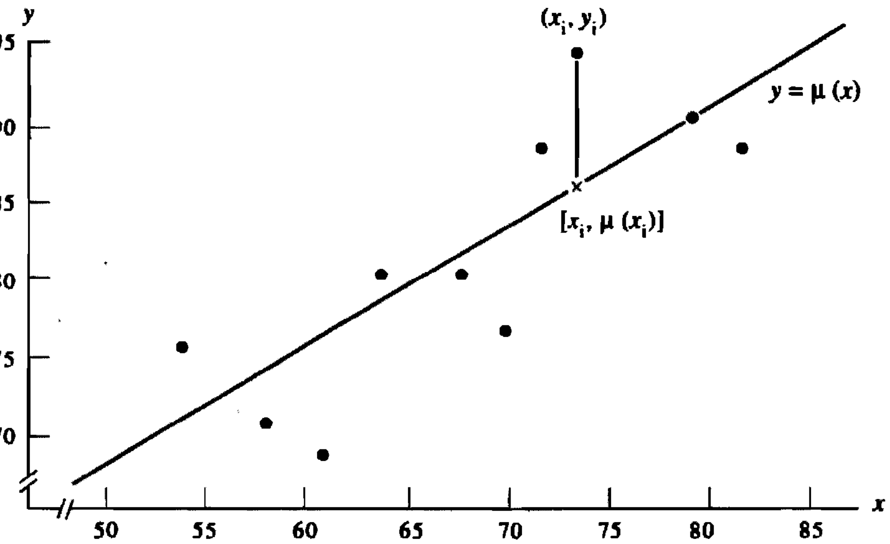

FIGURE 10.1

product of the individual probability density functions; that is, the likelihood function equals

$$L(\alpha, \beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{[y_i - \alpha - \beta(x_i - \bar{x})]^2}{2\sigma^2}\right\}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - \alpha - \beta(x_i - \bar{x})]^2\right\}$$

To maximize $L(\alpha, \beta, \sigma^2)$, or, equivalently, to minimize

$$-\ln L(\alpha, \beta, \sigma^2) = \frac{n}{2}\ln(2\pi\sigma^2) + \frac{\sum_{i=1}^{n}[y_i - \alpha - \beta(x_i - \bar{x})]^2}{2\sigma^2},$$

we must select α and β to minimize

$$H(\alpha, \beta) = \sum_{i=1}^{n}[y_i - \alpha - \beta(x_i - \bar{x})]^2.$$

Since $|y_i - \alpha - \beta(x_i - \bar{x})| = |y_i - \mu(x_i)|$ is the vertical distance from the point (x_i, y_i) to the line $y = \mu(x)$, we note that $H(\alpha, \beta)$ represents the sum of the squares of those distances. Thus selecting α and β so that the sum of the squares is minimized means that we are fitting the straight line to the data by the *method of least squares*.

To minimize $H(\alpha, \beta)$, we find the two first partial derivatives

$$\frac{\partial H(\alpha, \beta)}{\partial \alpha} = 2\sum_{i=1}^{n}[y_i - \alpha - \beta(x_i - \bar{x})](-1)$$

and

$$\frac{\partial H(\alpha, \beta)}{\partial \beta} = 2\sum_{i=1}^{n}[y_i - \alpha - \beta(x_i - \bar{x})][-(x_i - \bar{x})].$$

Setting $\partial H(\alpha, \beta)/\partial \alpha = 0$, we obtain

$$\sum_{i=1}^{n} y_i - n\alpha - \beta\sum_{i=1}^{n}(x_i - \bar{x}) = 0.$$

Since

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0,$$

we have that

$$\sum_{i=1}^{n} y_i - n\alpha = 0$$

and thus

$$\hat{\alpha} = \bar{Y}.$$

The equation $\partial H(\alpha, \beta)/\partial\beta = 0$ yields, with α replaced by $\bar{y}$,

$$\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) - \beta \sum_{i=1}^{n} (x_i - \bar{x})^2 = 0$$

or, equivalently,

$$\hat{\beta} = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} Y_i(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}.$$

To find the maximum likelihood estimator of σ^2, consider the partial derivative

$$\frac{\partial[-\ln L(\alpha, \beta, \sigma^2)]}{\partial(\sigma^2)} = \frac{n}{2\sigma^2} - \frac{\sum_{i=1}^{n} [y_i - \alpha - \beta(x_i - \bar{x})]^2}{2(\sigma^2)^2}.$$

Setting this equal to zero and replacing α and β by their solutions $\hat{\alpha}$ and $\hat{\beta}$, we obtain

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} [Y_i - \hat{\alpha} - \hat{\beta}(x_i - \bar{x})]^2.$$

Of course, due to invariance, $\widehat{\sigma^2} = \hat{\sigma}^2$.

Since $\hat{\alpha}$ is a linear function of independent and normally distributed random variables, $\hat{\alpha}$ has a normal distribution with mean

$$E(\hat{\alpha}) = E\left(\frac{1}{n} \sum_{i=1}^{n} Y_i\right) = \frac{1}{n} \sum_{i=1}^{n} E(Y_i)$$

$$= \frac{1}{n} \sum_{i=1}^{n} [\alpha + \beta(x_i - \bar{x})] = \alpha,$$

and variance

$$\operatorname{var}(\hat{\alpha}) = \sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 \operatorname{var}(Y_i) = \frac{\sigma^2}{n}.$$

The estimator $\hat{\beta}$ is also a linear function of $Y_1, Y_2, \ldots, Y_n$ and hence has a normal distribution with mean

$$E(\hat{\beta}) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})E(Y_i)}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$= \frac{\sum_{i=1}^{n} (x_i - \bar{x})[\alpha + \beta(x_i - \bar{x})]}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$= \frac{\alpha \sum_{i=1}^{n} (x_i - \bar{x}) + \beta \sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \beta$$

and variance

$$\text{var}(\hat{\beta}) = \sum_{i=1}^{n} \left[\frac{x_i - \bar{x}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right]^2 \text{var}(Y_i)$$

$$= \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\left[\sum_{i=1}^{n} (x_i - \bar{x})^2\right]^2} \sigma^2 = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}.$$

It can be shown (Exercise 10.27) that

$$\sum_{i=1}^{n} [Y_i - \alpha - \beta(x_i - \bar{x})]^2 = \sum_{i=1}^{n} \{(\hat{\alpha} - \alpha) + (\hat{\beta} - \beta)(x_i - \bar{x}) + [Y_i - \hat{\alpha} - \hat{\beta}(x_i - \bar{x})]\}^2$$

$$= n(\hat{\alpha} - \alpha)^2 + (\hat{\beta} - \beta)^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 + n\widehat{\sigma^2}.$$

or, for brevity,

$$Q = Q_1 + Q_2 + Q_3.$$

Here Q, Q_1, Q_2, and Q_3 are real quadratic forms in the variables

$$Y_i - \alpha - \beta(x_i - \bar{x}), \qquad i = 1, 2, \ldots, n.$$

In this equation, Q represents the sum of the squares of n independent random variables that have normal distributions with means zero and variances σ^2. Thus Q/σ^2 has a chi-square distribution with n degrees of freedom. Each of the random variables $\sqrt{n}(\hat{\alpha} - \alpha)/\sigma$ and $\sqrt{\sum_{1}^{n} (x_i - \bar{x})^2}(\hat{\beta} - \beta)/\sigma$ has a normal distribution with zero mean and unit variance; thus each of Q_1/σ^2 and Q_2/σ^2 has a chi-square distribution with 1 degree of freedom. Since Q_3 is nonnegative, we have, in accordance with the theorem of Section 10.1, that Q_1, Q_2, and Q_3 are independent, so that Q_3/σ^2 has a chi-square distribution with $n-1-1=n-2$ degrees of freedom. Then each of the random variables

$$T_1 = \frac{[\sqrt{n}(\hat{\alpha}-\alpha)]/\sigma}{\sqrt{Q_3/[\sigma^2(n-2)]}} = \frac{\hat{\alpha}-\alpha}{\sqrt{\hat{\sigma}^2/(n-2)}}$$

and

$$T_2 = \frac{\left[\sqrt{\sum_1^n (x_i-\bar{x})^2}(\hat{\beta}-\beta)\right]\Big/\sigma}{\sqrt{Q_3/[\sigma^2(n-2)]}} = \frac{\hat{\beta}-\beta}{\sqrt{n\hat{\sigma}^2\Big/\left[(n-2)\sum_1^n (x_i-\bar{x})^2\right]}}$$

has a t-distribution with $n-2$ degrees of freedom. These facts enable us to obtain confidence intervals for α and β. The fact that $n\hat{\sigma}^2/\sigma^2$ has a chi-square distribution with $n-2$ degrees of freedom provides a means of determining a confidence interval for σ^2. These are some of the statistical inferences about the parameters to which reference was made in the introductory remarks of this section.

Remark. The more discerning reader should quite properly question our constructions of T_1 and T_2 immediately above. We know that the *squares* of the linear forms are independent of $Q_3 = n\hat{\sigma}^2$, but we do not know, at this time, that the linear forms themselves enjoy this independence. This problem arises again in Section 10.7. In Exercise 10.47, a more general problem is proposed, of which the present case is a special instance.

EXERCISES

10.26. Students' scores on the mathematics portion of the ACT examination, x, and on the final examination in first-semester calculus (200 points possible), y, are given.

(a) Calculate the least squares regression line for these data.
(b) Plot the points and the least squares regression line on the same graph.
(c) Find point estimates for α, β, and σ^2.
(d) Find 95 percent confidence intervals for α and β under the usual assumptions.

x	y	x	y
25	138	20	100
20	84	25	143
26	104	26	141
26	112	28	161
28	88	25	124
28	132	31	118
29	90	30	168
32	183		

10.27. Show that

$$\sum_{i=1}^{n} [Y_i - \alpha - \beta(x_i - \bar{x})]^2 = n(\hat{\alpha} - \alpha)^2 + (\hat{\beta} - \beta)^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 + \sum_{i=1}^{n} [Y_i - \hat{\alpha} - \hat{\beta}(x_i - \bar{x})]^2.$$

10.28. Let the independent random variables $Y_1, Y_2, \ldots, Y_n$ have, respectively, the probability density functions $N(\beta x_i, \gamma^2 x_i^2)$, $i = 1, 2, \ldots, n$, where the given numbers $x_1, x_2, \ldots, x_n$ are not all equal and no one is zero. Find the maximum likelihood estimators of β and γ^2.

10.29. Let the independent random variables $Y_1, \ldots, Y_n$ have the joint p.d.f.

$$L(\alpha, \beta, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{1}^{n} [y_i - \alpha - \beta(x_i - \bar{x})]^2\right\},$$

where the given numbers $x_1, x_2, \ldots, x_n$ are not all equal. Let $H_0 : \beta = 0$ (α and σ^2 unspecified). It is desired to use a likelihood ratio test to test H_0 against all possible alternatives. Find λ and see whether the test can be based on a familiar statistic.

Hint: In the notation of this section show that

$$\sum_{1}^{n} (Y_i - \hat{\alpha})^2 = Q_3 + \hat{\beta}^2 \sum_{1}^{n} (x_i - \bar{x})^2.$$

10.30. Using the notation of Section 10.2, assume that the means μ_j satisfy a linear function of j, namely $\mu_j = c + d[j - (b + 1)/2]$. Let independent random samples of size a be taken from the b normal distributions with common unknown variance σ^2.

(a) Show that the maximum likelihood estimators of c and d are, respectively, $\hat{c} = \bar{X}_{..}$ and

$$\hat{d} = \frac{\sum_{j=1}^{b} [j - (b + 1)/2](\bar{X}_{.j} - \bar{X}_{..})}{\sum_{j=1}^{b} [j - (b + 1)/2]^2}.$$

(b) Show that

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij}-\bar{X}_{..})^2=\sum_{i=1}^{a}\sum_{j=1}^{b}\left[X_{ij}-\bar{X}_{..}-\hat{d}\left(j-\frac{b+1}{2}\right)\right]^2$$
$$+\hat{d}^2\sum_{j=1}^{b}a\left(j-\frac{b+1}{2}\right)^2.$$

(c) Argue that the two terms in the right-hand member of part (b), once divided by σ^2, are independent random variables with chi-square distributions provided that $d = 0$.

(d) What F-statistic would be used to test the equality of the means, that is, $H_0: d = 0$?

10.7 A Test of Independence

Let X and Y have a bivariate normal distribution with means μ_1 and μ_2, positive variances σ_1^2 and σ_2^2, and correlation coefficient ρ. We wish to test the hypothesis that X and Y are independent. Because two jointly normally distributed random variables are independent if and only if $\rho = 0$, we test the hypothesis $H_0: \rho = 0$ against the hypothesis $H_1: \rho \neq 0$. A likelihood ratio test will be used. Let (X_1, Y_1), $(X_2, Y_2), \ldots, (X_n, Y_n)$ denote a random sample of size $n > 2$ from the bivariate normal distribution; that is, the joint p.d.f. of these $2n$ random variables is given by

$$f(x_1, y_1)f(x_2, y_2)\cdots f(x_n, y_n).$$

Although it is fairly difficult to show, the statistic that is defined by the likelihood ratio λ is a function of the statistic

$$R=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}.$$

This statistic R is called the *correlation coefficient* of the random sample. The likelihood ratio principle, which calls for the rejection of H_0 if $\lambda \le \lambda_0$, is equivalent to the computed value of $|R| \ge c$. That is, if the absolute value of the correlation coefficient of the sample is too large, we reject the hypothesis that the correlation coefficient of the distribution is equal to zero. To determine a value of c for a satisfactory significance level, it will be necessary to obtain the distribution of R, or a function of R, when H_0 is true. This will now be done.

Let $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n, n > 2$, where $x_1, x_2, \ldots, x_n$ and $\bar{x} = \sum_1^n x_i/n$ are fixed numbers such that $\sum_1^n (x_i - \bar{x})^2 > 0$. Consider the conditional p.d.f. of $Y_1, Y_2, \ldots, Y_n$, given that $X_1 = x_1$, $X_2 = x_2, \ldots, X_n = x_n$. Because $Y_1, Y_2, \ldots, Y_n$ are independent and, with $\rho = 0$, are also independent of $X_1, X_2, \ldots, X_n$, this conditional p.d.f. is given by

$$\left(\frac{1}{\sqrt{2\pi}\sigma_2}\right)^n \exp\left[-\frac{\sum_1^n (y_i - \mu_2)^2}{2\sigma_2^2}\right].$$

Let R_c be the correlation coefficient, given $X_1 = x_1$, $X_2 = x_2, \ldots, X_n = x_n$, so that

$$\frac{R_c\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

is like $\hat{\beta}$ of Section 10.6 and has mean zero when $\rho = 0$. Thus, referring to T_2 of Section 10.6, we see that

$$\frac{R_c\sqrt{\Sigma(Y_i - \bar{Y})^2}/\sqrt{\Sigma(x_i - \bar{x})^2}}{\sqrt{\dfrac{\sum_{i=1}^n \{Y_i - \bar{Y} - [R_c\sqrt{\Sigma(Y_i - \bar{Y})^2}/\sqrt{\Sigma(x_i - \bar{x})^2}](x_i - \bar{x})\}^2}{(n-2)\Sigma(x_i - \bar{x})^2}}} = \frac{R_c\sqrt{n-2}}{\sqrt{1 - R_c^2}} \tag{1}$$

has, given $X_1 = x_1, \ldots, X_n = x_n$, a conditional t-distribution with $n - 2$ degrees of freedom. Note that the p.d.f., say $g(t)$, of this t-distribution does not depend upon $x_1, x_2, \ldots, x_n$. Now the joint p.d.f. of $X_1, X_2, \ldots, X_n$ and $R\sqrt{n-2}/\sqrt{1-R^2}$, where

$$R = \frac{\sum_1^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_1^n (X_i - \bar{X})^2 \sum_1^n (Y_i - \bar{Y})^2}},$$

is the product of $g(t)$ and the joint p.d.f. of $X_1, X_2, \ldots, X_n$. Integration on $x_1, x_2, \ldots, x_n$ yields the marginal p.d.f. of

$R\sqrt{n-2}/\sqrt{1-R^2}$; because $g(t)$ does not depend upon $x_1, x_2, \ldots, x_n$ it is obvious that this marginal p.d.f. is $g(t)$, the conditional p.d.f. of $R_c\sqrt{n-2}/\sqrt{1-R_c^2}$. The change-of-variable technique can now be used to find the p.d.f. of R.

Remarks. Since R has, when $\rho = 0$, a conditional distribution that does not depend upon $x_1, x_2, \ldots, x_n$ (and hence that conditional distribution is, in fact, the marginal distribution of R), we have the remarkable fact that R is independent of $X_1, X_2, \ldots, X_n$. It follows that R is independent of *every function* of $X_1, X_2, \ldots, X_n$ alone, that is, a function that does not depend upon any Y_i. In like manner, R is independent of every function of $Y_1, Y_2, \ldots, Y_n$ alone. Moreover, a careful review of the argument reveals that nowhere did we use the fact that X has a normal marginal distribution. Thus, if X and Y are independent, and if Y has a normal distribution, then R has the same conditional distribution whatever be the distribution of X, subject to the condition $\sum_1^n (x_i - \bar{x})^2 > 0$. Moreover, if $\Pr\left[\sum_1^n (X_i - \bar{X})^2 > 0\right] = 1$, then R has the same marginal distribution whatever be the distribution of X.

If we write $T = R\sqrt{n-2}/\sqrt{1-R^2}$, where T has a t-distribution with $n - 2 > 0$ degrees of freedom, it is easy to show, by the change-of-variable technique (Exercise 10.34), that the p.d.f. of R is given by

$$\begin{aligned} g(r) &= \frac{\Gamma[(n-1)/2]}{\Gamma(\frac{1}{2})\Gamma[(n-2)/2]}(1-r^2)^{(n-4)/2}, \qquad -1 < r < 1, \\ &= 0 \qquad \text{elsewhere.} \end{aligned} \tag{2}$$

We have now solved the problem of the distribution of R, when $\rho = 0$ and $n > 2$, or, perhaps more conveniently, that of $R\sqrt{n-2}/\sqrt{1-R^2}$. The likelihood ratio test of the hypothesis $H_0 : \rho = 0$ against all alternatives $H_1 : \rho \neq 0$ may be based either on the statistic R or on the statistic $R\sqrt{n-2}/\sqrt{1-R^2} = T$, although the latter is easier to use. In either case the significance level of the test is

$$\alpha = \Pr(|R| \geq c_1; H_0) = \Pr(|T| \geq c_2; H_0),$$

where the constants c_1 and c_2 are chosen so as to give the desired value of α.

Remark. It is also possible to obtain an approximate test of size α by using the fact that

$$W = \frac{1}{2}\ln\left(\frac{1+R}{1-R}\right)$$

has an approximate normal distribution with mean $\frac{1}{2}\ln[(1+\rho)/(1-\rho)]$ and variance $1/(n-3)$. We accept this statement without proof. Thus a test of $H_0: \rho = 0$ can be based on the statistic

$$Z = \frac{\frac{1}{2}\ln[(1+R)/(1-R)] - \frac{1}{2}\ln[(1+\rho)/(1-\rho)]}{\sqrt{1/(n-3)}},$$

with $\rho = 0$ so that $\frac{1}{2}\ln[(1+\rho)/(1-\rho)] = 0$. However, using W, we can also test hypotheses like $H_0: \rho = \rho_0$ against $H_1: \rho \neq \rho_0$, where ρ_0 is not necessarily zero. In that case the hypothesized mean of W is

$$\frac{1}{2}\ln\left(\frac{1+\rho_0}{1-\rho_0}\right).$$

EXERCISES

10.31. Show that

$$R = \frac{\sum_1^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_1^n (X_i - \bar{X})^2 \sum_1^n (Y_i - \bar{Y})^2}} = \frac{\sum_1^n X_i Y_i - n\bar{X}\bar{Y}}{\sqrt{\left(\sum_1^n X_i^2 - n\bar{X}^2\right)\left(\sum_1^n Y_i^2 - n\bar{Y}^2\right)}}.$$

10.32. A random sample of size $n = 6$ from a bivariate normal distribution yields a value of the correlation coefficient of 0.89. Would we accept or reject, at the 5 percent signficance level, the hypothesis that $\rho = 0$?

10.33. Verify Equation (1) of this section.

10.34. Verify the p.d.f. (2) of this section.

10.8 The Distributions of Certain Quadratic Forms

Remark. It is essential that the reader have the background of the multivariate normal distribution as given in Section 4.10 to understand Sections 10.8 and 10.9.

Let X_i, $i = 1, 2, \ldots, n$, denote independent random variables which are $N(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, n$, respectively. Then $Q = \sum_1^n (X_i - \mu_i)^2/\sigma_i^2$ is $\chi^2(n)$. Now Q is a quadratic form in the $X_i - \mu_i$ and Q is seen to be, apart from the coefficient $-\frac{1}{2}$, the random variable which is defined by the exponent on the number e in the joint p.d.f. of $X_1, X_2, \ldots, X_n$. We shall now show that this result can be generalized.

Let $X_1, X_2, \ldots, X_n$ have a multivariate normal distribution with p.d.f.

$$\frac{1}{(2\pi)^{n/2}\sqrt{|\mathbf{V}|}} \exp\left[-\frac{(\mathbf{x}-\boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}\right],$$

where, as usual, the covariance matrix $\mathbf{V}$ is positive definite. We shall show that the random variable Q (a quadratic form in the $X_i - \mu_i$), which is defined by $(\mathbf{x}-\boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})$, is $\chi^2(n)$. We have for the m.g.f. $M(t)$ of Q the integral

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{n/2}\sqrt{|\mathbf{V}|}}$$
$$\times \exp\left[t(\mathbf{x}-\boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu}) - \frac{(\mathbf{x}-\boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}\right] dx_1 \cdots dx_n$$
$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{n/2}\sqrt{|\mathbf{V}|}}$$
$$\times \exp\left[-\frac{(\mathbf{x}-\boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})(1-2t)}{2}\right] dx_1 \cdots dx_n.$$

With $\mathbf{V}^{-1}$ positive definite, the integral is seen to exist for all real values of $t < \frac{1}{2}$. Moreover, $(1-2t)\mathbf{V}^{-1}$, $t < \frac{1}{2}$, is a positive definite matrix and, since $|(1-2t)\mathbf{V}^{-1}| = (1-2t)^n|\mathbf{V}^{-1}|$, it follows that

$$\frac{1}{(2\pi)^{n/2}\sqrt{|\mathbf{V}|/(1-2t)^n}} \exp\left[-\frac{(\mathbf{x}-\boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})(1-2t)}{2}\right]$$

can be treated as a multivariate normal p.d.f. If we multiply our integrand by $(1-2t)^{n/2}$, we have this multivariate p.d.f. Thus the m.g.f. of Q is given by

$$M(t) = \frac{1}{(1-2t)^{n/2}}, \qquad t < \tfrac{1}{2},$$

and Q is $\chi^2(n)$, as we wished to show. This fact is the basis of the chi-square tests that were discussed in Chapter 6.

The remarkable fact that the random variable which is defined by $(\mathbf{x}-\boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ is $\chi^2(n)$ stimulates a number of questions about quadratic forms in normally distributed variables. We would like to treat this problem in complete generality, but limitations of space forbid this, and we find it necessary to restrict ourselves to some special cases.

Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a distribution which is $N(0, \sigma^2)$, $\sigma^2 > 0$. Let $\mathbf{X}' = [X_1, X_2, \ldots, X_n]$ and let $\mathbf{A}$ denote an arbitrary $n \times n$ real symmetric matrix. We shall investigate the distribution of the quadratic form $\mathbf{X}'\mathbf{A}\mathbf{X}$. For instance, we know that $\mathbf{X}'\mathbf{I}_n\mathbf{X}/\sigma^2 = \mathbf{X}'\mathbf{X}/\sigma^2 = \sum_1^n X_i^2/\sigma^2$ is $\chi^2(n)$. First we shall find the m.g.f.of $\mathbf{X}'\mathbf{A}\mathbf{X}/\sigma^2$. Then we shall investigate the conditions that must be imposed upon the real symmetric matrix $\mathbf{A}$ if $\mathbf{X}'\mathbf{A}\mathbf{X}/\sigma^2$ is to have a chi-square distribution. This m.g.f. is given by

$$M(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(\frac{t\mathbf{x}'\mathbf{A}\mathbf{x}}{\sigma^2} - \frac{\mathbf{x}'\mathbf{x}}{2\sigma^2}\right) dx_1 \cdots dx_n$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{\mathbf{x}'(\mathbf{I} - 2t\mathbf{A})\mathbf{x}}{2\sigma^2}\right] dx_1 \cdots dx_n,$$

where $\mathbf{I} = \mathbf{I}_n$. The matrix $\mathbf{I} - 2t\mathbf{A}$ is positive definite if we take $|t|$ sufficiently small, say $|t| < h$, $h > 0$. Moreover, we can treat

$$\frac{1}{(2\pi)^{n/2}\sqrt{|(\mathbf{I} - 2t\mathbf{A})^{-1}\sigma^2|}} \exp\left[-\frac{\mathbf{x}'(\mathbf{I} - 2t\mathbf{A})\mathbf{x}}{2\sigma^2}\right]$$

as a multivariate normal p.d.f. Now $|(\mathbf{I} - 2t\mathbf{A})^{-1}\sigma^2|^{1/2} = \sigma^n/|\mathbf{I} - 2t\mathbf{A}|^{1/2}$. If we multiply our integrand by $|\mathbf{I} - 2t\mathbf{A}|^{1/2}$, we have this multivariate p.d.f. Hence the m.g.f. of $\mathbf{X}'\mathbf{A}\mathbf{X}/\sigma^2$ is given by

$$M(t) = |\mathbf{I} - 2t\mathbf{A}|^{-1/2}, \qquad |t| < h. \tag{1}$$

It proves useful to express this m.g.f. in a different form. To do this, let $a_1, a_2, \ldots, a_n$ denote the characteristic numbers of $\mathbf{A}$ and let $\mathbf{L}$ denote an $n \times n$ orthogonal matrix such that $\mathbf{L}'\mathbf{A}\mathbf{L} = \text{diag}\,[a_1, a_2, \ldots, a_n]$. Thus

$$\mathbf{L}'(\mathbf{I} - 2t\mathbf{A})\mathbf{L} = \begin{bmatrix} 1 - 2ta_1 & 0 & \cdots & 0 \\ 0 & 1 - 2ta_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 - 2ta_n \end{bmatrix}.$$

Then

$$\prod_{i=1}^{n} (1 - 2ta_i) = |\mathbf{L}'(\mathbf{I} - 2t\mathbf{A})\mathbf{L}| = |\mathbf{I} - 2t\mathbf{A}|.$$

Accordingly, we can write $M(t)$, as given in Equation (1), in the form

$$M(t) = \left[\prod_{i=1}^{n} (1 - 2ta_i) \right]^{-1/2}, \qquad |t| < h. \tag{2}$$

Let r, $0 < r \le n$, denote the rank of the real symmetric matrix $\mathbf{A}$. Then exactly r of the real numbers $a_1, a_2, \ldots, a_n$, say $a_1, \ldots, a_r$, are not zero and exactly $n - r$ of these numbers, say $a_{r+1}, \ldots, a_n$, are zero. Thus we can write the m.g.f. of $\mathbf{X}'\mathbf{A}\mathbf{X}/\sigma^2$ as

$$M(t) = [(1 - 2ta_1)(1 - 2ta_2) \cdots (1 - 2ta_r)]^{-1/2}.$$

Now that we have found, in suitable form, the m.g.f. of our random variable, let us turn to the question of the conditions that must be imposed if $\mathbf{X}'\mathbf{A}\mathbf{X}/\sigma^2$ is to have a chi-square distribution. Assume that $\mathbf{X}'\mathbf{A}\mathbf{X}/\sigma^2$ is $\chi^2(k)$. Then

$$M(t) = [(1 - 2ta_1)(1 - 2ta_2) \cdots (1 - 2ta_r)]^{-1/2} = (1 - 2t)^{-k/2},$$

or, equivalently,

$$(1 - 2ta_1)(1 - 2ta_2) \cdots (1 - 2ta_r) = (1 - 2t)^k, \qquad |t| < h.$$

Because the positive integers r and k are the degrees of these polynomials, and because these polynomials are equal for infinitely many values of t, we have $k = r$, the rank of $\mathbf{A}$. Moreover, the uniqueness of the factorization of a polynomial implies that $a_1 = a_2 = \cdots = a_r = 1$. If each of the nonzero characteristic numbers of a real symmetric matrix is one, the matrix is idempotent, that is, $\mathbf{A}^2 = \mathbf{A}$, and conversely (see Exercise 10.38). Accordingly, if $\mathbf{X}'\mathbf{A}\mathbf{X}/\sigma^2$ has a chi-square distribution, then $\mathbf{A}^2 = \mathbf{A}$ and the random variable is $\chi^2(r)$, where r is the rank of $\mathbf{A}$. Conversely, if $\mathbf{A}$ is of rank r, $0 < r \le n$, and if $\mathbf{A}^2 = \mathbf{A}$, then $\mathbf{A}$ has exactly r characteristic numbers that are equal to one, and the remaining $n - r$ characteristic numbers are equal to zero. Thus the m.g.f. of $\mathbf{X}'\mathbf{A}\mathbf{X}/\sigma^2$ is given by $(1 - 2t)^{-r/2}$, $t < \frac{1}{2}$, and $\mathbf{X}'\mathbf{A}\mathbf{X}/\sigma^2$ is $\chi^2(r)$. This establishes the following theorem.

Theorem 2. *Let Q denote a random variable which is a quadratic form in the observations of a random sample of size n from a distribution which is $N(0, \sigma^2)$. Let $\mathbf{A}$ denote the symmetric matrix of Q and let r, $0 < r \le n$, denote the rank of $\mathbf{A}$. Then Q/σ^2 is $\chi^2(r)$ if and only if $\mathbf{A}^2 = \mathbf{A}$.*

Remark. If the normal distribution in Theorem 2 is $N(\mu, \sigma^2)$, the condition $\mathbf{A}^2 = \mathbf{A}$ remains a necessary and sufficient condition that Q/σ^2 have a chi-square distribution. In general, however, Q/σ^2 is not $\chi^2(r)$ but, instead, Q/σ^2 has a noncentral chi-square distribution if $\mathbf{A}^2 = \mathbf{A}$. The number of degrees of freedom is r, the rank of $\mathbf{A}$, and the noncentrality parameter is $\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}/\sigma^2$, where $\boldsymbol{\mu}' = [\mu, \mu, \ldots, \mu]$. Since $\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} = \mu^2 \sum_{i,j} a_{ij}$, where $\mathbf{A} = [a_{ij}]$, then, if $\mu \neq 0$, the conditions $\mathbf{A}^2 = \mathbf{A}$ and $\sum_{i,j} a_{ij} = 0$ are necessary and sufficient conditions that Q/σ^2 be central $\chi^2(r)$. Moreover, the theorem may be extended to a quadratic form in random variables which have a multivariate normal distribution with positive definite covariance matrix $\mathbf{V}$; here the necessary and sufficient condition that Q have a chi-square distribution is $\mathbf{AVA} = \mathbf{A}$.

EXERCISES

10.35. Let $Q = X_1X_2 - X_3X_4$, where X_1, X_2, X_3, X_4 is a random sample of size 4 from a distribution which is $N(0, \sigma^2)$. Show that Q/σ^2 does not have a chi-square distribution. Find the m.g.f. of Q/σ^2.

10.36. Let $\mathbf{X}' = [X_1, X_2]$ be bivariate normal with matrix of means $\boldsymbol{\mu}' = [\mu_1, \mu_2]$ and positive definite covariance matrix $\mathbf{V}$. Let

$$Q_1 = \frac{X_1^2}{\sigma_1^2(1-\rho^2)} - 2\rho\frac{X_1X_2}{\sigma_1\sigma_2(1-\rho^2)} + \frac{X_2^2}{\sigma_2^2(1-\rho^2)}.$$

Show that Q_1 is $\chi^2(r, \theta)$ and find r and θ. When and only when does Q_1 have a central chi-square distribution?

10.37. Let $\mathbf{X}' = [X_1, X_2, X_3]$ denote a random sample of size 3 from a distribution that is $N(4, 8)$ and let

$$\mathbf{A} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$

Justify the assertion that $\mathbf{X}'\mathbf{A}\mathbf{X}/\sigma^2$ is $\chi^2(2, 6)$.

10.38. Let $\mathbf{A}$ be a real symmetric matrix. Prove that each of the nonzero characteristic numbers of $\mathbf{A}$ is equal to 1 if and only if $\mathbf{A}^2 = \mathbf{A}$.

Hint: Let $\mathbf{L}$ be an orthogonal matrix such that $\mathbf{L}'\mathbf{A}\mathbf{L} = \text{diag}\,[a_1, a_2, \ldots, a_n]$ and note that $\mathbf{A}$ is idempotent if and only if $\mathbf{L}'\mathbf{A}\mathbf{L}$ is idempotent.

10.39. The sum of the elements on the principal diagonal of a square matrix $\mathbf{A}$ is called the trace of $\mathbf{A}$ and is denoted by tr $\mathbf{A}$.

(a) If $\mathbf{B}$ is $n \times m$ and $\mathbf{C}$ is $m \times n$, prove that tr $(\mathbf{BC})$ = tr $(\mathbf{CB})$.

(b) If $\mathbf{A}$ is a square matrix and if $\mathbf{L}$ is an orthogonal matrix, use the result of part (a) to show that tr $(\mathbf{L}'\mathbf{A}\mathbf{L})$ = tr $\mathbf{A}$.

(c) If $\mathbf{A}$ is a real symmetric idempotent matrix, use the result of part (b) to prove that the rank of $\mathbf{A}$ is equal to tr $\mathbf{A}$.

10.40. Let $\mathbf{A} = [a_{ij}]$ be a real symmetric matrix. Prove that $\sum_j \sum_i a_{ij}^2$ is equal to the sum of the squares of the characteristic numbers of $\mathbf{A}$.

Hint: If $\mathbf{L}$ is an orthogonal matrix, show that $\sum_j \sum_i a_{ij}^2 = \operatorname{tr}(\mathbf{A}^2) = \operatorname{tr}(\mathbf{L}'\mathbf{A}^2\mathbf{L}) = \operatorname{tr}[(\mathbf{L}'\mathbf{A}\mathbf{L})(\mathbf{L}'\mathbf{A}\mathbf{L})]$.

10.41. Let $\bar{X}$ and S^2 denote, respectively, the mean and the variance of a random sample of size n from a distribution which is $N(0, \sigma^2)$.

(a) If $\mathbf{A}$ denotes the symmetric matrix of $n\bar{X}^2$, show that $\mathbf{A} = (1/n)\mathbf{P}$, where $\mathbf{P}$ is the $n \times n$ matrix, each of whose elements is equal to one.

(b) Demonstrate that $\mathbf{A}$ is idempotent and that the tr $\mathbf{A} = 1$. Thus $n\bar{X}^2/\sigma^2$ is $\chi^2(1)$.

(c) Show that the symmetric matrix $\mathbf{B}$ of nS^2 is $\mathbf{I} - (1/n)\mathbf{P}$.

(d) Demonstrate that $\mathbf{B}$ is idempotent and that tr $\mathbf{B} = n - 1$. Thus nS^2/σ^2 is $\chi^2(n-1)$, as previously proved otherwise.

(e) Show that the product matrix $\mathbf{AB}$ is the zero matrix.

10.9 The Independence of Certain Quadratic Forms

We have previously investigated the independence of linear functions of normally distributed variables (see Exercise 4.132). In this section we shall prove some theorems about the independence of quadratic forms. As we remarked on p. 483, we shall confine our attention to normally distributed variables that constitute a random sample of size n from a distribution that is $N(0, \sigma^2)$.

Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a distribution which is $N(0, \sigma^2)$. Let $\mathbf{A}$ and $\mathbf{B}$ denote two real symmetric matrices, each of order n. Let $\mathbf{X}' = [X_1, X_2, \ldots, X_n]$ and consider the two quadratic forms $\mathbf{X}'\mathbf{A}\mathbf{X}$ and $\mathbf{X}'\mathbf{B}\mathbf{X}$. We wish to show that these quadratic forms are independent if and only if $\mathbf{AB} = \mathbf{0}$, the zero matrix. We shall first compute the m.g.f. $M(t_1, t_2)$ of $\mathbf{X}'\mathbf{A}\mathbf{X}/\sigma^2$ and $\mathbf{X}'\mathbf{B}\mathbf{X}/\sigma^2$. We have

$$M(t_1, t_2) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(\frac{t_1\mathbf{x}'\mathbf{A}\mathbf{x}}{\sigma^2} + \frac{t_2\mathbf{x}'\mathbf{B}\mathbf{x}}{\sigma^2} - \frac{\mathbf{x}'\mathbf{x}}{2\sigma^2}\right) dx_1 \cdots dx_n$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(-\frac{\mathbf{x}'(\mathbf{I} - 2t_1\mathbf{A} - 2t_2\mathbf{B})\mathbf{x}}{2\sigma^2}\right) dx_1 \cdots dx_n.$$

The matrix $\mathbf{I} - 2t_1\mathbf{A} - 2t_2\mathbf{B}$ is positive definite if we take $|t_1|$ and $|t_2|$ sufficiently small, say $|t_1| < h_1$, $|t_2| < h_2$, where $h_1, h_2 > 0$. Then, as on p. 483, we have

$$M(t_1, t_2) = |\mathbf{I} - 2t_1\mathbf{A} - 2t_2\mathbf{B}|^{-1/2}, \qquad |t_1| < h_1, \quad |t_2| < h_2.$$

Let us assume that $\mathbf{X}'\mathbf{A}\mathbf{X}/\sigma^2$ and $\mathbf{X}'\mathbf{B}\mathbf{X}/\sigma^2$ are independent (so that likewise are $\mathbf{X}'\mathbf{A}\mathbf{X}$ and $\mathbf{X}'\mathbf{B}\mathbf{X}$) and prove that $\mathbf{AB} = \mathbf{0}$. Thus we assume that

$$M(t_1, t_2) = M(t_1, 0)M(0, t_2) \tag{1}$$

for all t_1 and t_2 for which $|t_i| < h_i$, $i = 1, 2$. Identity (1) is equivalent to the identity

$$|\mathbf{I} - 2t_1\mathbf{A} - 2t_2\mathbf{B}| = |\mathbf{I} - 2t_1\mathbf{A}||\mathbf{I} - 2t_2\mathbf{B}|, \qquad |t_i| < h_i, \quad i = 1, 2. \tag{2}$$

Let $r > 0$ denote the rank of $\mathbf{A}$ and let $a_1, a_2, \ldots, a_r$ denote the r nonzero characteristic numbers of $\mathbf{A}$. There exists an orthogonal matrix $\mathbf{L}$ such that

$$\mathbf{L}'\mathbf{A}\mathbf{L} = \left[\begin{array}{cccc|c} a_1 & 0 & \cdots & 0 & \\ 0 & a_2 & \cdots & 0 & \mathbf{0} \\ \vdots & \vdots & & \vdots & \\ 0 & 0 & \cdots & a_r & \\ \hline & \mathbf{0} & & & \mathbf{0} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{C}_{11} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] = \mathbf{C}$$

for a suitable ordering of $a_1, a_2, \ldots, a_r$. Then $\mathbf{L}'\mathbf{B}\mathbf{L}$ may be written in the identically partitioned form

$$\mathbf{L}'\mathbf{B}\mathbf{L} = \left[\begin{array}{c|c} \mathbf{D}_{11} & \mathbf{D}_{12} \\ \hline \mathbf{D}_{21} & \mathbf{D}_{22} \end{array}\right] = \mathbf{D}.$$

The identity (2) may be written as

$$|\mathbf{L}'||\mathbf{I} - 2t_1\mathbf{A} - 2t_2\mathbf{B}||\mathbf{L}| = |\mathbf{L}'||\mathbf{I} - 2t_1\mathbf{A}||\mathbf{L}||\mathbf{L}'||\mathbf{I} - 2t_2\mathbf{B}||\mathbf{L}|, \tag{2'}$$

or as

$$|\mathbf{I} - 2t_1\mathbf{C} - 2t_2\mathbf{D}| = |\mathbf{I} - 2t_1\mathbf{C}||\mathbf{I} - 2t_2\mathbf{D}|. \tag{3}$$

The coefficient of $(-2t_1)^r$ in the right-hand member of Equation (3) is seen by inspection to be $a_1a_2 \cdots a_r|\mathbf{I} - 2t_2\mathbf{D}|$. It is not so easy to find the coefficient of $(-2t_1)^r$ in the left-hand member of Equation (3).

Conceive of expanding this determinant in terms of minors of order r formed from the first r columns. One term in this expansion is the product of the minor of order r in the upper left-hand corner, namely, $|\mathbf{I}_r - 2t_1\mathbf{C}_{11} - 2t_2\mathbf{D}_{11}|$, and the minor of order $n-r$ in the lower right-hand corner, namely, $|\mathbf{I}_{n-r} - 2t_2\mathbf{D}_{22}|$. Moreover, this product is the only term in the expansion of the determinant that involves $(-2t_1)^r$. Thus the coefficient of $(-2t_1)^r$ in the left-hand member of Equation (3) is $a_1 a_2 \cdots a_r|\mathbf{I}_{n-r} - 2t_2\mathbf{D}_{22}|$. If we equate these coefficients of $(-2t_1)^r$, we have, for all t_2, $|t_2| < h_2$,

$$|\mathbf{I} - 2t_2\mathbf{D}| = |\mathbf{I}_{n-r} - 2t_2\mathbf{D}_{22}|. \tag{4}$$

Equation (4) implies that the nonzero characteristic numbers of the matrices $\mathbf{D}$ and $\mathbf{D}_{22}$ are the same (see Exercise 10.49). Recall that the sum of the squares of the characteristic numbers of a symmetric matrix is equal to the sum of the squares of the elements of that matrix (see Exercise 10.40). Thus the sum of the squares of the elements of matrix $\mathbf{D}$ is equal to the sum of the squares of the elements of $\mathbf{D}_{22}$. Since the elements of the matrix $\mathbf{D}$ are real, it follows that each of the elements of $\mathbf{D}_{11}$, $\mathbf{D}_{12}$, and $\mathbf{D}_{21}$ is zero. Accordingly, we can write $\mathbf{D}$ in the form

$$\mathbf{D} = \mathbf{L}'\mathbf{B}\mathbf{L} = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{D}_{22} \end{array}\right].$$

Thus $\mathbf{CD} = \mathbf{L}'\mathbf{A}\mathbf{L}\mathbf{L}'\mathbf{B}\mathbf{L} = \mathbf{0}$ and $\mathbf{L}'\mathbf{A}\mathbf{B}\mathbf{L} = \mathbf{0}$ and $\mathbf{AB} = \mathbf{0}$, as we wished to prove.

To complete the proof of the theorem, we assume that $\mathbf{AB} = \mathbf{0}$. We are to show that $\mathbf{X}'\mathbf{A}\mathbf{X}/\sigma^2$ and $\mathbf{X}'\mathbf{B}\mathbf{X}/\sigma^2$ are independent. We have, for all real values of t_1 and t_2,

$$(\mathbf{I} - 2t_1\mathbf{A})(\mathbf{I} - 2t_2\mathbf{B}) = \mathbf{I} - 2t_1\mathbf{A} - 2t_2\mathbf{B},$$

since $\mathbf{AB} = \mathbf{0}$. Thus

$$|\mathbf{I} - 2t_1\mathbf{A} - 2t_2\mathbf{B}| = |\mathbf{I} - 2t_1\mathbf{A}||\mathbf{I} - 2t_2\mathbf{B}|.$$

Since the m.g.f. of $\mathbf{X}'\mathbf{A}\mathbf{X}/\sigma^2$ and $\mathbf{X}'\mathbf{B}\mathbf{X}/\sigma^2$ is given by

$$M(t_1, t_2) = |\mathbf{I} - 2t_1\mathbf{A} - 2t_2\mathbf{B}|^{-1/2}, \qquad |t_i| < h_i, \quad i = 1, 2,$$

we have

$$M(t_1, t_2) = M(t_1, 0)M(0, t_2),$$

and the proof of the following theorem is complete.

Theorem 3. *Let Q_1 and Q_2 denote random variables which are quadratic forms in the observations of a random sample of size n from a distribution which is $N(0, \sigma^2)$. Let* $\mathbf{A}$ *and* $\mathbf{B}$ *denote, respectively, the real symmetric matrices of Q_1 and Q_2. The random variables Q_1 and Q_2 are independent if and only if* $\mathbf{AB} = \mathbf{0}$.

Remark. Theorem 3 remains valid if the random sample is from a distribution which is $N(\mu, \sigma^2)$, whatever be the real value of μ. Moreover, Theorem 2 may be extended to quadratic forms in random variables that have a joint multivariate normal distribution with a positive definite covariance matrix $\mathbf{V}$. The necessary and sufficient condition for the independence of two such quadratic forms with symmetric matrices $\mathbf{A}$ and $\mathbf{B}$ then becomes $\mathbf{AVB} = \mathbf{0}$. In our Theorem 2, we have $\mathbf{V} = \sigma^2\mathbf{I}$, so that $\mathbf{AVB} = \mathbf{A}\sigma^2\mathbf{IB} = \sigma^2\mathbf{AB} = \mathbf{0}$.

We shall next prove Theorem 1 that was stated in Section 10.1.

Theorem 4. *Let $Q = Q_1 + \cdots + Q_{k-1} + Q_k$, where Q, $Q_1, \ldots, Q_{k-1}, Q_k$ are $k + 1$ random variables that are quadratic forms in the observations of a random sample of size n from a distribution which is $N(0, \sigma^2)$. Let Q/σ^2 be $\chi^2(r)$, let Q_i/σ^2 be $\chi^2(r_i)$, $i = 1, 2, \ldots, k - 1$, and let Q_k be nonnegative. Then the random variables $Q_1, Q_2, \ldots, Q_k$ are independent and, hence, Q_k/σ^2 is $\chi^2(r_k = r - r_1 - \cdots - r_{k-1})$.*

Proof. Take first the case of $k = 2$ and let the real symmetric matrices of Q, Q_1, and Q_2 be denoted, respectively, by $\mathbf{A}$, $\mathbf{A}_1$, $\mathbf{A}_2$. We are given that $Q = Q_1 + Q_2$ or, equivalently, that $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2$. We are also given that Q/σ^2 is $\chi^2(r)$ and that Q_1/σ^2 is $\chi^2(r_1)$. In accordance with Theorem 2, p. 484, we have $\mathbf{A}^2 = \mathbf{A}$ and $\mathbf{A}_1^2 = \mathbf{A}_1$. Since $Q_2 \geq 0$, each of the matrices $\mathbf{A}$, $\mathbf{A}_1$, and $\mathbf{A}_2$ is positive semidefinite. Because $\mathbf{A}^2 = \mathbf{A}$, we can find an orthogonal matrix $\mathbf{L}$ such that

$$\mathbf{L}'\mathbf{AL} = \left[\begin{array}{c|c} \mathbf{I}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right].$$

If then we multiply both members of $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2$ on the left by $\mathbf{L}'$ and on the right by $\mathbf{L}$, we have

$$\left[\begin{array}{c|c} \mathbf{I}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] = \mathbf{L}'\mathbf{A}_1\mathbf{L} + \mathbf{L}'\mathbf{A}_2\mathbf{L}.$$

Now each of $\mathbf{A}_1$ and $\mathbf{A}_2$, and hence each of $\mathbf{L}'\mathbf{A}_1\mathbf{L}$ and $\mathbf{L}'\mathbf{A}_2\mathbf{L}$ is positive semidefinite. Recall that, if a real symmetric matrix is positive semidefinite, each element on the principal diagonal is positive or zero. Moreover, if an element on the principal diagonal is zero, then all elements in that row and all elements in that column are zero. Thus $\mathbf{L}'\mathbf{A}\mathbf{L} = \mathbf{L}'\mathbf{A}_1\mathbf{L} + \mathbf{L}'\mathbf{A}_2\mathbf{L}$ can be written as

$$\left[\begin{array}{c|c} \mathbf{I}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{G}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] + \left[\begin{array}{c|c} \mathbf{H}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right]. \tag{5}$$

Since $\mathbf{A}_1^2 = \mathbf{A}_1$, we have

$$(\mathbf{L}'\mathbf{A}_1\mathbf{L})^2 = \mathbf{L}'\mathbf{A}_1\mathbf{L} = \left[\begin{array}{c|c} \mathbf{G}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right].$$

If we multiply both members of Equation (5) on the left by the matrix $\mathbf{L}'\mathbf{A}_1\mathbf{L}$, we see that

$$\left[\begin{array}{c|c} \mathbf{G}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{G}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] + \left[\begin{array}{c|c} \mathbf{G}_r\mathbf{H}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right],$$

or, equivalently, $\mathbf{L}'\mathbf{A}_1\mathbf{L} = \mathbf{L}'\mathbf{A}_1\mathbf{L} + (\mathbf{L}'\mathbf{A}_1\mathbf{L})(\mathbf{L}'\mathbf{A}_2\mathbf{L})$. Thus $(\mathbf{L}'\mathbf{A}_1\mathbf{L}) \times (\mathbf{L}'\mathbf{A}_2\mathbf{L}) = \mathbf{0}$ and $\mathbf{A}_1\mathbf{A}_2 = \mathbf{0}$. In accordance with Theorem 3, Q_1 and Q_2 are independent. This independence immediately implies that Q_2/σ^2 is $\chi^2(r_2 = r - r_1)$. This completes the proof when $k = 2$. For $k > 2$, the proof may be made by induction. We shall merely indicate how this can be done by using $k = 3$. Take $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2 + \mathbf{A}_3$, where $\mathbf{A}^2 = \mathbf{A}$, $\mathbf{A}_1^2 = \mathbf{A}_1$, $\mathbf{A}_2^2 = \mathbf{A}_2$, and $\mathbf{A}_3$ is positive semidefinite. Write $\mathbf{A} = \mathbf{A}_1 + (\mathbf{A}_2 + \mathbf{A}_3) = \mathbf{A}_1 + \mathbf{B}_1$, say. Now $\mathbf{A}^2 = \mathbf{A}$, $\mathbf{A}_1^2 = \mathbf{A}_1$, and $\mathbf{B}_1$ is positive semidefinite. In accordance with the case of $k = 2$, we have $\mathbf{A}_1\mathbf{B}_1 = \mathbf{0}$, so that $\mathbf{B}_1^2 = \mathbf{B}_1$. With $\mathbf{B}_1 = \mathbf{A}_2 + \mathbf{A}_3$, where $\mathbf{B}_1^2 = \mathbf{B}_1$, $\mathbf{A}_2^2 = \mathbf{A}_2$, it follows from the case of $k = 2$ that $\mathbf{A}_2\mathbf{A}_3 = \mathbf{0}$ and $\mathbf{A}_3^2 = \mathbf{A}_3$. If we regroup by writing $\mathbf{A} = \mathbf{A}_2 + (\mathbf{A}_1 + \mathbf{A}_3)$, we obtain $\mathbf{A}_1\mathbf{A}_3 = \mathbf{0}$, and so on.

Remark. In our statement of Theorem 4 we took $X_1, X_2, \ldots, X_n$ to be observations of a random sample from a distribution which is $N(0, \sigma^2)$. We did this because our proof of Theorem 3 was restricted to that case. In fact, if $Q', Q'_1, \ldots, Q'_k$ are quadratic forms in any normal variables (including multivariate normal variables), if $Q' = Q'_1 + \cdots + Q'_k$, if $Q', Q'_1, \ldots, Q'_{k-1}$ are central or noncentral chi-square, and if Q'_k is nonnegative, then $Q'_1, \ldots, Q'_k$ are independent and Q'_k is either central or noncentral chi-square.

This section will conclude with a proof of a frequently quoted theorem due to Cochran.

Theorem 5. *Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution which is $N(0, \sigma^2)$. Let the sum of the squares of these observations be written in the form*

$$\sum_1^n X_i^2 = Q_1 + Q_2 + \cdots + Q_k,$$

where Q_j is a quadratic form in $X_1, X_2, \ldots, X_n$, with matrix $\mathbf{A}_j$ which has rank r_j, $j = 1, 2, \ldots, k$. The random variables $Q_1, Q_2, \ldots, Q_k$ are independent and Q_j/σ^2 is $\chi^2(r_j)$, $j = 1, 2, \ldots, k$, if and only if $\sum_1^k r_j = n$.

Proof. First assume the two conditions $\sum_1^k r_j = n$ and $\sum_1^n X_i^2 = \sum_1^k Q_j$ to be satisfied. The latter equation implies that $\mathbf{I} = \mathbf{A}_1 + \mathbf{A}_2 + \cdots + \mathbf{A}_k$. Let $\mathbf{B}_i = \mathbf{I} - \mathbf{A}_i$. That is, $\mathbf{B}_i$ is the sum of the matrices $\mathbf{A}_1, \ldots, \mathbf{A}_k$ exclusive of $\mathbf{A}_i$. Let R_i denote the rank of $\mathbf{B}_i$. Since the rank of the sum of several matrices is less than or equal to the sum of the ranks, we have $R_i \le \sum_1^k r_j - r_i = n - r_i$. However, $\mathbf{I} = \mathbf{A}_i + \mathbf{B}_i$, so that $n \le r_i + R_i$ and $n - r_i \le R_i$. Hence $R_i = n - r_i$. The characteristic numbers of $\mathbf{B}_i$ are the roots of the equation $|\mathbf{B}_i - \lambda\mathbf{I}| = 0$. Since $\mathbf{B}_i = \mathbf{I} - \mathbf{A}_i$, this equation can be written as $|\mathbf{I} - \mathbf{A}_i - \lambda\mathbf{I}| = 0$. Thus we have $|\mathbf{A}_i - (1 - \lambda)\mathbf{I}| = 0$. But each root of the last equation is one minus a characteristic number of $\mathbf{A}_i$. Since $\mathbf{B}_i$ has exactly $n - R_i = r_i$ characteristic numbers that are zero, then $\mathbf{A}_i$ has exactly r_i characteristic numbers that are equal to 1. However, r_i is the rank of $\mathbf{A}_i$. Thus each of the r_i nonzero characteristic numbers of $\mathbf{A}_i$ is 1. That is, $\mathbf{A}_i^2 = \mathbf{A}_i$ and thus Q_i/σ^2 is $\chi^2(r_i)$, $i = 1, 2, \ldots, k$. In accordance with Theorem 4, the random variables $Q_1, Q_2, \ldots, Q_k$ are independent.

To complete the proof of Theorem 5, take

$$\sum_1^n X_i^2 = Q_1 + Q_2 + \cdots + Q_k,$$

let $Q_1, Q_2, \ldots, Q_k$ be independent, and let Q_j/σ^2 be $\chi^2(r_j)$, $j = 1, 2, \ldots, k$. Then $\sum_1^k Q_j/\sigma^2$ is $\chi^2\left(\sum_1^k r_j\right)$. But $\sum_1^k Q_j/\sigma^2 = \sum_1^n X_i^2/\sigma^2$ is $\chi^2(n)$. Thus $\sum_1^k r_j = n$ and the proof is complete.

EXERCISES

10.42. Let X_1, X_2, X_3 be a random sample from the normal distribution $N(0, \sigma^2)$. Are the quadratic forms $X_1^2 + 3X_1X_2 + X_2^2 + X_1X_3 + X_3^2$ and $X_1^2 - 2X_1X_2 + \frac{2}{3}X_2^2 - 2X_1X_3 - X_3^2$ independent or dependent?

10.43. Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a distribution which is $N(0, \sigma^2)$. Prove that $\sum_1^n X_i^2$ and every quadratic form, which is nonidentically zero in $X_1, X_2, \ldots, X_n$, are dependent.

10.44. Let X_1, X_2, X_3, X_4 denote a random sample of size 4 from a distribution which is $N(0, \sigma^2)$. Let $Y = \sum_1^4 a_i X_i$, where a_1, a_2, a_3, and a_4 are real constants. If Y^2 and $Q = X_1X_2 - X_3X_4$ are independent, determine a_1, a_2, a_3, and a_4.

10.45. Let **A** be the real symmetric matrix of a quadratic form Q in the observations of a random sample of size n from a distribution which is $N(0, \sigma^2)$. Given that Q and the mean $\bar{X}$ of the sample are independent. What can be said of the elements of each row (column) of **A**?

Hint: Are Q and $\bar{X}^2$ independent?

10.46. Let $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$ be the matrices of $k > 2$ quadratic forms $Q_1, Q_2, \ldots, Q_k$ in the observations of a random sample of size n from a distribution which is $N(0, \sigma^2)$. Prove that the pairwise independence of these forms implies that they are mutually independent.

Hint: Show that $\mathbf{A}_i\mathbf{A}_j = \mathbf{0}$, $i \neq j$, permits $E[\exp(t_1Q_1 + t_2Q_2 + \cdots + t_kQ_k)]$ to be written as a product of the moment-generating functions of $Q_1, Q_2, \ldots, Q_k$.

10.47. Let $\mathbf{X}' = [X_1, X_2, \ldots, X_n]$, where $X_1, X_2, \ldots, X_n$ are observations of a random sample from a distribution which is $N(0, \sigma^2)$. Let $\mathbf{b}' = [b_1, b_2, \ldots, b_n]$ be a real nonzero matrix, and let **A** be a real symmetric matrix of order n. Prove that the linear form $\mathbf{b}'\mathbf{X}$ and the quadratic form $\mathbf{X}'\mathbf{A}\mathbf{X}$ are independent if and only if $\mathbf{b}'\mathbf{A} = \mathbf{0}$. Use this fact to prove that $\mathbf{b}'\mathbf{X}$ and $\mathbf{X}'\mathbf{A}\mathbf{X}$ are independent if and only if the two quadratic forms, $(\mathbf{b}'\mathbf{X})^2 = \mathbf{X}'\mathbf{b}\mathbf{b}'\mathbf{X}$ and $\mathbf{X}'\mathbf{A}\mathbf{X}$, are independent.

10.48. Let Q_1 and Q_2 be two nonnegative quadratic forms in the observations of a random sample from a distribution which is $N(0, \sigma^2)$. Show that another quadratic form Q is independent of $Q_1 + Q_2$ if and only if Q is independent of each of Q_1 and Q_2.

Hint: Consider the orthogonal transformation that diagonalizes the matrix of $Q_1 + Q_2$. After this transformation, what are the forms of the matrices of Q, Q_1, and Q_2 if Q and $Q_1 + Q_2$ are independent?

10.49. Prove that Equation (4) of this section implies that the nonzero characteristic numbers of the matrices **D** and $\mathbf{D}_{22}$ are the same.

Hint: Let $\lambda = 1/(2t_2)$, $t_2 \neq 0$, and show that Equation (4) is equivalent to $|\mathbf{D} - \lambda\mathbf{I}| = (-\lambda)^r|\mathbf{D}_{22} - \lambda\mathbf{I}_{n-r}|$.

10.50. Here Q_1 and Q_2 are quadratic forms in observations of a random sample from $N(0, 1)$. If Q_1 and Q_2 are independent and if $Q_1 + Q_2$ has a chi-square distribution, prove that Q_1 and Q_2 are chi-square variables.

10.51. Often in regression the mean of the random variable Y is a linear function of p-values $x_1, x_2, \ldots, x_p$, say $\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$, where $\boldsymbol{\beta}' = (\beta_1, \beta_2, \ldots, \beta_p)$ are the *regression coefficients*. Suppose that n values, $\mathbf{Y}' = (Y_1, Y_2, \ldots, Y_n)$, are observed for the x-values in $\mathbf{X} = (x_{ij})$, where $\mathbf{X}$ is an $n \times p$ *design matrix* and its ith row is associated with Y_i, $i = 1, 2, \ldots, n$. Assume that $\mathbf{Y}$ is multivariate normal with mean $\mathbf{X}\boldsymbol{\beta}$ and covariance matrix $\sigma^2\mathbf{I}$, where $\mathbf{I}$ is the $n \times n$ identity matrix.

(a) Note that $Y_1, Y_2, \ldots, Y_n$ are independent. Why?

(b) Since $\mathbf{Y}$ should approximately equal its mean $\mathbf{X}\boldsymbol{\beta}$, we estimate $\boldsymbol{\beta}$ by solving the *normal equations* $\mathbf{X}'\mathbf{Y} = \mathbf{X}'\mathbf{X}\boldsymbol{\beta}$ for $\boldsymbol{\beta}$. Assuming that $\mathbf{X}'\mathbf{X}$ is nonsingular, solve the equations to get $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Show that $\hat{\boldsymbol{\beta}}$ has a multivariate normal distribution with mean $\boldsymbol{\beta}$ and covariance matrix $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

(c) Show that

$$(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{X}'\mathbf{X})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}),$$

say $Q = Q_1 + Q_2$ for convenience.

(d) Show that Q_1/σ^2 is $\chi^2(p)$.

(e) Show that Q_1 and Q_2 are independent.

(f) Argue that Q_2/σ^2 is $\chi^2(n - p)$.

(g) Find c so that cQ_1/Q_2 has an F-distribution.

(h) The fact that a value d can be found so that $\Pr(cQ_1/Q_2 \le d) = 1 - \alpha$ could be used to find a $100(1 - \alpha)$ percent confidence ellipsoid for $\boldsymbol{\beta}$. Explain.

(i) If the coefficient matrix $\boldsymbol{\beta}$ has the prior distribution that is multivariate normal with mean matrix $\boldsymbol{\beta}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$, what is the posterior distribution of $\boldsymbol{\beta}$, given $\hat{\boldsymbol{\beta}}$?

10.52. Say that G.P.A. (Y) is thought to be a linear function of a "coded" high school rank (x_2) and a "coded" American College Testing score (x_3), namely, $\beta_1 + \beta_2 x_2 + \beta_3 x_3$. Note that all x_1 values equal 1. We observe the following five points:

x_1	x_2	x_3	Y
1	1	2	3
1	4	3	6
1	2	2	4
1	4	2	4
1	3	2	4

(a) Compute $\mathbf{X}'\mathbf{X}$ and $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

(b) Compute a 95 percent confidence ellipsoid for $\boldsymbol{\beta}' = (\beta_1, \beta_2, \beta_3)$.

ADDITIONAL EXERCISES

10.53. Let μ_1, μ_2, μ_3 be, respectively, the means of three normal distributions with a common but unknown variance σ^2. In order to test, at the $\alpha = 5$ percent significance level, the hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$ against all possible alternative hypotheses, we take an independent random sample of size 4 from each of these distributions. Determine whether we accept or reject H_0 if the observed values from these three distributions are, respectively,

X_1:	5	9	6	8
X_2:	11	13	10	12
X_3:	10	6	9	9

10.54. The driver of a diesel-powered automobile decided to test the quality of three types of diesel fuel sold in the area based on mpg. Test the null hypothesis that the three means are equal using the following data. Make the usual assumptions and take $\alpha = 0.05$.

Brand A:	38.7	39.2	40.1	38.9	
Brand B:	41.9	42.3	41.3		
Brand C:	40.8	41.2	39.5	38.9	40.3

10.55. We wish to compare compressive strengths of concrete corresponding to $a = 3$ different drying methods (treatments). Concrete is mixed in batches that are just large enough to produce three cylinders. Although care is taken to achieve uniformity, we expect some variability among the $b = 5$ batches used to obtain the following compressive strengths. (There is little reason to suspect interaction and hence only one observation is taken in each cell.)

	Batch				
Treatment	B_1	B_2	B_3	B_4	B_5
A_1	52	47	44	51	42
A_2	60	55	49	52	43
A_3	56	48	45	44	38

(a) Use the 5 percent significance level and test $H_A : \alpha_1 = \alpha_2 = \alpha_3 = 0$ against all alternatives.

(b) Use the 5 percent significance level and test $H_B : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ against all alternatives.

10.56. With $a = 3$ and $b = 4$, find μ, α_i, β_j, and γ_{ij}, if μ_{ij}, $i = 1, 2, 3$ and $j = 1, 2, 3, 4$, are given by

$$\begin{array}{cccc} 6 & 7 & 7 & 12 \\ 10 & 3 & 11 & 8 \\ 8 & 5 & 9 & 10 \end{array}$$

10.57. Two experiments gave the following results:

n	$\bar{x}$	$\bar{y}$	s_x	s_y	r
100	10	20	5	8	0.70
200	12	22	6	10	0.80

Calculate r for the combined sample.

10.58. Consider the following matrices: $\mathbf{Y}$ is $n \times 1$, $\boldsymbol{\beta}$ is $p \times 1$, $\mathbf{X}$ is $n \times p$ and of rank p. Let $\mathbf{Y}$ be $N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$. Discuss the joint p.d.f. of $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ and $\mathbf{Y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y}/\sigma^2$.

10.59. Fit $y = a + x$ to the data

x	0	1	2
y	1	3	4

by the method of least squares.

10.60. Fit by the method of least squares the plane $z = a + bx + cy$ to the five points (x, y, z): $(-1, -2, 5)$, $(0, -2, 4)$, $(0, 0, 4)$, $(1, 0, 2)$, $(2, 1, 0)$.

10.61. Let the 4×1 matrix $\mathbf{Y}$ be multivariate normal $N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where the 4×3 design matrix equals

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 2 \\ 1 & -1 & 2 \\ 1 & 0 & -3 \\ 1 & 0 & -1 \end{bmatrix}$$

and $\boldsymbol{\beta}$ is the 3×1 regression coefficient matrix.

(a) Find the mean matrix and the covariance matrix of $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

(b) If we observe $\mathbf{Y}'$ to be equal to $(6, 1, 11, 3)$, compute $\hat{\boldsymbol{\beta}}$.

10.62. Let the independent normal random variables $Y_1, Y_2, \ldots, Y_n$ have, respectively, the probability density functions $N(\mu, \gamma^2 x_i^2)$, $i = 1, 2, \ldots, n$, where the given $x_1, x_2, \ldots, x_n$ are not all equal and no one of which is zero. Discuss the test of the hypothesis $H_0 : \gamma = 1$, μ unspecified, against all alternatives $H_1 : \gamma \neq 1$, μ unspecified.

10.63. Let $Y_1, Y_2, \ldots, Y_n$ be n independent normal variables with common unknown variance σ^2. Let Y_i have mean βx_i, $i = 1, 2, \ldots, n$, where $x_1, x_2, \ldots, x_n$ are known but not all the same and β is an unknown constant. Find the likelihood ratio test for $H_0: \beta = 0$ against all alternatives. Show that this likelihood ratio test can be based on a statistic that has a well-known distribution.

10.64. Consider the multivariate normal p.d.f. $f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ where the known parameters equal either $\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1$ or $\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2$, respectively.

(a) If $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ is known to equal $\boldsymbol{\Sigma}$, classify $\mathbf{X}$ as being in the second of these distributions if

$$\frac{f(\mathbf{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma})}{f(\mathbf{x}; \boldsymbol{\mu}_2, \boldsymbol{\Sigma})} \le k;$$

otherwise, $\mathbf{X}$ is classified as being from the first distribution. Show that this rule is based upon a linear function of $\mathbf{X}$ and determine its distribution. This allows us to compute the probabilities of misclassification.

(b) If $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are different but known, show that

$$\frac{f(\mathbf{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)}{f(\mathbf{x}; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)} \le k$$

can be based upon a second degree polynomial in $\mathbf{X}$. When either $\boldsymbol{\Sigma}_1$ or $\boldsymbol{\Sigma}_2$ is the correct covariance matrix, does this expression have a chi-square distribution?

CHAPTER 11

Nonparametric Methods

11.1 Confidence Intervals for Distribution Quantiles

We shall first define the concept of a quantile of a distribution of a random variable of the continuous type. Let X be a random variable of the continuous type with p.d.f. $f(x)$ and distribution function $F(x)$. Let p denote a positive proper fraction and assume that the equation $F(x) = p$ has a unique solution for x. This unique root is denoted by the symbol ξ_p and is called the *quantile* (of the distribution) *of order* p. Thus $\Pr(X \le \xi_p) = F(\xi_p) = p$. For example, the quantile of order $\frac{1}{2}$ is the median of the distribution and $\Pr(X \le \xi_{0.5}) = F(\xi_{0.5}) = \frac{1}{2}$.

In Chapter 6 we computed the probability that a certain random interval includes a special point. Frequently, this special point was a parameter of the distribution of probability under consideration. Thus we are led to the notion of an interval estimate of a parameter. If the parameter happens to be a quantile of the distribution, and if we work with certain functions of the order statistics, it will be seen that this method of statistical inference is applicable to all distri-