CHAPTER 4

The Distributions and Uses of Sample Correlation Coefficients

4.1. INTRODUCTION

In Chapter 2, in which the multivariate normal distribution was introduced, it was shown that a measure of dependence between two normal variates is the correlation coefficient $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$. In a conditional distribution of $X_1, \dots, X_q$ given $X_{q+1} = x_{q+1}, \dots, X_p = x_p$, the partial correlation $\rho_{ij \cdot q+1, \dots, p}$ measures the dependence between X_i and X_j. The third kind of correlation discussed was the multiple correlation which measures the relationship between one variate and a set of others. In this chapter we treat the sample equivalents of these quantities; they are point estimates of the population quantities. The distributions of the sample correlations are found. Tests of hypotheses and confidence intervals are developed.

In the cases of joint normal distributions these correlation coefficients are the natural measures of dependence. In the population they are the only parameters except for location (means) and scale (standard deviations) parameters. In the sample the correlation coefficients are derived as the reasonable estimates of the population correlations. Since the sample means and standard deviations are location and scale estimates, the sample correlations (that is, the standardized sample second moments) give all possible information about the population correlations. The sample correlations are the functions of the sufficient statistics that are invariant with respect to location and scale transformations; the population correlations are the functions of the parameters that are invariant with respect to these transformations.

An Introduction to Multivariate Statistical Analysis, Third Edition. By T. W. Anderson
ISBN 0-471-36091-0.

In *regression theory* or least squares, one variable is considered random or *dependent*, and the others fixed or *independent*. In correlation theory we consider several variables as random and treat them symmetrically. If we start with a joint normal distribution and hold all variables fixed except one, we obtain the least squares model because the expected value of the random variable in the conditional distribution is a linear function of the variables held fixed. The sample regression coefficients obtained in least squares are functions of the sample variances and correlations.

In testing independence we shall see that we arrive at the same tests in either case (i.e., in the joint normal distribution or in the conditional distribution of least squares). The probability theory under the null hypothesis is the same. The distribution of the test criterion when the null hypothesis is not true differs in the two cases. If all variables may be considered random, one uses correlation theory as given here; if only one variable is random, one uses least squares theory (which is considered in some generality in Chapter 8).

In Section 4.2 we derive the distribution of the sample correlation coefficient, first when the corresponding population correlation coefficient is 0 (the two normal variables being independent) and then for any value of the population coefficient. The Fisher z-transform yields a useful approximate normal distribution. Exact and approximate confidence intervals are developed. In Section 4.3 we carry out the same program for partial correlations, that is, correlations in conditional normal distributions. In Section 4.4 the distributions and other properties of the sample multiple correlation coefficient are studied. In Section 4.5 the asymptotic distributions of these correlations are derived for elliptically contoured distributions. A stochastic representation for a class of such distributions is found.

4.2. CORRELATION COEFFICIENT OF A BIVARIATE SAMPLE

4.2.1. The Distribution When the Population Correlation Coefficient Is Zero; Tests of the Hypothesis of Lack of Correlation

In Section 3.2 it was shown that if one has a sample (of p-component vectors) $x_1, \dots, x_N$ from a normal distribution, the maximum likelihood estimator of the correlation between X_i and X_j (two components of the random vector X) is

$$r_{ij} = \frac{\sum_{\alpha=1}^{N}(x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)}{\sqrt{\sum_{\alpha=1}^{N}(x_{i\alpha} - \bar{x}_i)^2}\sqrt{\sum_{\alpha=1}^{N}(x_{j\alpha} - \bar{x}_j)^2}}, \tag{1}$$

where $x_{i\alpha}$ is the ith component of $\boldsymbol{x}_\alpha$ and

$$\bar{x}_i = \frac{1}{N} \sum_{\alpha=1}^{N} x_{i\alpha}. \tag{2}$$

In this section we shall find the distribution of r_{ij} when the population correlation between X_i and X_j is zero, and we shall see how to use the sample correlation coefficient to test the hypothesis that the population coefficient is zero.

For convenience we shall treat r_{12}; the same theory holds for each r_{ij}. Since r_{12} depends only on the first two coordinates of each $\boldsymbol{x}_\alpha$, to find the distribution of r_{12} we need only consider the joint distribution of (x_{11}, x_{21}), $(x_{12}, x_{22}), \dots, (x_{1N}, x_{2N})$. We can reformulate the problems to be considered here, therefore, in terms of a bivariate normal distribution. Let $\boldsymbol{x}_1^*, \dots, \boldsymbol{x}_N^*$ be observation vectors from

$$N\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho \\ \sigma_2 \sigma_1 \rho & \sigma_2^2 \end{pmatrix}\right]. \tag{3}$$

We shall consider

$$r = \frac{a_{12}}{\sqrt{a_{11}}\sqrt{a_{22}}}, \tag{4}$$

where

$$a_{ij} = \sum_{\alpha=1}^{N} (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j), \qquad i, j = 1, 2, \tag{5}$$

and $\bar{x}_i$ is defined by (2), $x_{i\alpha}$ being the ith component of $\boldsymbol{x}_\alpha^*$.

From Section 3.3 we see that a_{11}, a_{12}, and a_{22} are distributed like

$$a_{ij} = \sum_{\alpha=1}^{n} z_{i\alpha} z_{j\alpha}, \qquad i, j = 1, 2, \tag{6}$$

where $n = N - 1$, $(z_{1\alpha}, z_{2\alpha})$ is distributed according to

$$N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho \\ \sigma_2 \sigma_1 \rho & \sigma_2^2 \end{pmatrix}\right], \tag{7}$$

and the pairs $(z_{11}, z_{21}), \dots, (z_{1N}, z_{2N})$ are independently distributed.

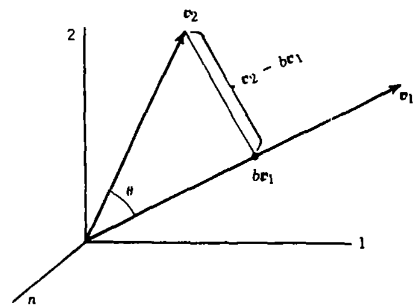

Figure 4.1

Define the n-component vector $v_i = (z_{i1}, \dots, z_{in})'$, $i = 1, 2$. These two vectors can be represented in an n-dimensional space; see Figure 4.1. The correlation coefficient is the cosine of the angle, say θ, between v_1 and v_2. (See Section 3.2.) To find the distribution of $\cos\theta$ we shall first find the distribution of $\cot\theta$. As shown in Section 3.2, if we let $b = v_2'v_1/v_1'v_1$, then $v_2 - bv_1$ is orthogonal to v_1 and

$$\cot\theta = \frac{b\|v_1\|}{\|v_2 - bv_1\|}. \tag{8}$$

If v_1 is fixed, we can rotate coordinate axes so that the first coordinate axis lies along v_1. Then bv_1 has only the first coordinate different from zero, and $v_2 - bv_1$ has this first coordinate equal to zero. We shall show that $\cot\theta$ is proportional to a t-variable when $\rho = 0$.

We use the following lemma.

Lemma 4.2.1. *If $Y_1, \dots, Y_n$ are independently distributed, if $Y_\alpha = (Y_\alpha^{(1)\prime}, Y_\alpha^{(2)\prime})'$ has the density $f(y_\alpha)$, and if the conditional density of $Y_\alpha^{(2)}$ given $Y_\alpha^{(1)} = y_\alpha^{(1)}$ is $f(y_\alpha^{(2)}|y_\alpha^{(1)})$, $\alpha = 1, \dots, n$, then in the conditional distribution of $Y_1^{(2)}, \dots, Y_n^{(2)}$ given $Y_1^{(1)} = y_1^{(1)}, \dots, Y_n^{(1)} = y_n^{(1)}$, the random vectors $Y_1^{(2)}, \dots, Y_n^{(2)}$ are independent and the density of $Y_\alpha^{(2)}$ is $f(y_\alpha^{(2)}|y_\alpha^{(1)})$, $\alpha = 1, \dots, n$.*

Proof. The marginal density of $Y_1^{(1)}, \dots, Y_n^{(1)}$ is $\prod_{\alpha=1}^n f_1(y_\alpha^{(1)})$, where $f_1(y_\alpha^{(1)})$ is the marginal density of $Y_\alpha^{(1)}$, and the conditional density of $Y_1^{(2)}, \dots, Y_n^{(2)}$ given $Y_1^{(1)} = y_1^{(1)}, \dots, Y_n^{(1)} = y_n^{(1)}$ is

$$\frac{\prod_{\alpha=1}^n f(y_\alpha)}{\prod_{\alpha=1}^n f_1(y_\alpha^{(1)})} = \prod_{\alpha=1}^n \frac{f(y_\alpha)}{f_1(y_\alpha^{(1)})} = \prod_{\alpha=1}^n f(y_\alpha^{(2)}|y_\alpha^{(1)}). \tag{9}$$ ∎

Write $V_i = (Z_{i1}, \dots, Z_{in})'$, $i = 1, 2$, to denote random vectors. The conditional distribution of $Z_{2\alpha}$ given $Z_{1\alpha} = z_{1\alpha}$ is $N(\beta z_{1\alpha}, \sigma^2)$, where $\beta = \rho\sigma_2/\sigma_1$ and $\sigma^2 = \sigma_2^2(1 - \rho^2)$. (See Section 2.5.) The density of V_2 given $V_1 = v_1$ is $N(\beta v_1, \sigma^2 I)$ since the $Z_{2\alpha}$ are independent. Let $b = V_2' v_i / v_1' v_1$ $(= a_{21}/a_{11})$, so that $bv_1'(V_2 - bv_1) = 0$, and let $U = (V_2 - bv_1)'(V_2 - bv_1) = V_2'V_2 - b^2 v_1' v_1$ $(= a_{22} - a_{12}^2/a_{11})$. Then $\cot\theta = b\sqrt{a_{11}/U}$. The rotation of coordinate axes involves choosing an $n \times n$ orthogonal matrix C with first row $(1/c)v_1'$, where $c^2 = v_1' v_1$.

We now apply Theorem 3.3.1 with $X_\alpha = Z_{2\alpha}$. Let $Y_\alpha = \sum_\beta c_{\alpha\beta} Z_{2\beta}$, $\alpha = 1, \dots, n$. Then $Y_1, \dots, Y_n$ are independently normally distributed with variance σ^2 and means

$$(10) \qquad \mathscr{E} Y_1 = \sum_{\gamma=1}^{n} c_{1\gamma} \beta z_{1\gamma} = \frac{\beta}{c} \sum_{\gamma=1}^{n} z_{1\gamma}^2 = \beta c,$$

$$(11) \qquad \mathscr{E} Y_\alpha = \sum_{\gamma=1}^{n} c_{\alpha\gamma} \beta z_{1\gamma} = \beta c \sum_{\gamma=1}^{n} c_{\alpha\gamma} c_{1\gamma} = 0, \qquad \alpha \neq 1.$$

We have $b = \sum_{\alpha=1}^n Z_{2\alpha} z_{1\alpha} / \sum_{\alpha=1}^n z_{1\alpha}^2 = c \sum_{\alpha=1}^n Z_{2\alpha} c_{1\alpha} / c^2 = Y_1/c$ and, from Lemma 3.3.1,

$$(12) \qquad U = \sum_{\alpha=1}^{n} Z_{2\alpha}^2 - b^2 \sum_{\alpha=1}^{n} z_{1\alpha}^2 = \sum_{\alpha=1}^{n} Y_\alpha^2 - Y_1^2 = \sum_{\alpha=2}^{n} Y_\alpha^2,$$

which is independent of b. Then U/σ^2 has a χ^2-distribution with $n - 1$ degrees of freedom.

Lemma 4.2.2. *If $(Z_{1\alpha}, Z_{2\alpha})$, $\alpha = 1, \dots, n$, are independent, each pair with density (7), then the conditional distributions of $b = \sum_{\alpha=1}^n Z_{2\alpha} Z_{1\alpha} / \sum_{\alpha=1}^n Z_{1\alpha}^2$ and $U/\sigma^2 = \sum_{\alpha=1}^n (Z_{2\alpha} - bZ_{1\alpha})^2/\sigma^2$ given $Z_{1\alpha} = z_{1\alpha}$, $\alpha = 1, \dots, n$, are $N(\beta, \sigma^2/c^2)$ $(c^2 = \sum_{\alpha=1}^n z_{1\alpha}^2)$ and χ^2 with $n - 1$ degrees of freedom, respectively; and b and U are independent.*

If $\rho = 0$, then $\beta = 0$, and b is distributed conditionally according to $N(0, \sigma^2/c^2)$, and

$$(13) \qquad \frac{cb/\sigma}{\sqrt{\dfrac{U/\sigma^2}{n-1}}} = \frac{cb}{\sqrt{\dfrac{U}{n-1}}}$$

has a conditional t-distribution with $n-1$ degrees of freedom. (See Problem 4.27.) However, this random variable is

$$(14)\qquad \sqrt{n-1}\,\frac{\sqrt{a_{11}}\,a_{12}/a_{11}}{\sqrt{a_{22}-a_{12}^2/a_{11}}} = \sqrt{n-1}\,\frac{a_{12}/\sqrt{a_{11}a_{22}}}{\sqrt{1-\left[a_{12}^2/(a_{11}a_{22})\right]}} = \sqrt{n-1}\,\frac{r}{\sqrt{1-r^2}}.$$

Thus $\sqrt{n-1}\,r/\sqrt{1-r^2}$ has a conditional t-distribution with $n-1$ degrees of freedom. The density of t is

$$(15)\qquad \frac{\Gamma(\frac{1}{2}n)}{\sqrt{n-1}\,\Gamma[\frac{1}{2}(n-1)]\sqrt{\pi}}\left(1+\frac{t^2}{n-1}\right)^{-\frac{1}{2}n},$$

and the density of $W = r/\sqrt{1-r^2}$ is

$$(16)\qquad \frac{\Gamma(\frac{1}{2}n)}{\Gamma[\frac{1}{2}(n-1)]\sqrt{\pi}}(1+w^2)^{-\frac{1}{2}n}.$$

Since $w = r(1-r^2)^{-\frac{1}{2}}$, we have $dw/dr = (1-r^2)^{-\frac{3}{2}}$. Therefore the density of r is (replacing n by $N-1$)

$$(17)\qquad \frac{\Gamma[\frac{1}{2}(N-1)]}{\Gamma[\frac{1}{2}(N-2)]\sqrt{\pi}}(1-r^2)^{\frac{1}{2}(N-4)}.$$

It should be noted that (17) is the conditional density of r for v_1 fixed. However, since (17) does not depend on v_1, it is also the marginal density of r.

Theorem 4.2.1. *Let $X_1, \dots, X_N$ be independent, each with distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. If $\rho_{ij} = 0$, the density of r_{ij} defined by* (1) *is* (17).

From (17) we see that the density is symmetric about the origin. For $N > 4$, it has a mode at $r = 0$ and its order of contact with the r-axis at ± 1 is $\frac{1}{2}(N-5)$ for N odd and $\frac{1}{2}N-3$ for N even. Since the density is even, the odd moments are zero; in particular, the mean is zero. The even moments are found by integration (letting $x = r^2$ and using the definition of the beta function). That $\mathscr{E}r^{2m} = \Gamma[\frac{1}{2}(N-1)]\Gamma(m+\frac{1}{2})/\{\sqrt{\pi}\,\Gamma[\frac{1}{2}(N-1)+m]\}$ and in particular that the variance is $1/(N-1)$ may be verified by the reader.

The most important use of Theorem 4.2.1 is to find significance points for testing the hypothesis that a pair of variables are not correlated. Consider the hypothesis

$$H: \rho_{ij} = 0 \tag{18}$$

for some particular pair (i, j). It would seem reasonable to reject this hypothesis if the corresponding sample correlation coefficient were very different from zero. Now how do we decide what we mean by "very different"?

Let us suppose we are interested in testing H against the alternative hypotheses $\rho_{ij} > 0$. Then we reject H if the sample correlation coefficient r_{ij} is greater than some number r_0. The probability of rejecting H when H is true is

$$\int_{r_0}^{1} k_N(r)\, dr, \tag{19}$$

where $k_N(r)$ is (17), the density of a correlation coefficient based on N observations. We choose r_0 so (19) is the desired significance level. If we test H against alternatives $\rho_{ij} < 0$, we reject H when $r_{ij} < -r_0$.

Now suppose we are interested in alternatives $\rho_{ij} \neq 0$; that is, ρ_{ij} may be either positive or negative. Then we reject the hypothesis H if $r_{ij} > r_1$ or $r_{ij} < -r_1$. The probability of rejection when H is true is

$$\int_{-1}^{-r_1} k_N(r)\, dr + \int_{r_1}^{1} k_N(r)\, dr. \tag{20}$$

The number r_1 is chosen so that (20) is the desired significance level.

The significance points r_1 are given in many books, including Table VI of Fisher and Yates (1942); the index n in Table VI is equal to our $N-2$. Since $\sqrt{N-2}\, r/\sqrt{1-r^2}$ has the t-distribution with $N-2$ degrees of freedom, t-tables can also be used. Against alternatives $\rho_{ij} \neq 0$, reject H if

$$\sqrt{N-2}\, \frac{|r_{ij}|}{\sqrt{1-r_{ij}^2}} > t_{N-2}(\alpha), \tag{21}$$

where $t_{N-2}(\alpha)$ is the two-tailed significance point of the t-statistic with $N-2$ degrees of freedom for significance level α. Against alternatives $\rho_{ij} > 0$, reject H if

$$\sqrt{N-2}\, \frac{r_{ij}}{\sqrt{1-r_{ij}^2}} > t_{N-2}(2\alpha). \tag{22}$$

From (13) and (14) we see that $\sqrt{N-2}\,r/\sqrt{1-r^2}$ is the proper statistic for testing the hypothesis that the regression of V_2 on v_1 is zero. In terms of the original observation $\{x_{i\alpha}\}$, we have

$$(23)\quad \sqrt{N-2}\,\frac{r}{\sqrt{1-r^2}} = \frac{b\sqrt{\sum_{\alpha=1}^{N}(x_{1\alpha}-\bar{x}_1)^2}}{\sqrt{\sum_{\alpha=1}^{N}\left[x_{2\alpha}-\bar{x}_2-b(x_{1\alpha}-\bar{x}_1)\right]^2/(N-2)}},$$

where $b=\sum_{\alpha=1}^{N}(x_{2\alpha}-\bar{x}_2)(x_{1\alpha}-\bar{x}_1)/\sum_{\alpha=1}^{N}(x_{1\alpha}-\bar{x}_1)^2$ is the least squares regression coefficient of $x_{2\alpha}$ on $x_{1\alpha}$. It is seen that the test of $\rho_{12}=0$ is equivalent to the test that the regression of X_2 on x_1 is zero (i.e., that $\rho_{12}\sigma_2/\sigma_1=0$).

To illustrate this procedure we consider the example given in Section 3.2. Let us test the null hypothesis that the effects of the two drugs are uncorrelated against the alternative that they are positively correlated. We shall use the 5% level of significance. For $N=10$, the 5% significance point (r_0) is 0.5494. Our observed correlation coefficient of 0.7952 is significant; we reject the hypothesis that the effects of the two drugs are independent.

4.2.2. The Distribution When the Population Correlation Coefficient Is Nonzero; Tests of Hypotheses and Confidence Intervals

To find the distribution of the sample correlation coefficient when the population coefficient is different from zero, we shall first derive the joint density of a_{11}, a_{12}, and a_{22}. In Section 4.2.1 we saw that, conditional on $\boldsymbol{v}_1$ held fixed, the random variables $b=a_{12}/a_{11}$ and $U/\sigma^2=(a_{22}-a_{12}^2/a_{11})/\sigma^2$ are distributed independently according to $N(\beta,\sigma^2/c^2)$ and the χ^2-distribution with $n-1$ degrees of freedom, respectively. Denoting the density of the χ^2-distribution by $g_{n-1}(u)$, we write the conditional density of b and U as $n(b|\beta,\sigma^2/a_{11})g_{n-1}(u/\sigma^2)/\sigma^2$. The joint density of V_1, b, and U is $n(v_1|\mathbf{0},\sigma_1^2 I)n(b|\beta,\sigma^2/a_{11})g_{n-1}(u/\sigma^2)/\sigma^2$. The marginal density of $V_1'V_1/\sigma_1^2=a_{11}/\sigma_1^2$ is $g_n(u)$; that is, the density of a_{11} is

$$(24)\qquad \frac{1}{\sigma_1^2}g_n\left(\frac{a_{11}}{\sigma_1^2}\right)=\int_{v_1'v_1=a_{11}}\cdots\int n\left(v_1|\mathbf{0},\sigma_1^2 I\right)dW,$$

where dW is the proper volume element.

The integration is over the sphere $v_1'v_1=a_{11}$; thus, dW is an element of area on this sphere. (See Problem 7.1 for the use of angular coordinates in defining dW.) Thus the joint density of b, U, and a_{11} is

$$(25)\quad \underset{v_1'v_1=a_{11}}{\int\cdots\int} n(b|\beta,\sigma^2/a_{11})g_{n-1}(u/\sigma^2)\frac{1}{\sigma^2}n(v_1|\mathbf{0},\sigma_1^2\boldsymbol{I})\,dW$$

$$= \frac{g_n(a_{11}/\sigma_1^2)n(b|\beta,\sigma^2/a_{11})g_{n-1}(u/\sigma^2)}{\sigma_1^2\sigma^2}$$

$$= \frac{(a_{11})^{\frac{1}{2}n-1}}{(2\sigma_1^2)^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}\exp\left(-\frac{a_{11}}{2\sigma_1^2}\right)\frac{\sqrt{a_{11}}}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{a_{11}}{2\sigma^2}(b-\beta)^2\right]$$

$$\cdot\frac{1}{(2\sigma^2)^{\frac{1}{2}(n-1)}\Gamma[\frac{1}{2}(n-1)]}u^{\frac{1}{2}(n-3)}\exp\left(-\frac{u}{2\sigma^2}\right).$$

Now let $b=a_{12}/a_{11}$, $U=a_{22}-a_{12}^2/a_{11}$. The Jacobian is

$$(26)\qquad \left|\frac{\partial(b,u)}{\partial(a_{12},a_{22})}\right| = \begin{vmatrix} \frac{1}{a_{11}} & 0 \\ -2\frac{a_{12}}{a_{11}} & 1 \end{vmatrix} = \frac{1}{a_{11}}.$$

Thus the density of a_{11}, a_{12}, and a_{22} for $a_{11}\ge 0$, $a_{22}\ge 0$, and $a_{11}a_{22}-a_{12}^2\ge 0$ is

$$(27)\qquad \frac{a_{11}^{\frac{1}{2}(n-3)}\left(\dfrac{a_{11}a_{22}-a_{12}^2}{a_{11}}\right)^{\frac{1}{2}(n-3)}e^{-\frac{1}{2}Q}}{2^n\sigma_1^n\sigma_2^n(1-\rho^2)^{\frac{1}{2}n}\sqrt{\pi}\,\Gamma(\frac{1}{2}n)\Gamma[\frac{1}{2}(n-1)]},$$

where

$$(28)\quad Q=\frac{a_{11}}{\sigma_1^2}+\frac{a_{11}}{\sigma^2}\left(\frac{a_{12}^2}{a_{11}^2}-2\rho\frac{\sigma_1\sigma_2}{\sigma_1^2}\frac{a_{12}}{a_{11}}+\frac{\rho^2\sigma_1^2\sigma_2^2}{\sigma_1^4}\right)+\frac{1}{\sigma^2}\left(a_{22}-\frac{a_{12}^2}{a_{11}}\right)$$

$$=a_{11}\left[\frac{1}{\sigma_1^2}+\frac{\rho^2\sigma_1^2\sigma_2^2}{\sigma_1^4\sigma_2^2(1-\rho^2)}\right]-2a_{12}\frac{\rho\sigma_2}{\sigma_1\sigma_2^2(1-\rho^2)}+\frac{a_{22}}{\sigma_2^2(1-\rho^2)}$$

$$=\frac{1}{1-\rho^2}\left(\frac{a_{11}}{\sigma_1^2}-2\rho\frac{a_{12}}{\sigma_1\sigma_2}+\frac{a_{22}}{\sigma_2^2}\right).$$

The density can be written

$$\frac{|A|^{\frac{1}{2}(n-3)}e^{-\frac{1}{2}Q}}{2^n|\Sigma|^{\frac{1}{2}n}\sqrt{\pi}\,\Gamma(\frac{1}{2}n)\Gamma[\frac{1}{2}(n-1)]} \tag{29}$$

for A positive definite, and 0 otherwise. This is a special case of the Wishart density derived in Chapter 7.

We want to find the density of

$$r=\frac{a_{12}}{\sqrt{a_{11}a_{22}}}=\frac{a_{12}/(\sigma_1\sigma_2)}{\sqrt{(a_{11}/\sigma_1^2)(a_{22}/\sigma_2^2)}}=\frac{a_{12}^*}{\sqrt{a_{11}^*a_{22}^*}}, \tag{30}$$

where $a_{11}^*=a_{11}/\sigma_1^2$, $a_{22}^*=a_{22}/\sigma_2^2$, and $a_{12}^*=a_{12}/(\sigma_1\sigma_2)$. The transformation is equivalent to setting $\sigma_1=\sigma_2=1$. Then the density of a_{11}, a_{22}, and $r=a_{12}/\sqrt{a_{11}a_{22}}$ $(da_{12}=dr\sqrt{a_{11}a_{22}})$ is

$$\frac{a_{11}^{\frac{1}{2}n-1}a_{22}^{\frac{1}{2}n-1}(1-r^2)^{\frac{1}{2}(n-3)}e^{-\frac{1}{2}Q}}{2^n(1-\rho^2)^{\frac{1}{2}n}\sqrt{\pi}\,\Gamma(\frac{1}{2}n)\Gamma[\frac{1}{2}(n-1)]}, \tag{31}$$

where

$$Q=\frac{a_{11}-2\rho r\sqrt{a_{11}}\sqrt{a_{22}}+a_{22}}{1-\rho^2}. \tag{32}$$

To find the density of r, we must integrate (31) with respect to a_{11} and a_{22} over the range 0 to ∞. There are various ways of carrying out the integration, which result in different expressions for the density. The method we shall indicate here is straightforward. We expand part of the exponential:

$$\exp\left[\frac{\rho r\sqrt{a_{11}}\sqrt{a_{22}}}{(1-\rho^2)}\right]=\sum_{\alpha=0}^{\infty}\frac{\left(\rho r\sqrt{a_{11}}\sqrt{a_{22}}\right)^\alpha}{\alpha!(1-\rho^2)^\alpha}. \tag{33}$$

Then the density (31) is

$$\frac{(1-r^2)^{\frac{1}{2}(n-3)}}{(1-\rho^2)^{\frac{1}{2}n}2^n\sqrt{\pi}\,\Gamma(\frac{1}{2}n)\Gamma[\frac{1}{2}(n-1)]}\sum_{\alpha=0}^{\infty}\frac{(\rho r)^\alpha}{\alpha!(1-\rho^2)^\alpha} \tag{34}$$
$$\cdot\left\{\exp\left[-\frac{a_{11}}{2(1-\rho^2)}\right]a_{11}^{(n+\alpha)/2-1}\right\}\left\{\exp\left[-\frac{a_{22}}{2(1-\rho^2)}\right]a_{22}^{(n+\alpha)/2-1}\right\}.$$

Since

$$\int_0^\infty a^{\frac{1}{2}(n+\alpha)-1}\exp\left[-\frac{a}{2(1-\rho^2)}\right]da = \Gamma\left[\tfrac{1}{2}(n+\alpha)\right]\left[2(1-\rho^2)\right]^{\frac{1}{2}(n+\alpha)}, \tag{35}$$

the integral of (34) (term-by-term integration is permissible) is

$$\begin{aligned}(36)\qquad &\frac{(1-r^2)^{\frac{1}{2}(n-3)}}{(1-\rho^2)^{\frac{1}{2}n}2^n\sqrt{\pi}\,\Gamma(\frac{1}{2}n)\Gamma[\frac{1}{2}(n-1)]}\\ &\quad\cdot\sum_{\alpha=0}^{\infty}\frac{(\rho r)^\alpha}{\alpha!(1-\rho^2)^\alpha}\Gamma^2\left[\tfrac{1}{2}(n+\alpha)\right]2^{n+\alpha}(1-\rho^2)^{n+\alpha}\\ &=\frac{(1-\rho^2)^{\frac{1}{2}n}(1-r^2)^{\frac{1}{2}(n-3)}}{\sqrt{\pi}\,\Gamma(\frac{1}{2}n)\Gamma[\frac{1}{2}(n-1)]}\sum_{\alpha=0}^{\infty}\frac{(2\rho r)^\alpha}{\alpha!}\Gamma^2\left[\tfrac{1}{2}(n+\alpha)\right].\end{aligned}$$

The *duplication formula for the gamma function* is

$$\Gamma(2z) = \frac{2^{2z-1}\Gamma(z)(z+\frac{1}{2})}{\sqrt{\pi}}. \tag{37}$$

It can be used to modify the constant in (36).

Theorem 4.2.2. *The correlation coefficient in a sample of N from a bivariate normal distribution with correlation ρ is distributed with density*

$$\frac{2^{n-2}(1-\rho^2)^{\frac{1}{2}n}(1-r^2)^{\frac{1}{2}(n-3)}}{(n-2)!\,\pi}\sum_{\alpha=0}^{\infty}\frac{(2\rho r)^\alpha}{\alpha!}\Gamma^2\left[\tfrac{1}{2}(n+\alpha)\right], \tag{38}$$

$$-1 \le r \le 1,$$

where $n = N-1$.

The distribution of r was first found by Fisher (1915). He also gave as another form of the density,

$$\frac{(1-\rho^2)^{\frac{1}{2}n}(1-r^2)^{\frac{1}{2}(n-3)}}{\pi(n-2)!}\left[\frac{d^{n-1}}{dx^{n-1}}\left\{\frac{\cos^{-1}(-x)}{\sqrt{1-x^2}}\right\}\bigg|_{x=r\rho}\right]. \tag{39}$$

See Problem 4.24.

Hotelling (1953) has made an exhaustive study of the distribution of r. He has recommended the following form:

$$\text{(40)} \qquad \frac{n-1}{\sqrt{2\pi}} \frac{\Gamma(n)}{\Gamma(n+\frac{1}{2})} (1-\rho^2)^{\frac{1}{2}n} (1-r^2)^{\frac{1}{2}(n-3)} \cdot (1-\rho r)^{-n+\frac{1}{2}} F\left(\tfrac{1}{2},\tfrac{1}{2};\, n+\tfrac{1}{2};\, \frac{1+\rho r}{2}\right),$$

where

$$\text{(41)} \qquad F(a,b;c;x) = \sum_{j=0}^{\infty} \frac{\Gamma(a+j)}{\Gamma(a)} \frac{\Gamma(b+j)}{\Gamma(b)} \frac{\Gamma(c)}{\Gamma(c+j)} \frac{x^j}{j!}$$

is a *hypergeometric function*. (See Problem 4.25.) The series in (40) converges more rapidly than the one in (38). Hotelling discusses methods of integrating the density and also calculates moments of r.

The cumulative distribution of r,

$$\text{(42)} \qquad \Pr\{r \le r^*\} = F(r^*|N,\rho),$$

has been tabulated by David (1938) for† $\rho = 0(.1).9$, $N = 3(1)25, 50, 100, 200, 400$, and $r^* = -1(.05)1$. (David's n is our N.) It is clear from the density (38) that $F(r^*|N,\rho) = 1 - F(-r^*|N,-\rho)$ because the density for r, ρ is equal to the density for $-r, -\rho$. These tables can be used for a number of statistical procedures.

First, we consider the problem of using a sample to test the hypothesis

$$\text{(43)} \qquad H\colon \rho = \rho_0.$$

If the alternatives are $\rho > \rho_0$, we reject the hypothesis if the sample correlation coefficient is greater than r_0, where r_0 is chosen so $1 - F(r_0|N,\rho_0) = \alpha$, the significance level. If the alternatives are $\rho < \rho_0$, we reject the hypothesis if the sample correlation coefficient is less than r_0', where r_0' is chosen so $F(r_0'|N,\rho_0) = \alpha$. If the alternatives are $\rho \ne \rho_0$, the region of rejection is $r > r_1$ and $r < r_1'$, where r_1 and r_1' are chosen so $[1 - F(r_1|N,\rho_0)] + F(r_1'|N,\rho_0) = \alpha$. David suggests that r_1 and r_1' be chosen so $[1 - F(r_1|N,\rho_0)] = F(r_1'|N,\rho_0) = \frac{1}{2}\alpha$. She has shown (1937) that for $N \ge 10$, $|\rho| \le 0.8$ this critical region is nearly the region of an unbiased test of H, that is, a test whose power function has its minimum at ρ_0.

It should be pointed out that any test based on r is invariant under transformations of location and scale, that is, $x_{i\alpha}^* = b_i x_{i\alpha} + c_i$, $b_i > 0$, $i = 1, 2$,

† $\rho = 0(.1).9$ means $\rho = 0, 0.1, 0.2, \ldots, 0.9$.

Table 4.1. A Power Function

ρ	Probability
-1.0	0.0000
-0.8	0.0000
-0.6	0.0004
-0.4	0.0032
-0.2	0.0147
0.0	0.0500
0.2	0.1376
0.4	0.3215
0.6	0.6235
0.8	0.9279
1.0	1.0000

$\alpha = 1, \dots, N$; and r is essentially the only invariant of the sufficient statistics (Problem 3.7). The above procedure for testing $H: \rho = \rho_0$ against alternatives $\rho > \rho_0$ is uniformly most powerful among all invariant tests. (See Problems 4.16, 4.17, and 4.18.)

As an example suppose one wishes to test the hypothesis that $\rho = 0.5$ against alternatives $\rho \neq 0.5$ at the 5% level of significance using the correlation observed in a sample of 15. In David's tables we find (by interpolation) that $F(0.027 | 15, 0.5) = 0.025$ and $F(0.805 | 15, 0.5) = 0.975$. Hence we reject the hypothesis if our sample r is less than 0.027 or greater than 0.805.

Secondly, we can use David's tables to compute the power function of a test of correlation. If the region of rejection of H is $r > r_1$ and $r < r_1'$, the power of the test is a function of the true correlation ρ, namely $[1 - F(r_1 | N, \rho) + [F(r_1' | N, \rho)]$; this is the probability of rejecting the null hypothesis when the population correlation is ρ.

As an example consider finding the power function of the test for $\rho = 0$ considered in the preceding section. The rejection region (one-sided) is $r \geq 0.5494$ at the 5% significance level. The probabilities of rejection are given in Table 4.1. The graph of the power function is illustrated in Figure 4.2.

Thirdly, David's computations lead to confidence regions for ρ. For given N, r_1' (defining a significance point) is a function of ρ, say $f_1(\rho)$, and r_1 is another function of ρ, say $f_2(\rho)$, such that

$$\Pr\{f_1(\rho) < r < f_2(\rho) | \rho\} = 1 - \alpha. \tag{44}$$

Clearly, $f_1(\rho)$ and $f_2(\rho)$ are monotonically increasing functions of ρ if r_1 and r_1' are chosen so $1 - F(r_1 | N, \rho) = \frac{1}{2}\alpha = F(r_1' | N, \rho)$. If $\rho = f_i^{-1}(r)$ is the

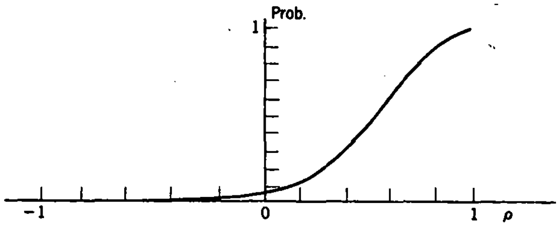

Figure 4.2. A power function.

inverse of $r = f_i(\rho)$, $i = 1, 2$, then the inequality $f_1(\rho) < r$ is equivalent to† $\rho < f_1^{-1}(r)$, and $r < f_2(\rho)$ is equivalent to $f_2^{-1}(r) < \rho$. Thus (44) can be written

$$\Pr\{f_2^{-1}(r) < \rho < f_1^{-1}(r) \mid \rho\} = 1 - \alpha. \tag{45}$$

This equation says that the probability is $1 - \alpha$ that we draw a sample such that the interval $(f_2^{-1}(r), f_1^{-1}(r))$ covers the parameter ρ. Thus this interval is a confidence interval for ρ with confidence coefficient $1 - \alpha$. For a given N and α the curves $r = f_1(\rho)$ and $r = f_2(\rho)$ appear as in Figure 4.3. In testing the hypothesis $\rho = \rho_0$, the intersection of the line $\rho = \rho_0$ and the two curves gives the significance points r_1 and r_1'. In setting up a confidence region for ρ on the basis of a sample correlation r^*, we find the limits $f_2^{-1}(r^*)$ and

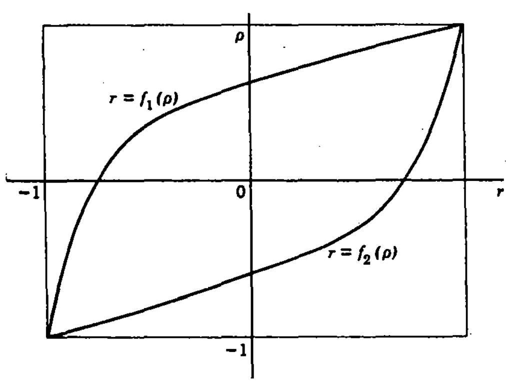

Figure 4.3

†The point $(f_1(\rho), \rho)$ on the first curve is to the left of (r, ρ), and the point $(r, f_1^{-1}(r))$ is above (r, ρ).

$f_1^{-1}(r^*)$ by the intersection of the line $r = r^*$ with the two curves. David gives these curves for $\alpha = 0.1$, 0.05, 0.02, and 0.01 for various values of N. One-sided confidence regions can be obtained by using only one inequality above.

The tables of $F(r|N, \rho)$ can also be used instead of the curves for finding the confidence interval. Given the sample value r^*, $f_1^{-1}(r^*)$ is the value of ρ such that $\frac{1}{2}\alpha = \Pr\{r \le r^* | \rho\} = F(r^* | N, \rho)$, and similarly $f_2^{-1}(r^*)$ is the value of ρ such that $\frac{1}{2}\alpha = \Pr\{r \ge r^* | \rho\} = 1 - F(r^* | N, \rho)$. The interval between these two values of ρ, $(f_2^{-1}(r^*), f_1^{-1}(r^*))$, is the confidence interval.

As an example, consider the confidence interval with confidence coefficient 0.95 based on the correlation of 0.7952 observed in a sample of 10. Using Graph II of David, we find the two limits are 0.34 and 0.94. Hence we state that $0.34 < \rho < 0.94$ with confidence 95%.

Definition 4.2.1. *Let $L(x, \boldsymbol{\theta})$ be the likelihood function of the observation vector $\mathbf{x}$ and the parameter vector $\boldsymbol{\theta} \in \Omega$. Let a null hypothesis be defined by a proper subset ω of Ω. The* likelihood ratio criterion *is*

$$\lambda(x) = \frac{\sup_{\boldsymbol{\theta} \in \omega} L(x, \boldsymbol{\theta})}{\sup_{\boldsymbol{\theta} \in \Omega} L(x, \boldsymbol{\theta})}. \tag{46}$$

The likelihood ratio test *is the procedure of rejecting the null hypothesis when $\lambda(x)$ is less than a predetermined constant.*

Intuitively, one rejects the null hypothesis if the density of the observations under the most favorable choice of parameters in the null hypothesis is much less than the density under the most favorable unrestricted choice of the parameters. Likelihood ratio tets have some desirable features; see Lehmann (1959), for example. Wald (1943) has proved some favorable asymptotic properties. For most hypotheses concerning the multivariate normal distribution, likelihood ratio tests are appropriate and often are optimal.

Let us consider the likelihood ratio test of the hypothesis that $\rho = \rho_0$ based on a sample $x_1, \dots, x_N$ from the bivariate normal distribution. The set Ω consists of μ_1, μ_2, σ_1, σ_2, and ρ such that $\sigma_1 > 0$, $\sigma_2 > 0$, $-1 < \rho < 1$. The set ω is the subset for which $\rho = \rho_0$. The likelihood maximized in Ω is (by Lemmas 3.2.2 and 3.2.3)

$$\max_{\Omega} L = \frac{N^N e^{-N}}{(2\pi)^N (1 - r^2)^{\frac{1}{2}N} a_{11}^{N/2} a_{22}^{N/2}}. \tag{47}$$

Under the null hypothesis the likelihood function is

$$(48)\qquad \frac{1}{(2\pi)^N(1-\rho_0^2)^{\frac{1}{2}N}(\sigma^2)^N}\exp\left[-\frac{a_{11}/\tau+\tau a_{22}-2\rho_0 a_{12}}{2\sigma^2(1-\rho_0^2)}\right],$$

where $\sigma^2=\sigma_1\sigma_2$ and $\tau=\sigma_1/\sigma_2$. The maximum of (48) with respect to τ occurs at $\hat{\tau}=\sqrt{a_{11}}\,/\sqrt{a_{22}}$. The concentrated likelihood is

$$(49)\qquad \frac{1}{(2\pi)^N(1-\rho_0^2)^{\frac{1}{2}N}(\sigma^2)^N}\exp\left[-\frac{\sqrt{a_{11}}\sqrt{a_{22}}(1-\rho_0 r)}{\sigma^2(1-\rho_0^2)}\right];$$

the maximum of (49) occurs at

$$(50)\qquad \hat{\sigma}^2=\frac{a_{11}^{\frac{1}{2}}a_{22}^{\frac{1}{2}}(1-\rho_0 r)}{N(1-\rho_0^2)}.$$

The likelihood ratio criterion is, therefore,

$$(51)\qquad \frac{\max_\omega L}{\max_\Omega L}=\frac{(1-\rho_0^2)^{\frac{1}{2}N}(1-r^2)^{\frac{1}{2}N}}{(1-\rho_0 r)^N}=\left[\frac{(1-\rho_0^2)(1-r^2)}{(1-\rho_0 r)^2}\right]^{\frac{1}{2}N}.$$

The likelihood ratio test is $(1-\rho_0^2)(1-r^2)(1-\rho_0 r)^{-2}<c$, where c is chosen so the probability of the inequality when samples are drawn from normal populations with correlation ρ_0 is the prescribed significance level. The critical region can be written equivalently as

$$(52)\qquad (\rho_0^2 c-\rho_0^2+1)r^2-2\rho_0 cr+c-1+\rho_0^2>0,$$

or

$$(53)\qquad \begin{aligned} r &> \frac{\rho_0 c+(1-\rho_0^2)\sqrt{1-c}}{\rho_0^2 c+1-\rho_0^2},\\ r &< \frac{\rho_0 c-(1-\rho_0^2)\sqrt{1-c}}{\rho_0^2 c+1-\rho_0^2}. \end{aligned}$$

Thus the likelihood ratio test of $H:\rho=\rho_0$ against alternatives $\rho\neq\rho_0$ has a rejection region of the form $r>r_1$ and $r<r_1'$; but r_1 and r_1' are not chosen so that the probability of each inequality is $\alpha/2$ when H is true, but are taken to be of the form given in (53), where c is chosen so that the probability of the two inequalities is α.

4.2.3. The Asymptotic Distribution of a Sample Correlation Coefficient and Fisher's z

In this section we shall show that as the sample size increases, a sample correlation coefficient tends to be normally distributed. The distribution of a particular function of a sample correlation, Fisher's z [Fisher (1921)], which has a variance approximately independent of the population correlation, tends to normality faster.

We are particularly interested in the sample correlation coefficient

$$r(n) = \frac{A_{ij}(n)}{\sqrt{A_{ii}(n) A_{jj}(n)}} \tag{54}$$

for some i and j, $i \neq j$. This can also be written

$$r(n) = \frac{C_{ij}(n)}{\sqrt{C_{ii}(n) C_{jj}(n)}}, \tag{55}$$

where $C_{gh}(n) = A_{gh}(n)/\sqrt{\sigma_{gg}\sigma_{hh}}$. The set $C_{ii}(n)$, $C_{jj}(n)$, and $C_{ij}(n)$ is distributed like the distinct elements of the matrix

$$\sum_{\alpha=1}^{n} \begin{pmatrix} Z_{i\alpha}^* \\ Z_{j\alpha}^* \end{pmatrix} (Z_{i\alpha}^*, Z_{j\alpha}^*) = \sum_{\alpha=1}^{n} \begin{pmatrix} Z_{i\alpha}/\sqrt{\sigma_{ii}} \\ Z_{j\alpha}/\sqrt{\sigma_{jj}} \end{pmatrix} \left(Z_{i\alpha}/\sqrt{\sigma_{ii}}, Z_{j\alpha}/\sqrt{\sigma_{jj}} \right), \tag{56}$$

where the $(Z_{i\alpha}^*, Z_{j\alpha}^*)$ are independent, each with distribution

$$N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right],$$

where

$$\rho = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}.$$

Let

$$U(n) = \frac{1}{n} \begin{pmatrix} C_{ii}(n) \\ C_{jj}(n) \\ C_{ij}(n) \end{pmatrix}, \tag{57}$$

$$b = \begin{pmatrix} 1 \\ 1 \\ \rho \end{pmatrix}. \tag{58}$$

Then by Theorem 3.4.4 the vector $\sqrt{n}\,[U(n)-b]$ has a limiting normal distribution with mean $\mathbf{0}$ and covariance matrix

$$\begin{pmatrix} 2 & 2\rho^2 & 2\rho \\ 2\rho^2 & 2 & 2\rho \\ 2\rho & 2\rho & 1+\rho^2 \end{pmatrix}. \tag{59}$$

Now we need the general theorem:

Theorem 4.2.3. *Let $\{U(n)\}$ be a sequence of m-component random vectors and b a fixed vector such that $\sqrt{n}\,[U(n)-b]$ has the limiting distribution $N(\mathbf{0},\mathbf{T})$ as $n\to\infty$. Let $f(u)$ be a vector-valued function of u such that each component $f_j(u)$ has a nonzero differential at $u=b$, and let $\partial f_j(u)/\partial u_i|_{u=b}$ be the i, jth component of $\mathbf{\Phi}_b$. Then $\sqrt{n}\,\{f[u(n)]-f(b)\}$ has the limiting distribution $N(\mathbf{0},\mathbf{\Phi}_b'\mathbf{T}\mathbf{\Phi}_b)$.*

Proof. See Serfling (1980), Section 3.3, or Rao (1973), Section 6a.2. A function $g(u)$ is said to have a differential at b or to be totally differentiable at b if the partial derivatives $\partial g(u)/\partial u_i$ exist at $u=b$ and for every $\varepsilon>0$ there exists a neighborhood $N_\varepsilon(b)$ such that

$$\left| g(u) - g(b) - \sum_{i=1}^{m} \frac{\partial g(u)}{\partial u_i}(u_i - b_i) \right| \le \varepsilon \|u - b\| \qquad \text{for all } u \in N_\varepsilon(b). \tag{60}$$ ■

It is clear that $U(n)$ defined by (57) with b and $\mathbf{T}$ defined by (58) and (59), respectively, satisfies the conditions of the theorem. The function

$$r = \frac{u_3}{\sqrt{u_1 u_2}} = u_3 u_1^{-\frac{1}{2}} u_2^{-\frac{1}{2}} \tag{61}$$

satisfies the conditions; the elements of $\mathbf{\Phi}_b$ are

$$\begin{aligned} \left.\frac{\partial r}{\partial u_1}\right|_{u=b} &= -\tfrac{1}{2} u_3 u_1^{-\frac{3}{2}} u_2^{-\frac{1}{2}}\big|_{u=b} = -\tfrac{1}{2}\rho, \\ \left.\frac{\partial r}{\partial u_2}\right|_{u=b} &= -\tfrac{1}{2} u_3 u_1^{-\frac{1}{2}} u_2^{-\frac{3}{2}}\big|_{u=b} = -\tfrac{1}{2}\rho, \\ \left.\frac{\partial r}{\partial u_3}\right|_{u=b} &= u_1^{-\frac{1}{2}} u_2^{-\frac{1}{2}}\big|_{u=b} = 1, \end{aligned} \tag{62}$$

and $f(b) = \rho$. The variance of the limiting distribution of $\sqrt{n}\,[r(n) - \rho]$ is

$$(63) \qquad (-\tfrac{1}{2}\rho, -\tfrac{1}{2}\rho, 1)\begin{pmatrix} 2 & 2\rho^2 & 2\rho \\ 2\rho^2 & 2 & 2\rho \\ 2\rho & 2\rho & 1+\rho^2 \end{pmatrix}\begin{pmatrix} -\tfrac{1}{2}\rho \\ -\tfrac{1}{2}\rho \\ 1 \end{pmatrix}$$

$$= (\rho - \rho^3, \rho - \rho^3, 1 - \rho^2)\begin{pmatrix} -\tfrac{1}{2}\rho \\ -\tfrac{1}{2}\rho \\ 1 \end{pmatrix}$$

$$= 1 - 2\rho^2 + \rho^4$$

$$= (1 - \rho^2)^2.$$

Thus we obtain the following:

Theorem 4.2.4. *If $r(n)$ is the sample correlation coefficient of a sample of N $(= n + 1)$ from a normal distribution with correlation ρ, then $\sqrt{n}\,[r(n) - \rho]/(1 - \rho^2)$ [or $\sqrt{N}\,[r(n) - \rho]/(1 - \rho^2)$] has the limiting distribution $N(0, 1)$.*

It is clear from Theorem 4.2.3 that if $f(x)$ is differentiable at $x = \rho$, then $\sqrt{n}\,[f(r) - f(\rho)]$ is asymptotically normally distributed with mean zero and variance

$$\left(\left.\frac{\partial f}{\partial x}\right|_{x=\rho}\right)^2 (1 - \rho^2)^2.$$

A useful function to consider is one whose asymptotic variance is constant (here unity) independent of the parameter ρ. This function satisfies the equation

$$(64) \qquad f'(\rho) = \frac{1}{1-\rho^2} = \frac{1}{2}\left(\frac{1}{1+\rho} + \frac{1}{1-\rho}\right).$$

Thus $f(\rho)$ is taken as $\frac{1}{2}[\log(1+\rho) - \log(1-\rho)] = \frac{1}{2}\log[(1+\rho)/(1-\rho)]$. The so-called Fisher's z is

$$(65) \qquad z = \tfrac{1}{2}\log\frac{1+r}{1-r} = \tanh^{-1} r,$$

where $r = \tanh z = (e^z - e^{-z})/(e^z + e^{-z})$. Let

$$(66) \qquad \zeta = \tfrac{1}{2}\log\frac{1+\rho}{1-\rho}.$$

Theorem 4.2.5. *Let z be defined by* (65), *where r is the correlation coefficient of a sample of N $(=n+1)$ from a bivariate normal distribution with correlation ρ; let ζ be defined by* (66). *Then $\sqrt{n}(z-\zeta)$ has a limiting normal distribution with mean* 0 *and variance* 1.

It can be shown that to a closer approximation

$$\mathscr{E}z \sim \zeta + \frac{\rho}{2n}. \tag{67}$$

$$\mathscr{E}(z-\zeta)^2 \sim \frac{1}{n-2} \sim \mathscr{E}\left(z-\zeta-\frac{\rho}{2n}\right)^2. \tag{68}$$

The latter follows from

$$\mathscr{E}(z-\zeta)^2 = \frac{1}{n} + \frac{8-\rho^2}{4n^2} + \cdots \tag{69}$$

and holds good for ρ^2/n^2 small. Hotelling (1953) gives moments of z to order n^{-3}. An important property of Fisher's z is that the approach to normality is much more rapid than for r. David (1938) makes some comparisons between the tabulated probabilities and the probabilities computed by assuming z is normally distributed. She recommends that for $N > 25$ one take z as normally distributed with mean and variance given by (67) and (68). Konishi (1978a, 1978b, 1979) has also studied z. [Ruben (1966) has suggested an alternative approach, which is more complicated, but possibly more accurate.]

We shall now indicate how Theorem 4.2.5 can be used.

a. Suppose we wish to test the hypothesis $\rho = \rho_0$ on the basis of a sample of N against the alternatives $\rho \neq \rho_0$. We compute r and then z by (65). Let

$$\zeta_0 = \tfrac{1}{2}\log\frac{1+\rho_0}{1-\rho_0}. \tag{70}$$

Then a region of rejection at the 5% significance level is

$$\sqrt{N-3}\,|z-\zeta_0| > 1.96. \tag{71}$$

A better region is

$$\sqrt{N-3}\left|z-\zeta_0-\frac{\frac{1}{2}\rho_0}{N-1}\right| > 1.96. \tag{72}$$

b. Suppose we have a sample of N_1 from one population and a sample of N_2 from a second population. How do we test the hypothesis that the two correlation coefficients are equal, $\rho_1 = \rho_2$? From Theorem 4.2.5 we know that if the null hypothesis is true then $z_1 - z_2$ [where z_1 and z_2 are defined by (65) for the two sample correlation coefficients] is asymptotically normally distributed with mean 0 and variance $1/(N_1 - 3) + 1/(N_2 - 3)$. As a critical region of size 5%, we use

$$\frac{|z_1 - z_2|}{\sqrt{1/(N_1 - 3) + 1/(N_2 - 3)}} > 1.96. \tag{73}$$

c. Under the conditions of paragraph b, assume that $\rho_1 = \rho_2 = \rho$. How do we use the results of both samples to give a joint estimate of ρ? Since z_1 and z_2 have variances $1/(N_1 - 3)$ and $1/(N_2 - 3)$, respectively, we can estimate ζ by

$$\frac{(N_1 - 3)z_1 + (N_2 - 3)z_2}{N_1 + N_2 - 6} \tag{74}$$

and convert this to an estimate of ρ by the inverse of (65).

d. Let r be the sample correlation from N observations. How do we obtain a confidence interval for ρ? We know that approximately

$$\Pr\{-1.96 \le \sqrt{N-3}\,(z - \zeta) \le 1.96\} = 0.95. \tag{75}$$

From this we deduce that $[-1.96/\sqrt{N-3} + z, 1.96/\sqrt{N-3} + z]$ is a confidence interval for ζ. From this we obtain an interval for ρ using the fact $\rho = \tanh \zeta = (e^{\zeta} - e^{-\zeta})/(e^{\zeta} + e^{-\zeta})$, which is a monotonic transformation. Thus a 95% confidence interval is

$$\tanh(z - 1.96/\sqrt{N-3}) \le \rho \le \tanh(z + 1.96/\sqrt{N-3}). \tag{76}$$

The *bootstrap* method has been developed to assess the variability of a sample quantity. See Efron (1982). We shall illustrate the method on the sample correlation coefficient, but it can be applied to other quantities studied in this book.

Suppose $x_1, \dots, x_N$ is a sample from some bivariate population not necessarily normal. The approach of the bootstrap is to consider these N vectors as a finite population of size N; a random vector X has the (discrete) probability

$$\Pr\{X = x_\alpha\} = \frac{1}{N}, \qquad \alpha = 1, \dots, N. \tag{77}$$

A random sample of size N drawn from this finite population has a probability distribution, and the correlation coefficient calculated from such a sample has a (discrete) probability distribution, say $p_N(r)$. The bootstrap proposes to use this distribution in place of the unobtainable distribution of the correlation coefficient of random samples from the parent population. However, it is prohibitively expensive to compute; instead $p_N(r)$ is estimated by the empirical distribution of r calculated from a large number of random samples from (77). Diaconis and Efron (1983) have given an example of $N = 15$; they find the empirical distribution closely resembles the actual distribution of r (essentially obtainable in this special case). An advantage of this approach is that it is not necessary to assume knowledge of the parent population; a disadvantage is the massive computation.

4.3. PARTIAL CORRELATION COEFFICIENTS; CONDITIONAL DISTRIBUTIONS

4.3.1. Estimation of Partial Correlation Coefficients

Partial correlation coefficients in normal distributions are correlation coefficients in conditional distributions. It was shown in Section 2.5 that if X is distributed according to $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$(1) \qquad X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

then the conditional distribution of $X^{(1)}$ given $X^{(2)} = x^{(2)}$ is $N[\boldsymbol{\mu}^{(1)} + \boldsymbol{\beta}(x^{(2)} - \boldsymbol{\mu}^{(2)}), \boldsymbol{\Sigma}_{11\cdot2}]$, where

$$(2) \qquad \boldsymbol{\beta} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1},$$

$$(3) \qquad \boldsymbol{\Sigma}_{11\cdot2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}.$$

The partial correlations of $X^{(1)}$ given $x^{(2)}$ are the correlations calculated in the usual way from $\boldsymbol{\Sigma}_{11\cdot2}$. In this section we are interested in statistical problems concerning these correlation coefficients.

First we consider the problem of estimation on the basis of a sample of N from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. What are the maximum likelihood estimators of the partial correlations of $X^{(1)}$ (of q components), $\rho_{ij\cdot q+1,\ldots,p}$? We know that the maximum likelihood estimator of $\boldsymbol{\Sigma}$ is $(1/N)A$, where

$$(4)\qquad \begin{aligned} A &= \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(x_\alpha - \bar{x})' \\ &= \sum_{\alpha=1}^{N} \begin{pmatrix} x_\alpha^{(1)} - \bar{x}^{(1)} \\ x_\alpha^{(2)} - \bar{x}^{(2)} \end{pmatrix} \left(x_\alpha^{(1)\prime} - \bar{x}^{(1)\prime}, x_\alpha^{(2)\prime} - \bar{x}^{(2)\prime} \right) \\ &= \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \end{aligned}$$

and $\bar{x} = (1/N)\Sigma_{\alpha=1}^{N} x_\alpha = (\bar{x}^{(1)\prime}\ \bar{x}^{(2)\prime})'$. The correspondence between $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{11\cdot 2}$, $\boldsymbol{\beta}$, and $\boldsymbol{\Sigma}_{22}$ is one-to-one by virtue of (2) and (3) and

$$(5)\qquad \boldsymbol{\Sigma}_{12} = \boldsymbol{\beta}\boldsymbol{\Sigma}_{22},$$

$$(6)\qquad \boldsymbol{\Sigma}_{11} = \boldsymbol{\Sigma}_{11\cdot 2} + \boldsymbol{\beta}\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}'.$$

We can now apply Corollary 3.2.1 to the effect that maximum likelihood estimators of functions of parameters are those functions of the maximum likelihood estimators of those parameters.

Theorem 4.3.1. *Let $x_1, \dots, x_N$ be a sample from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are partitioned as in* (1). *Define A by* (4) *and $(\bar{x}^{(1)\prime}\ \bar{x}^{(2)\prime}) = (1/N)\Sigma_{\alpha=1}^{N}(x_\alpha^{(1)\prime}\ x_\alpha^{(2)\prime})$. Then the maximum likelihood estimators of $\boldsymbol{\mu}^{(1)}$, $\boldsymbol{\mu}^{(2)}$, $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_{11\cdot 2}$, and $\boldsymbol{\Sigma}_{22}$ are $\hat{\boldsymbol{\mu}}^{(1)} = \bar{x}^{(1)}$, $\hat{\boldsymbol{\mu}}^{(2)} = \bar{x}^{(2)}$,*

$$(7)\qquad \hat{\boldsymbol{\beta}} = A_{12}A_{22}^{-1}, \qquad \hat{\boldsymbol{\Sigma}}_{11\cdot 2} = \frac{1}{N}\left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right),$$

and $\hat{\boldsymbol{\Sigma}}_{22} = (1/N)A_{22}$, respectively.

In turn, Corollary 3.2.1 can be used to obtain the maximum likelihood estimators of $\boldsymbol{\mu}^{(1)}$, $\boldsymbol{\mu}^{(2)}$, $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_{22}$, $\sigma_{ii\cdot q+1,\dots,p}$, $i = 1, \dots, q$, and $\rho_{ij\cdot q+1,\dots,p}$, $i, j = 1, \dots, q$. It follows that the maximum likelihood estimators of the partial correlation coefficients are

$$(8)\qquad \hat{\rho}_{ij\cdot q+1,\dots,p} = \frac{\hat{\sigma}_{ij\cdot q+1,\dots,p}}{\sqrt{\hat{\sigma}_{ii\cdot q+1,\dots,p}\,\hat{\sigma}_{jj\cdot q+1,\dots,p}}}, \qquad i, j = 1, \dots, q,$$

where $\hat{\sigma}_{ij\cdot q+1,\dots,p}$ is the i,jth element of $\hat{\boldsymbol{\Sigma}}_{11\cdot 2}$.

Theorem 4.3.2. *Let $x_1, \ldots, x_N$ be a sample of N from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The maximum likelihood estimators of $\rho_{ij\cdot q+1,\ldots,p}$, the partial correlations of the first q components conditional on the last $p-q$ components, are given by*

$$(9) \qquad \hat{\rho}_{ij\cdot q+1,\ldots,p} = \frac{a_{ij\cdot q+1,\ldots,p}}{\sqrt{a_{ii\cdot q+1,\ldots,p}\,a_{jj\cdot q+1,\ldots,p}}}, \qquad i,j = 1,\ldots,q,$$

where

$$(10) \qquad (a_{ij\cdot q+1,\ldots,p}) = A_{11} - A_{12}A_{22}^{-1}A_{21} = A_{11\cdot 2}.$$

The estimator $\hat{\rho}_{ij\cdot q+1,\ldots,p}$, denoted by $r_{ij\cdot q+1,\ldots,p}$, is called the *sample partial correlation coefficient between X_i and X_j holding $X_{q+1},\ldots,X_p$ fixed.* It is also called the sample partial correlation coefficient between X_i and X_j having taken account of $X_{q+1},\ldots,X_p$. Note that the calculations can be done in terms of (r_{ij}).

The matrix $A_{11\cdot 2}$ can also be represented as

$$(11) \qquad A_{11\cdot 2} = \sum_{\alpha=1}^{N}\left[x_\alpha^{(1)} - \bar{x}^{(1)} - \hat{\boldsymbol{\beta}}\left(x_\alpha^{(2)} - \bar{x}^{(2)}\right)\right]\left[x_\alpha^{(1)} - \bar{x}^{(1)} - \hat{\boldsymbol{\beta}}\left(x_\alpha^{(2)} - \bar{x}^{(2)}\right)\right]'$$
$$= A_{11} - \hat{\boldsymbol{\beta}}A_{22}\hat{\boldsymbol{\beta}}'.$$

The vector $x_\alpha^{(1)} - \bar{x}^{(1)} - \hat{\boldsymbol{\beta}}(x_\alpha^{(2)} - \bar{x}^{(2)})$ is the residual of $x_\alpha^{(1)}$ from its regression on $x_\alpha^{(2)}$ and 1. The partial correlations are simple correlations between these residuals. The definition can be used also when the distributions involved are not normal.

Two geometric interpretations of the above theory can be given. In p-dimensional space $x_1,\ldots,x_N$ represent N points. The sample regression function

$$(12) \qquad x^{(1)} = \bar{x}^{(1)} + \hat{\boldsymbol{\beta}}(x^{(2)} - \bar{x}^{(2)})$$

is a $(p-q)$-dimensional hyperplane which is the intersection of q $(p-1)$-dimensional hyperplanes,

$$(13) \qquad x_i = \bar{x}_i + \sum_{j=q+1}^{p}\hat{\beta}_{ij}(x_j - \bar{x}_j), \qquad i = 1,\ldots,q,$$

where x_i, x_j are running variables. Here $\hat{\beta}_{ij}$ is an element of $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\Sigma}}_{12}\hat{\boldsymbol{\Sigma}}_{22}^{-1} = A_{12}A_{22}^{-1}$. The ith row of $\hat{\boldsymbol{\beta}}$ is $(\hat{\beta}_{i,q+1},\ldots,\hat{\beta}_{ip})$. Each right-hand side of (13) is the least squares *regression* function of x_i on $x_{q+1},\ldots,x_p$; that is, if we project the points $x_1,\ldots,x_N$ on the coordinate hyperplane of $x_i, x_{q+1},\ldots,x_p$,

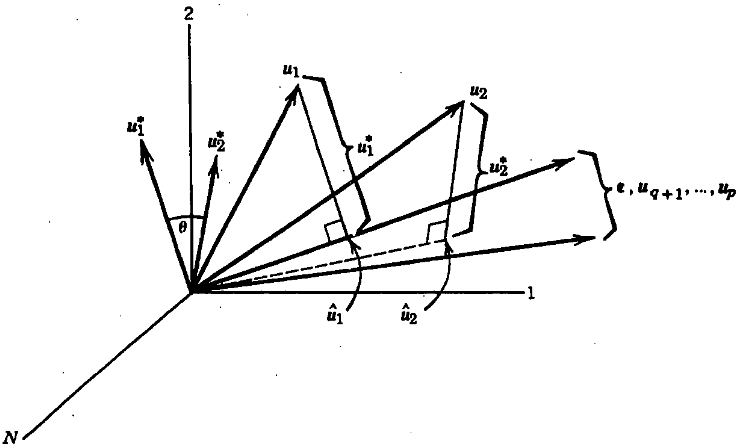

Figure 4.4

then (13) is the regression plane. The point with coordinates

$$x_i = \bar{x}_i + \sum_{j=q+1}^{p} \hat{\beta}_{ij}(x_{j\alpha} - \bar{x}_j), \qquad i = 1, \ldots, q, \tag{14}$$

$$x_j = x_{j\alpha}, \qquad j = q+1, \ldots, p,$$

is on the hyperplane (13). The difference in the ith coordinate of x_α and the point (14) is $y_{i\alpha} = x_{i\alpha} - [\bar{x}_i + \Sigma_{j=q+1}^{p} \hat{\beta}_{ij}(x_{j\alpha} - \bar{x}_j)]$ for $i = 1, \ldots, q$ and 0 for the other coordinates. Let $y'_\alpha = (y_{1\alpha}, \ldots, y_{q\alpha})$. These points can be represented as N points in a q-dimensional space. Then $A_{11\cdot 2} = \Sigma_{\alpha=1}^{N} y_\alpha y'_\alpha$.

We can also interpret the sample as p points in N-space (Figure 4.4). Let $u_j = (x_{j1}, \ldots, x_{jN})'$ be the jth point, and let $\varepsilon = (1, \ldots, 1)'$ be another point. The point with coordinates $\bar{x}_i, \ldots, \bar{x}_i$ is $\bar{x}_i \varepsilon$. The projection of u_i on the hyperplane spanned by $u_{q+1}, \ldots, u_p, \varepsilon$ is

$$\hat{u}_i = \bar{x}_i \varepsilon + \sum_{j=q+1}^{p} \hat{\beta}_{ij}(u_j - \bar{x}_j \varepsilon); \tag{15}$$

this is the point on the hyperplane that is at a minimum distance from u_i. Let u_i^* be the vector from $\hat{u}_i$ to u_i, that is, $u_i - \hat{u}_i$, or, equivalently, this vector translated so that one endpoint is at the origin. The set of vectors $u_1^*, \ldots, u_q^*$ are the projections of $u_1, \ldots, u_q$ on the hyperplane orthogonal to

$u_{q+1}, \ldots, u_p, \varepsilon$. Then $u_i^{*\prime} u_i^* = a_{ii\cdot q+1,\ldots,p}$, the length squared of u_i^* (i.e., the square of the distance of u from $\hat{u}_i$). Then $u_i^{*\prime} u_j^* / \sqrt{u_i^{*\prime} u_i^* u_j^{*\prime} u_j^*} = r_{ij\cdot q+1,\ldots,p}$ is the cosine of the angle between u_i^* and u_j^*.

As an example of the use of partial correlations we consider some data [Hooker (1907)] on yield of hay (X_1) in hundredweights per acrc, spring rainfall (X_2) in inches, and accumulated temperature above 42°F in the spring (X_3) for an English area over 20 years. The estimates of μ_i, σ_i $(= \sqrt{\sigma_{ii}})$, and ρ_{ij} are

$$\hat{\mu} = \bar{x} = \begin{pmatrix} 28.02 \\ 4.91 \\ 594 \end{pmatrix},$$

$$\begin{pmatrix} \hat{\sigma}_1 \\ \hat{\sigma}_2 \\ \hat{\sigma}_3 \end{pmatrix} = \begin{pmatrix} 4.42 \\ 1.10 \\ 85 \end{pmatrix}, \tag{16}$$

$$\begin{pmatrix} 1 & \hat{\rho}_{12} & \hat{\rho}_{13} \\ \hat{\rho}_{21} & 1 & \hat{\rho}_{23} \\ \hat{\rho}_{31} & \hat{\rho}_{32} & 1 \end{pmatrix} = \begin{pmatrix} 1.00 & 0.80 & -0.40 \\ 0.80 & 1.00 & -0.56 \\ -0.40 & -0.56 & 1.00 \end{pmatrix}.$$

From the correlations we observe that yield and rainfall are positively related, yield and temperature are negatively related, and rainfall and temperature are negatively related. What interpretation is to be given to the apparent negative relation between yield and temperature? Does high temperature tend to cause low yield, or is high temperature associated with low rainfall and hence with low yield? To answer this question we consider the correlation between yield and temperature when rainfall is held fixed; that is, we use the data given above to estimate the partial correlation between X_1 and X_3 with X_2 held fixed. It is†

$$r_{13\cdot 2} = \frac{\hat{\sigma}_{13\cdot 2}}{\sqrt{\hat{\sigma}_{11\cdot 2}\hat{\sigma}_{33\cdot 2}}} = 0.097. \tag{17}$$

Thus, if the effect of rainfall is removed, yield and temperature are positively correlated. The conclusion is that both high raninfall and high temperature increase hay yield, but in most years high rainfall occurs with low temperature and vice versa.

†We compute with $\hat{\Sigma}$ as if it were Σ.

4.3.2. The Distribution of the Sample Partial Correlation Coefficient

In order to test a hypothesis about a population partial correlation coefficient we want the distribution of the sample partial correlation coefficient. The partial correlations are computed from $A_{11\cdot 2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$ (as indicated in Theorem 4.3.1) in the same way that correlations are computed from A. To obtain the distribution of a simple correlation we showed that A was distributed as $\sum_{\alpha=1}^{N-1} \mathbf{Z}_\alpha \mathbf{Z}'_\alpha$, where $\mathbf{Z}_1, \dots, \mathbf{Z}_{N-1}$ are distributed independently according to $N(\mathbf{0}, \mathbf{\Sigma})$ and independent of $\bar{X}$ (Theorem 3.3.2). Here we want to show that $A_{11\cdot 2}$ is distributed as $\sum_{\alpha=1}^{N-1-(p-q)} U_\alpha U'_\alpha$, where $U_1, \dots, U_{N-1-(p-q)}$ are distributed independently according to $N(\mathbf{0}, \mathbf{\Sigma}_{11\cdot 2})$ and independently of $\hat{\boldsymbol{\beta}}$. The distribution of a partial correlation coefficient will follow from the characterization of the distribution of $A_{11\cdot 2}$. We state the theorem in a general form; it will be used in Chapter 8, where we treat regression in detail. The following corollary applies it to $A_{11\cdot 2}$, expressed in terms of residuals.

Theorem 4.3.3. *Suppose $Y_1, \dots, Y_m$ are independent with Y_α distributed according to $N(\Gamma \mathbf{w}_\alpha, \mathbf{\Phi})$, where $\mathbf{w}_\alpha$ is an r-component vector. Let $\boldsymbol{H} = \sum_{\alpha=1}^m \mathbf{w}_\alpha \mathbf{w}'_\alpha$, assumed nonsingular, $\boldsymbol{G} = \sum_{\alpha=1}^m Y_\alpha \mathbf{w}'_\alpha \boldsymbol{H}^{-1}$, and*

$$C = \sum_{\alpha=1}^{m} (Y_\alpha - G\mathbf{w}_\alpha)(Y_\alpha - G\mathbf{w}_\alpha)' = \sum_{\alpha=1}^{m} Y_\alpha Y'_\alpha - GHG'. \tag{18}$$

Then C is distributed as $\sum_{\alpha=1}^{m-r} U_\alpha U'_\alpha$, where $U_1, \dots, U_{m-r}$ are independently distributed according to $N(\mathbf{0}, \mathbf{\Phi})$ and independently of $\boldsymbol{G}$.

Proof. The rows of $Y = (Y_1, \dots, Y_m)$ are random vectors in an m-dimensional space, and the rows of $W = (\mathbf{w}_1, \dots, \mathbf{w}_m)$ are fixed vectors in that space. The idea of the proof is to rotate coordinate axes so that the last r axes are in the space spanned by the rows of W. Let $E_2 = FW$, where F is a square matrix such that $FHF' = I$. Then

$$\begin{aligned} E_2 E'_2 = FWW'F' &= F \sum_{\alpha=1}^{m} \mathbf{w}_\alpha \mathbf{w}'_\alpha F' \\ &= FHF' = I. \end{aligned} \tag{19}$$

Thus the m-component rows of E_2 are orthogonal and of unit length. It is possible to find an $(m-r) \times m$ matrix E_1 such that

$$E = \begin{pmatrix} E_1 \\ E_2 \end{pmatrix} \tag{20}$$

is orthogonal. (See Appendix, Lemma A.4.2.) Now let $U = YE'$ (i.e., $U_\alpha = \sum_{\beta=1}^m e_{\alpha\beta} Y_\beta$). By Theorem 3.3.1 the columns of $U = (U_1, \dots, U_m)$ are independently and normally distributed, each with covariance matrix $\mathbf{\Phi}$. The means are given by

$$\mathscr{E}U = \mathscr{E}YE' = \Gamma WE' \tag{21}$$
$$= \Gamma F^{-1} E_2 (E_1' \quad E_2')$$
$$= (\mathbf{0} \quad \Gamma F^{-1})$$

by orthogonality of E. To complete the proof we need to show that C transforms to $\sum_{\alpha=1}^{m-r} U_\alpha U_\alpha'$. We have

$$\sum_{\alpha=1}^{m} Y_\alpha Y_\alpha' = YY' = UEE'U' = UU' = \sum_{\alpha=1}^{m} U_\alpha U_\alpha'. \tag{22}$$

Note that

$$G = YW'H^{-1} = UEE_2'(F^{-1})'F'F \tag{23}$$
$$= U \begin{pmatrix} E_1 \\ E_2 \end{pmatrix} E_2' F$$
$$= U \begin{pmatrix} \mathbf{0} \\ I \end{pmatrix} F = U^{(2)} F,$$

where $U^{(2)} = (U_{m-r+1}, \dots, U_m)$. Then

$$GHG' = U^{(2)} FHF' U^{(2)\prime} = U^{(2)} U^{(2)\prime} = \sum_{\alpha=m-r+1}^{m} U_\alpha U_\alpha'. \tag{24}$$

Thus C is

$$\sum_{\alpha=1}^{m} Y_\alpha Y_\alpha' - GHG' = \sum_{\alpha=1}^{m} U_\alpha U_\alpha' - \sum_{\alpha=m-r+1}^{m} U_\alpha U_\alpha' = \sum_{\alpha=1}^{m-r} U_\alpha U_\alpha'. \tag{25}$$

This proves the theorem. ■

It follows from the above considerations that when $\Gamma = \mathbf{0}$, the $\mathscr{E}U = \mathbf{0}$, and we obtain the following:

Corollary 4.3.1. *If* $\Gamma = \mathbf{0}$, *the matrix* GHG' *defined in Theorem* 4.3.3 *is distributed as* $\sum_{\alpha=m-r+1}^{m} U_\alpha U_\alpha'$, *where* $U_{m-r+1}, \dots, U_m$ *are independently distributed, each according to* $N(\mathbf{0}, \mathbf{\Phi})$.

We now find the distribution of $A_{11\cdot2}$ in the same form. It was shown in Theorem 3.3.1 that A is distributed as $\sum_{\alpha=1}^{N-1} Z_\alpha Z'_\alpha$, where $Z_1, \ldots, Z_{N-1}$ are independent, each with distribution $N(\mathbf{0}, \mathbf{\Sigma})$. Let Z_α be partitioned into two subvectors of q and $p-q$ components, respectively:

$$(26) \qquad Z_\alpha = \begin{pmatrix} Z_\alpha^{(1)} \\ Z_\alpha^{(2)} \end{pmatrix}.$$

Then $A_{ij} = \sum_{\alpha=1}^{N} Z_\alpha^{(i)} Z_\alpha^{(j)\prime}$. By Lemma 4.2.1, conditionally on $Z_1^{(2)} = z_1^{(2)}, \ldots, Z_{N-1}^{(2)} = z_{N-1}^{(2)}$, the random vectors $Z_1^{(1)}, \ldots, Z_{N-1}^{(1)}$ are independently distributed, with $Z_\alpha^{(1)}$ distributed according to $N(\boldsymbol{\beta} z_\alpha^{(2)}, \mathbf{\Sigma}_{11\cdot2})$, where $\boldsymbol{\beta} = \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}$ and $\mathbf{\Sigma}_{11\cdot2} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$. Now we apply Theorem 4.3.3 with $Z_\alpha^{(1)} = Y_\alpha$, $z_\alpha^{(2)} = w_\alpha$, $N-1 = m$, $p-q = r$, $\boldsymbol{\beta} = \mathbf{\Gamma}$, $\mathbf{\Sigma}_{11\cdot2} = \mathbf{\Phi}$, $A_{11} = \sum_{\alpha=1}^{N-1} Y_\alpha Y'_\alpha$, $A_{12}A_{22}^{-1} = G$, $A_{22} = H$. We find that the conditional distribution of $A_{11} - (A_{12}A_{22}^{-1})A_{22}(A_{22}^{-1}A'_{12}) = A_{11\cdot2}$ given $Z_\alpha^{(2)} = z_\alpha^{(2)}$, $\alpha = 1, \ldots, N-1$, is that of $\sum_{\alpha=1}^{N-1-(p-q)} U_\alpha U'_\alpha$, where $U_1, \ldots, U_{N-1-(p-q)}$ are independent, each with distribution $N(\mathbf{0}, \mathbf{\Sigma}_{11\cdot2})$. Since this distribution does not depend on $\{z_\alpha^{(2)}\}$, we obtain the following theorem:

Theorem 4.3.4. *The matrix* $A_{11\cdot2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$ *is distributed as* $\sum_{\alpha=1}^{N-1-(p-q)} U_\alpha U'_\alpha$, *where* $U_1, \ldots, U_{N-1-(p-q)}$ *are independently distributed, each according to* $N(\mathbf{0}, \mathbf{\Sigma}_{11\cdot2})$, *and independently of* A_{12} *and* A_{22}.

Corollary 4.3.2. *If* $\mathbf{\Sigma}_{12} = \mathbf{0}$ $(\boldsymbol{\beta} = \mathbf{0})$, *then* $A_{11\cdot2}$ *is distributed as* $\sum_{\alpha=1}^{N-1-(p-q)} U_\alpha U'_\alpha$ *and* $A_{12}A_{22}^{-1}A_{21}$ *is distributed as* $\sum_{\alpha=N-(p-q)}^{N-1} U_\alpha U'_\alpha$, *where* $U_1, \ldots, U_{N-1}$ *are independently distributed, each according to* $N(\mathbf{0}, \mathbf{\Sigma}_{11\cdot2})$.

Now it follows that the distribution of $r_{ij\cdot q+1,\ldots,p}$ based on N observations is the same as that of a simple correlation coefficient based on $N-(p-q)$ observations with a corresponding population correlation value of $\rho_{ij\cdot q+1,\ldots,p}$.

Theorem 4.3.5. *If the* cdf *of* r_{ij} *based on a sample of N from a normal distribution with correlation* ρ_{ij} *is denoted by* $F(r|N, \rho_{ij})$, *then the* cdf *of the sample partial correlation* $r_{ij\cdot q+1,\ldots,p}$ *based on a sample of N from a normal distribution with partial correlation coefficient* $\rho_{ij\cdot q+1,\ldots,p}$ *is* $F[r|N-(p-q), \rho_{ij\cdot q+1,\ldots,p}]$.

This distribution was derived by Fisher (1924).

4.3.3. Tests of Hypotheses and Confidence Regions for Partial Correlation Coefficients

Since the distribution of a sample partial correlation $r_{ij\cdot q+1,\ldots,p}$ based on a sample of N from a distribution with population correlation $\rho_{ij\cdot q+1,\ldots,p}$

equal to a certain value, ρ, say, is the same as the distribution of a simple correlation r based on a sample of size $N-(p-q)$ from a distribution with the corresponding population correlation of ρ, all statistical inference procedures for the simple population correlation can be used for the partial correlation. The procedure for the partial correlation is exactly the same except that N is replaced by $N-(p-q)$. To illustrate this rule we give two examples.

Example 1. Suppose that on the basis of a sample of size N we wish to obtain a confidence interval for $\rho_{ij\cdot q+1,\dots,p}$. The sample partial correlation is $r_{ij\cdot q+1,\dots,p}$. The procedure is to use David's charts for $N-(p-q)$. In the example at the end of Section 4.3.1, we might want to find a confidence interval for $\rho_{12\cdot 3}$ with confidence coefficient 0.95. The sample partial correlation is $r_{12\cdot 3} = 0.759$. We use the chart (or table) for $N-(p-q) = 20-1 = 19$. The interval is $0.50 < \rho_{12\cdot 3} < 0.88$.

Example 2. Suppose that on the basis of a sample of size N we use Fisher's z for an approximate significance test of $\rho_{ij\cdot q+1,\dots,p} = \rho_0$ against two-sided alternatives. We let

$$z = \tfrac{1}{2}\log\frac{1+r_{ij\cdot q+1,\dots,p}}{1-r_{ij\cdot q+1,\dots,p}}, \tag{27}$$
$$\zeta_0 = \tfrac{1}{2}\log\frac{1+\rho_0}{1-\rho_0}.$$

Then $\sqrt{N-(p-q)-3}\,(z-\zeta_0)$ is compared with the significance points of the standardized normal distribution. In the example at the end of Section 4.3.1, we might wish to test the hypothesis $\rho_{13\cdot 2} = 0$ at the 0.05 level. Then $\zeta_0 = 0$ and $\sqrt{20-1-3}\,(0.0973) = 0.3892$. This value is clearly nonsignificant ($|0.3892| < 1.96$), and hence the data do not indicate rejection of the null hypothesis.

To answer the question whether two variables x_1 and x_2 are related when both may be related to a vector $x^{(2)} = (x_3, \dots, x_p)$, two approaches may be used. One is to consider the regression of x_1 on x_2 and $x^{(2)}$ and test whether the regression of x_1 on x_2 is 0. Another is to test whether $\rho_{12\cdot 3,\dots,p} = 0$. Problems 4.43–4.47 show that these approaches lead to exactly the same test.

4.4. THE MULTIPLE CORRELATION COEFFICIENT

4.4.1. Estimation of the Multiple Correlation Coefficient

The population multiple correlation between one variate and a set of variates was defined in Section 2.5. For the sake of convenience in this section we shall treat the case of the multiple correlation between X_1 and the vector

$X^{(2)} = (X_2, \ldots, X_p)'$; we shall not need subscripts on R. The variables can always be numbered so that the desired multiple correlation is this one (any irrelevant variables being omitted). Then the multiple correlation in the population is

$$(1) \qquad \bar{R} = \frac{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}}{\sqrt{\sigma_{11}\boldsymbol{\beta}'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}}} = \sqrt{\frac{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}}{\sigma_{11}}} = \sqrt{\frac{\boldsymbol{\sigma}'_{(1)}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{(1)}}{\sigma_{11}}},$$

where $\boldsymbol{\beta}$, $\boldsymbol{\sigma}_{(1)}$, and $\boldsymbol{\Sigma}_{22}$ are defined by

$$(2) \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma}'_{(1)} \\ \boldsymbol{\sigma}_{(1)} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

$$(3) \qquad \boldsymbol{\beta} = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{(1)}.$$

Given a sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ $(N > p)$, we estimate $\boldsymbol{\Sigma}$ by $S = [N/(N-1)]\hat{\boldsymbol{\Sigma}}$ or

$$(4) \qquad \hat{\boldsymbol{\Sigma}} = \frac{1}{N}A = \frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})' = \begin{pmatrix} \hat{\sigma}_{11} & \hat{\boldsymbol{\sigma}}'_{(1)} \\ \hat{\boldsymbol{\sigma}}_{(1)} & \hat{\boldsymbol{\Sigma}}_{22} \end{pmatrix},$$

and we estimate $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\Sigma}}_{22}^{-1}\hat{\boldsymbol{\sigma}}_{(1)} = A_{22}^{-1}\boldsymbol{a}_{(1)}$. We define the *sample multiple correlation coefficient* by

$$(5) \qquad R = \sqrt{\frac{\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\Sigma}}_{22}\hat{\boldsymbol{\beta}}}{\hat{\sigma}_{11}}} = \sqrt{\frac{\hat{\boldsymbol{\sigma}}'_{(1)}\hat{\boldsymbol{\Sigma}}_{22}^{-1}\hat{\boldsymbol{\sigma}}_{(1)}}{\hat{\sigma}_{11}}} = \sqrt{\frac{\boldsymbol{a}'_{(1)}A_{22}^{-1}\boldsymbol{a}_{(1)}}{a_{11}}}.$$

That this is the maximum likelihood estimator of $\bar{R}$ is justified by Corollary 3.2.1, since we can define $\bar{R}, \boldsymbol{\sigma}_{(1)}, \boldsymbol{\Sigma}_{22}$ as a one-to-one transformation of $\boldsymbol{\Sigma}$. Another expression for R [see (16) of Section 2.5] follows from

$$(6) \qquad 1 - R^2 = \frac{|\hat{\boldsymbol{\Sigma}}|}{\hat{\sigma}_{11}|\hat{\boldsymbol{\Sigma}}_{22}|} = \frac{|A|}{a_{11}|A_{22}|}.$$

The quantities R and $\hat{\boldsymbol{\beta}}$ have properties in the sample that are similar to those $\bar{R}$ and $\boldsymbol{\beta}$ have in the population. We have analogs of Theorems 2.5.2, 2.5.3, and 2.5.4. Let $\hat{x}_{1\alpha} = \bar{x}_1 + \hat{\boldsymbol{\beta}}'(\boldsymbol{x}_\alpha^{(2)} - \bar{\boldsymbol{x}}^{(2)})$, and $x_{1\alpha}^* = x_{1\alpha} - \hat{x}_{1\alpha}$ be the residual.

Theorem 4.4.1. *The residuals $x_{1\alpha}^*$ are uncorrelated in the sample with the components of $\boldsymbol{x}_\alpha^{(2)}$, $\alpha = 1, \ldots, N$. For every vector $\boldsymbol{a}$*

$$(7) \qquad \sum_{\alpha=1}^{N}\left[x_{1\alpha} - \bar{x}_1 - \hat{\boldsymbol{\beta}}'\left(\boldsymbol{x}_\alpha^{(2)} - \bar{\boldsymbol{x}}^{(2)}\right)\right]^2 \leq \sum_{\alpha=1}^{N}\left[x_{1\alpha} - \bar{x}_1 - \boldsymbol{a}'\left(\boldsymbol{x}_\alpha^{(2)} - \bar{\boldsymbol{x}}^{(2)}\right)\right]^2.$$

The sample correlation between $x_{1\alpha}$ and $a'x_\alpha^{(2)}$, $\alpha = 1, \dots, N$, is maximized for $a = \hat{\beta}$, and that maximum correlation is R.

Proof. Since the sample mean of the residuals is 0, the vector of sample covariances between $x^*_{1\alpha}$ and $x_\alpha^{(2)}$ is proportional to

$$(8) \qquad \sum_{\alpha=1}^{N} \left[(x_{1\alpha} - \bar{x}_1) - \hat{\beta}'\left(x_\alpha^{(2)} - \bar{x}^{(2)}\right)\right]\left(x_\alpha^{(2)} - \bar{x}^{(2)}\right)' = a'_{(1)} - \hat{\beta}'A_{22} = \mathbf{0}.$$

The right-hand side of (7) can be written as the left-hand side plus

$$(9) \qquad \sum_{\alpha=1}^{N} \left[(\hat{\beta} - a)'\left(x_\alpha^{(2)} - \bar{x}^{(2)}\right)\right]^2 = (\hat{\beta} - a)' \sum_{\alpha=1}^{N} \left(x_\alpha^{(2)} - \bar{x}^{(2)}\right)\left(x_\alpha^{(2)} - \bar{x}^{(2)}\right)'(\hat{\beta} - a),$$

which is 0 if and only if $a = \hat{\beta}$. To prove the third assertion we consider the vector a for which $\sum_{\alpha=1}^{N}[a'(x_\alpha^{(2)} - \bar{x}^{(2)})]^2 = \sum_{\alpha=1}^{N}[\hat{\beta}'(x_\alpha^{(2)} - \bar{x}^{(2)})]^2$, since the correlation is unchanged when the linear function is mulitplied by a positive constant. From (7) we obtain

$$(10) \qquad a_{11} - 2\sum_{\alpha=1}^{N} (x_{1\alpha} - \bar{x}_1)\hat{\beta}'\left(x_\alpha^{(2)} - \bar{x}^{(2)}\right) + \sum_{\alpha=1}^{N} \left[\hat{\beta}'\left(x_\alpha^{(2)} - \bar{x}^{(2)}\right)\right]^2$$
$$\leq a_{11} - 2\sum_{\alpha=1}^{N} (x_{1\alpha} - \bar{x}_1)a'\left(x_\alpha^{(2)} - \bar{x}^{(2)}\right) + \sum_{\alpha=1}^{N} \left[a'\left(x_\alpha^{(2)} - \bar{x}^{(2)}\right)\right]^2,$$

from which we deduce

$$(11) \qquad \frac{\sum_{\alpha=1}^{N}(x_{1\alpha} - \bar{x}_1)\left(x_\alpha^{(2)} - \bar{x}^{(2)}\right)'a}{\sqrt{a_{11}}\sqrt{\sum_{\alpha=1}^{N}\left[a'\left(x_\alpha^{(2)} - \bar{x}^{(2)}\right)\right]^2}} \leq \frac{\sum_{\alpha=1}^{N}(x_{1\alpha} - \bar{x}_1)\left(x_\alpha^{(2)} - \bar{x}^{(2)}\right)'\hat{\beta}}{\sqrt{a_{11}}\sqrt{\sum_{\alpha=1}^{N}\left[\hat{\beta}'\left(x_\alpha^{(2)} - \bar{x}^{(2)}\right)\right]^2}}$$
$$= \frac{a'_{(1)}\hat{\beta}}{\sqrt{a_{11}}\sqrt{\hat{\beta}'A_{22}\hat{\beta}}},$$

which is (5). ∎

Thus $\bar{x}_1 + \hat{\beta}'(x_\alpha^{(2)} - \bar{x}^{(2)})$ is the best linear predictor of $x_{1\alpha}$ in the sample, and $\hat{\beta}'x_\alpha^{(2)}$ is the linear function of $x_\alpha^{(2)}$ that has maximum sample correlation with $x_{1\alpha}$. The minimum sum of squares of deviations [the left-hand side of (7)] is

$$\sum_{\alpha=1}^{N}\left[(x_{1\alpha}-\bar{x}_1)-\hat{\boldsymbol{\beta}}'\left(\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)}\right)\right]^2 = a_{11}-\hat{\boldsymbol{\beta}}'A_{22}\hat{\boldsymbol{\beta}} \tag{12}$$

$$= a_{11}-\boldsymbol{a}_{(1)}'A_{22}^{-1}\boldsymbol{a}_{(1)}$$

$$= a_{11\cdot 2}$$

as defined in Section 4.3 with $q=1$. The maximum likelihood estimator of $\sigma_{11\cdot 2}$ is $\hat{\sigma}_{11\cdot 2}=a_{11\cdot 2}/N$. It follows that

$$\hat{\sigma}_{11\cdot 2}=(1-R^2)\,\hat{\sigma}_{11}. \tag{13}$$

Thus $1-R^2$ measures the proportional reduction in the variance by using residuals. We can say that R^2 is the fraction of the variance explained by $\boldsymbol{x}^{(2)}$. The larger R^2 is, the more the variance is decreased by use of the explanatory variables in $\boldsymbol{x}^{(2)}$.

In p-dimensional space $\boldsymbol{x}_1,\dots,\boldsymbol{x}_N$ represent N points. The sample regression function $x_1=\bar{x}_1+\hat{\boldsymbol{\beta}}'(\boldsymbol{x}^{(2)}-\bar{\boldsymbol{x}}^{(2)})$ is the $(p-1)$-dimensional hyperplane that minimizes the squared deviations of the points from the hyperplane, the deviations being calculated in the x_1-direction. The hyperplane goes through the point $\bar{\boldsymbol{x}}$.

In N-dimensional space the rows of $(\boldsymbol{x}_1,\dots,\boldsymbol{x}_N)$ represent p points. The N-component vector with αth component $x_{i\alpha}-\bar{x}_i$ is the projection of the vector with αth component $x_{i\alpha}$ on the plane orthogonal to the equiangular line. We have p such vectors; $\boldsymbol{a}'(\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)})$ is the αth component of a vector in the hyperplane spanned by the last $p-1$ vectors. Since the right-hand side of (7) is the squared distance between the first vector and the linear combination of the last $p-1$ vectors, $\hat{\boldsymbol{\beta}}'(\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)})$ is a component of the vector which minimizes this squared distance. The interpretation of (8) is that the vector with αth component $(x_{1\alpha}-\bar{x}_1)-\hat{\boldsymbol{\beta}}'(\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)})$ is orthogonal to each of the last $p-1$ vectors. Thus the vector with αth component $\hat{\boldsymbol{\beta}}'(\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)})$ is the projection of the first vector on the hyperplane. See Figure 4.5. The length squared of the projection vector is

$$\sum_{\alpha=1}^{N}\left[\hat{\boldsymbol{\beta}}'\left(\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)}\right)\right]^2=\hat{\boldsymbol{\beta}}'A_{22}\hat{\boldsymbol{\beta}}=\boldsymbol{a}_{(1)}'A_{22}^{-1}\boldsymbol{a}_{(1)}, \tag{14}$$

and the length squared of the first vector is $\sum_{\alpha=1}^{N}(x_{1\alpha}-\bar{x}_1)^2=a_{11}$. Thus R is the cosine of the angle between the first vector and its projection.

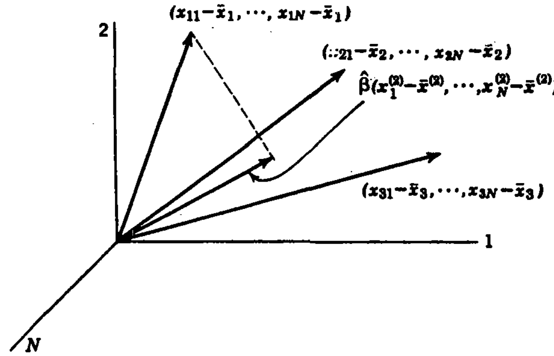

Figure 4.5

In Section 3.2 we saw that the simple correlation coefficient is the cosine of the angle between the two vectors involved (in the plane orthogonal to the equiangular line). The property of R that it is the maximum correlation between $x_{1\alpha}$ and linear combinations of the components of $x_\alpha^{(2)}$ corresponds to the geometric property that R is the cosine of the smallest angle between the vector with components $x_{1\alpha} - \bar{x}_1$ and a vector in the hyperplane spanned by the other $p-1$ vectors.

The geometric interpretations are in terms of the vectors in the $(N-1)$-dimensional hyperplane orthogonal to the equiangular line. It was shown in Section 3.3 that the vector $(x_{i1} - \bar{x}_i, \dots, x_{iN} - \bar{x}_i)$ in this hyperplane can be designated as $(z_{i1}, \dots, z_{i,N-1})$, where the $z_{i\alpha}$ are the coordinates referred to an $(N-1)$-dimensional coordinate system in the hyperplane. It was shown that the new coordinates are obtained from the old by the transformation $z_{i\alpha} = \sum_{\beta=1}^N b_{\alpha\beta} x_{i\beta}$, $\alpha = 1, \dots, N$, where $\boldsymbol{B} = (b_{\alpha\beta})$ is an orthogonal matrix with last row $(1/\sqrt{N}, \dots, 1/\sqrt{N})$. Then

$$a_{ij} = \sum_{\alpha=1}^{N} (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j) = \sum_{\alpha=1}^{N-1} z_{i\alpha} z_{j\alpha}. \tag{15}$$

It will be convenient to refer to the multiple correlation defined in terms of $z_{i\alpha}$ as the *multiple correlation without subtracting the means*.

The population multiple correlation $\bar{R}$ is essentially the only function of the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ that is invariant under changes of location, changes of scale of X_1, and nonsingular linear transformations of $X^{(2)}$, that is, transformations $X_1^* = cX_1 + d$, $X^{(2)*} = CX^{(2)} + d$. Similarly, the sample multiple correlation coefficient R is essentially the only function of $\bar{x}$ and $\hat{\Sigma}$, the sufficient set of statistics for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, that is invariant under these transformations. Just as the simple correlation r is a measure of association between two scalar variables in a sample, the multiple correlation R is a measure of association between a scalar variable and a vector variable in a sample.

4.4.2. Distribution of the Sample Multiple Correlation Coefficient When the Population Multiple Correlation Coefficient Is Zero

From (5) we have

$$R^2 = \frac{\boldsymbol{a}'_{(1)} A_{22}^{-1} \boldsymbol{a}_{(1)}}{a_{11}}; \tag{16}$$

then

$$1 - R^2 = 1 - \frac{\boldsymbol{a}'_{(1)} A_{22}^{-1} \boldsymbol{a}_{(1)}}{a_{11}} = \frac{a_{11} - \boldsymbol{a}'_{(1)} A_{22}^{-1} \boldsymbol{a}_{(1)}}{a_{11}} = \frac{a_{11\cdot 2}}{a_{11}}, \tag{17}$$

and

$$\frac{R^2}{1-R^2} = \frac{\boldsymbol{a}'_{(1)} A_{22}^{-1} \boldsymbol{a}_{(1)}}{a_{11\cdot 2}}. \tag{18}$$

For $q = 1$, Corollary 4.3.2 states that when $\boldsymbol{\beta} = \mathbf{0}$, that is, when $\overline{R} = 0$, $a_{11\cdot 2}$ is distributed as $\sum_{\alpha=1}^{N-p} V_\alpha^2$ and $\boldsymbol{a}'_{(1)} A_{22}^{-1} \boldsymbol{a}_{(1)}$ is distributed as $\sum_{\alpha=N-p+1}^{N-1} V_\alpha^2$, where $V_1, \dots, V_{N-1}$ are independent, each with distribution $N(0, \sigma_{11\cdot 2})$. Then $a_{11\cdot 2}/\sigma_{11\cdot 2}$ and $\boldsymbol{a}'_{(1)} A_{22}^{-1} \boldsymbol{a}_{(1)}/\sigma_{11\cdot 2}$ are distributed independently as χ^2-variables with $N-p$ and $p-1$ degrees of freedom, respectively. Thus

$$\begin{aligned} \frac{R^2}{1-R^2} \cdot \frac{N-p}{p-1} &= \frac{\boldsymbol{a}'_{(1)} A_{22}^{-1} \boldsymbol{a}_{(1)}/\sigma_{11\cdot 2}}{a_{11\cdot 2}/\sigma_{11\cdot 2}} \cdot \frac{N-p}{p-1} \\ &= \frac{\chi_{p-1}^2}{\chi_{N-p}^2} \cdot \frac{N-p}{p-1} \\ &= F_{p-1,N-p} \end{aligned} \tag{19}$$

has the F-distribution with $p-1$ and $N-p$ degrees of freedom. The density of F is

(20)

$$\frac{\Gamma[\frac{1}{2}(N-1)]}{\Gamma[\frac{1}{2}(p-1)]\Gamma[\frac{1}{2}(N-p)]} \left(\frac{p-1}{N-p}\right)^{\frac{1}{2}(p-1)} f^{\frac{1}{2}(p-1)-1} \left(1 + \frac{p-1}{N-p} f\right)^{-\frac{1}{2}(N-1)}.$$

Thus the density of

$$R = \sqrt{\frac{\frac{p-1}{N-p}F_{p-1,N-p}}{1+\frac{p-1}{N-p}F_{p-1,N-p}}} \tag{21}$$

is

$$2\frac{\Gamma[\frac{1}{2}(N-1)]}{\Gamma[\frac{1}{2}(p-1)]\Gamma[\frac{1}{2}(N-p)]}R^{p-2}(1-R^2)^{\frac{1}{2}(N-p)-1}, \qquad 0 \le R \le 1. \tag{22}$$

Theorem 4.4.2. *Let R be the sample multiple correlation coefficient [defined by (5)] between X_1 and $X^{(2)\prime} = (X_2, \dots, X_p)$ based on a sample of N from $N(\mu, \Sigma)$. If $\bar{R} = 0$ [that is, if $(\sigma_{12}, \dots, \sigma_{1p})' = 0 = \beta$], then $[R^2/(1-R^2)]\cdot[(N-p)/(p-1)]$ is distributed as F with $p-1$ and $N-p$ degrees of freedom.*

It should be noticed that $p-1$ is the number of components of $X^{(2)}$ and that $N-p = N-(p-1)-1$. If the multiple correlation is between a component X_i and q other components, the numbers are q and $N-q-1$.

It might be observed that $R^2/(1-R^2)$ is the quantity that arises in regression (or least squares) theory for testing the hypothesis that the regression of X_1 on $X_2, \dots, X_p$ is zero.

If $\bar{R} \neq 0$, the distribution of R is much more difficult to derive. This distribution will be obtained in Section 4.4.3.

Now let us consider the statistical problem of testing the hypothesis $H: \bar{R} = 0$ on the basis of a sample of N from $N(\mu, \Sigma)$. [$\bar{R}$ is the population multiple correlation between X_1 and $(X_2, \dots, X_p)$.] Since $\bar{R} \ge 0$, the alternatives considered are $\bar{R} > 0$.

Let us derive the likelihood ratio test of this hypothesis. The likelihood function is

$$L(\mu^*, \Sigma^*) = \frac{1}{(2\pi)^{\frac{1}{2}pN}|\Sigma^*|^{\frac{1}{2}N}}\exp\left[-\frac{1}{2}\sum_{\alpha=1}^{N}(x_\alpha - \mu^*)'\Sigma^{*-1}(x_\alpha - \mu^*)\right]. \tag{23}$$

The observations are given; L is a function of the indeterminates μ^*, Σ^*. Let ω be the region in the parameter space Ω specified by the null hypothesis. The likelihood ratio criterion is

$$\lambda = \frac{\max_{\mu^*, \Sigma^* \in \omega} L(\mu^*, \Sigma^*)}{\max_{\mu^*, \Sigma^* \in \Omega} L(\mu^*, \Sigma^*)}. \tag{24}$$

Here Ω is the space of $\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*$ positive definite, and ω is the region in this space where $\bar{R} = \sqrt{\boldsymbol{\sigma}_{(1)}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{(1)}} / \sqrt{\sigma_{11}} = 0$, that is, where $\boldsymbol{\sigma}_{(1)}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{(1)} = 0$. Because $\boldsymbol{\Sigma}_{22}^{-1}$ is positive definite, this condition is equivalent to $\boldsymbol{\sigma}_{(1)} = \mathbf{0}$. The maximum of $L(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ over Ω occurs at $\boldsymbol{\mu}^* = \hat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}}$ and $\boldsymbol{\Sigma}^* = \hat{\boldsymbol{\Sigma}} = (1/N)A = (1/N)\sum_{\alpha=1}^N(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})'$ and is

$$\max_{\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^* \in \Omega} L(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*) = \frac{N^{-\frac{1}{2}pN} e^{\frac{1}{2}pN}}{(2\pi)^{\frac{1}{2}pN}|A|^{\frac{1}{2}N}}. \tag{25}$$

In ω the likelihood function is

$$\begin{aligned} L\left(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^* | \boldsymbol{\sigma}_{(1)}^* = \mathbf{0}\right) = {} & \frac{1}{(2\pi)^{\frac{1}{2}N}\sigma_{11}^{*\frac{1}{2}N}} \exp\left[-\frac{1}{2}\sum_{\alpha=1}^N (x_{1\alpha} - \mu_1^*)^2/\sigma_{11}^*\right] \\ & \cdot \frac{1}{(2\pi)^{\frac{1}{2}(p-1)N}|\boldsymbol{\Sigma}_{22}^*|^{\frac{1}{2}N}} \exp\left[-\frac{1}{2}\sum_{\alpha=1}^N \left(\boldsymbol{x}_\alpha^{(2)} - \boldsymbol{\mu}^{(2)*}\right)'\boldsymbol{\Sigma}_{22}^{*-1}\left(\boldsymbol{x}_\alpha^{(2)} - \boldsymbol{\mu}^{(2)*}\right)\right]. \end{aligned} \tag{26}$$

The first factor is maximized at $\mu_1^* = \hat{\mu}_1 = \bar{x}_1$ and $\sigma_{11}^* = \hat{\sigma}_{11}^* = (1/N)a_{11}$, and the second factor is maximized at $\boldsymbol{\mu}^{(2)*} = \hat{\boldsymbol{\mu}}^{(2)} = \bar{\boldsymbol{x}}^{(2)}$ and $\boldsymbol{\Sigma}_{22}^* = \hat{\boldsymbol{\Sigma}}_{22} = (1/N)A_{22}$. The value of the maximized function is

$$\max_{\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^* \in \omega} L(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*) = \frac{N^{\frac{1}{2}N} e^{-\frac{1}{2}N}}{(2\pi)^{\frac{1}{2}N} a_{11}^{\frac{1}{2}N}} \cdot \frac{N^{\frac{1}{2}(p-1)N} e^{-\frac{1}{2}(p-1)N}}{(2\pi)^{\frac{1}{2}(p-1)N}|A_{22}|^{\frac{1}{2}N}}. \tag{27}$$

Thus the likelihood ratio criterion is [see (6)]

$$\lambda = \frac{|A|^{\frac{1}{2}N}}{a_{11}^{\frac{1}{2}N}|A_{22}|^{\frac{1}{2}N}} = (1 - R^2)^{\frac{1}{2}N}. \tag{28}$$

The likelihood ratio test consists of the critical region $\lambda < \lambda_0$, where λ_0 is chosen so the probability of this inequality when $\bar{R} = 0$ is the significance level α. An equivalent test is

$$1 - \lambda^{2/N} = R^2 > 1 - \lambda_0^{2/N}. \tag{29}$$

Since $[R^2/(1 - R^2)][(N-p)/(p-1)]$ is a monotonic function of R, an equivalent test involves this ratio being larger than a constant. When $\bar{R} = 0$, this ratio has an $F_{p-1, N-p}$-distribution. Hence, the critical region is

$$\frac{R^2}{1 - R^2} \cdot \frac{N-p}{p-1} > F_{p-1, N-p}(\alpha), \tag{30}$$

where $F_{p-1, N-p}(\alpha)$ is the (upper) significance point corresponding to the α significance level.

Theorem 4.4.3. *Given a sample* $x_1, \dots, x_N$ *from* $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *the likelihood ratio test at significance level* α *for the hypothesis* $\bar{R} = 0$, *where* $\bar{R}$ *is the population multiple correlation coefficient between* X_1 *and* $(X_2, \dots, X_p)$, *is given by* (30), *where* R *is the sample multiple correlation coefficient defined by* (5).

As an example consider the data given at the end of Section 4.3.1. The sample multiple correlation coefficient is found from

$$(31)\qquad 1 - R^2 = \frac{\begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{23} \\ r_{32} & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1.00 & 0.80 & -0.40 \\ 0.80 & 1.00 & -0.56 \\ -0.40 & -0.56 & 1.00 \end{vmatrix}}{\begin{vmatrix} 1.00 & -0.56 \\ -0.56 & 1.00 \end{vmatrix}} = 0.357.$$

Thus R is 0.802. If we wish to test the hypothesis at the 0.01 level that hay yield is independent of spring rainfall and temperature, we compare the observed $[R^2/(1-R^2)][(20-3)/(3-1)] = 15.3$ with $F_{2,17}(0.01) = 6.11$ and find the result significant; that is, we reject the null hypothesis.

The test of independence between X_1 and $(X_2, \dots, X_p) = X^{(2)\prime}$ is equivalent to the test that if the regression of X_1 on $x^{(2)}$ (that is, the conditional expected value of X_1 given $X_2 = x_2, \dots, X_p = x_p$) is $\mu_1 + \boldsymbol{\beta}'(x^{(2)} - \boldsymbol{\mu}^{(2)})$, the vector of regression coefficients is $\mathbf{0}$. Here $\hat{\boldsymbol{\beta}} = A_{22}^{-1}a_{(1)}$ is the usual least squares estimate of $\boldsymbol{\beta}$ with expected value $\boldsymbol{\beta}$ and covariance matrix $\sigma_{11\cdot 2}A_{22}^{-1}$ (when the $X_\alpha^{(2)}$ are fixed), and $a_{11\cdot 2}/(N-p)$ is the usual estimate of $\sigma_{11\cdot 2}$. Thus [see (18)]

$$(32)\qquad \frac{R^2}{1-R^2} \cdot \frac{N-p}{p-1} = \frac{\hat{\boldsymbol{\beta}}' A_{22} \hat{\boldsymbol{\beta}}}{a_{11\cdot 2}} \cdot \frac{N-p}{p-1}$$

is the usual F-statistic for testing the hypothesis that the regression of X_1 on $x_2, \dots, x_p$ is 0. In this book we are primarily interested in the multiple correlation coefficient as a measure of association between one variable and a vector of variables when both are random. We shall not treat problems of univariate regression. In Chapter 8 we study regression when the dependent variable is a vector.

Adjusted Multiple Correlation Coefficient

The expression (17) is the ratio of $a_{11\cdot 2}$, the sum of squared deviations from the fitted regression, to a_{11}, the sum of squared deviations around the mean. To obtain unbiased estimators of σ_{11} when $\boldsymbol{\beta} = \mathbf{0}$ we would divide these quantities by their numbers of degrees of freedom, $N-p$ and $N-1$, respectively. Accordingly we can define an *adjusted multiple correlation coefficient* R^* by

$$(33)\qquad 1 - R^{*2} = \frac{a_{11\cdot 2}/(N-p)}{a_{11}/(N-1)} = \frac{N-1}{N-p}(1-R^2),$$

which is equivalent to

$$(34)\qquad R^{*2} = R^2 - \frac{p-1}{N-p}(1-R^2).$$

This quantity is smaller than R^2 (unless $p=1$ or $R^2=1$). A possible merit to it is that it takes account of p; the idea is that the larger p is relative to N, the greater the tendency of R^2 to be large by chance.

4.4.3. Distribution of the Sample Multiple Correlation Coefficient When the Population Multiple Correlation Coefficient Is Not Zero

In this subsection we shall find the distribution of R when the null hypothesis $\overline{R}=0$ is not true. We shall find that the distribution depends only on the population multiple correlation coefficient $\overline{R}$.

First let us consider the conditional distribution of $R^2/(1-R^2) = \boldsymbol{a}'_{(1)}A_{22}^{-1}\boldsymbol{a}_{(1)}/a_{11\cdot 2}$ given $\boldsymbol{Z}_\alpha^{(2)} = \boldsymbol{z}_\alpha^{(2)}$, $\alpha = 1,\dots,n$. Under these conditions $Z_{11},\dots,Z_{1n}$ are independently distributed, $Z_{1\alpha}$ according to $N(\boldsymbol{\beta}'\boldsymbol{z}_\alpha^{(2)}, \sigma_{11\cdot 2})$, where $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{(1)}$ and $\sigma_{11\cdot 2} = \sigma_{11} - \boldsymbol{\sigma}'_{(1)}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{(1)}$. The conditions are those of Theorem 4.3.3 with $Y_\alpha = Z_{1\alpha}$, $\boldsymbol{\Gamma} = \boldsymbol{\beta}'$, $\boldsymbol{w}_\alpha = \boldsymbol{z}_\alpha^{(2)}$, $r = p-1$, $\boldsymbol{\Phi} = \sigma_{11\cdot 2}$, $m=n$. Then $a_{11\cdot 2} = a_{11} - \boldsymbol{a}'_{(1)}A_{22}^{-1}\boldsymbol{a}_{(1)}$ corresponds to $\sum_{\alpha=1}^m Y_\alpha Y'_\alpha - \boldsymbol{GHG}'$, and $a_{11\cdot 2}/\sigma_{11\cdot 2}$ has a χ^2-distribution with $n-(p-1)$ degrees of freedom. $\boldsymbol{a}'_{(1)}A_{22}^{-1}\boldsymbol{a}_{(1)} = (A_{22}^{-1}\boldsymbol{a}_{(1)})'A_{22}(A_{22}^{-1}\boldsymbol{a}_{(1)})$ corresponds to $\boldsymbol{GHG}'$ and is distributed as $\sum_\alpha U_\alpha^2$, $\alpha = n-(p-1)+1,\dots,n$, where $\operatorname{Var}(U_\alpha) = \sigma_{11\cdot 2}$ and

$$(35)\qquad \mathscr{E}(U_{n-p+2},\dots,U_n) = \boldsymbol{\Gamma F}^{-1},$$

where $\boldsymbol{FHF}' = \boldsymbol{I}$ $[\boldsymbol{H} = \boldsymbol{F}^{-1}(\boldsymbol{F}')^{-1}]$. Then $\boldsymbol{a}'_{(1)}A_{22}^{-1}\boldsymbol{a}_{(1)}/\sigma_{11\cdot 2}$ is distributed as $\sum_\alpha(U_\alpha/\sqrt{\sigma_{11\cdot 2}})^2$, where $\operatorname{Var}(U_\alpha/\sqrt{\sigma_{11\cdot 2}}) = 1$ and

$$(36)\qquad \sum_{\alpha=n-p+2}^{n}\left(\frac{\mathscr{E}U_\alpha}{\sqrt{\sigma_{11\cdot 2}}}\right)^2 = \frac{1}{\sigma_{11\cdot 2}}\boldsymbol{\Gamma F}^{-1}(\boldsymbol{\Gamma F}^{-1})' = \frac{\boldsymbol{\Gamma H \Gamma}'}{\sigma_{11\cdot 2}}$$
$$= \frac{\boldsymbol{\beta}'A_{22}\boldsymbol{\beta}}{\sigma_{11\cdot 2}}.$$

Thus (conditionally) $\boldsymbol{a}'_{(1)}A_{22}^{-1}\boldsymbol{a}_{(1)}/\sigma_{11\cdot 2}$ has a noncentral χ^2-distribution with

$p-1$ degrees of freedom and noncentrality parameter $\beta' A_{22}\beta/\sigma_{11\cdot 2}$. (See Theorem 5.4.1.) We are led to the following theorem:

Theorem 4.4.4. *Let R be the sample multiple correlation coefficient between $X_{(1)}$ and $X^{(2)\prime} = (X_2, \dots, X_p)$ based on N observations $(x_{11}, x_1^{(2)}), \dots, (x_{1N}, x_N^{(2)})$. The conditional distribution of $[R^2/(1-R^2)][N-p)/(p-1)]$ given $x_\alpha^{(2)}$ fixed is noncentral F with $p-1$ and $N-p$ degrees of freedom and noncentrality parameter $\beta' A_{22}\beta/\sigma_{11\cdot 2}$.*

The conditional density (from Theorem 5.4.1) of $F = [R^2/(1-R^2)][(N-p)/(p-1)]$ is

$$(37)\qquad \frac{(p-1)\exp\left[-\frac{1}{2}\beta' A_{22}\beta/\sigma_{11\cdot 2}\right]}{(N-p)\Gamma\left[\frac{1}{2}(N-p)\right]} \cdot \sum_{\alpha=0}^{\infty} \frac{\left(\dfrac{\beta' A_{22}\beta}{2\sigma_{11\cdot 2}}\right)^{\alpha} \left[\dfrac{(p-1)f}{N-p}\right]^{\frac{1}{2}(p-1)+\alpha-1} \Gamma\left[\frac{1}{2}(N-1)+\alpha\right]}{\alpha!\,\Gamma\left[\frac{1}{2}(p-1)+\alpha\right]\left[1+\dfrac{(p-1)f}{N-p}\right]^{\frac{1}{2}(N-1)+\alpha}},$$

and the conditional density of $W = R^2$ is $(df = [(N-p)/(p-1)](1-w)^{-2}\,dw)$

$$(38)\qquad \frac{\exp\left[-\frac{1}{2}\beta' A_{22}\beta/\sigma_{11\cdot 2}\right]}{\Gamma\left[\frac{1}{2}(N-p)\right]}(1-w)^{\frac{1}{2}(N-p)-1} \cdot \sum_{\alpha=0}^{\infty} \frac{\left(\dfrac{\beta' A_{22}\beta}{2\sigma_{11\cdot 2}}\right)^{\alpha} w^{\frac{1}{2}(p-1)+\alpha-1}\,\Gamma\left[\frac{1}{2}(N-1)+\alpha\right]}{\alpha!\,\Gamma\left[\frac{1}{2}(p-1)+\alpha\right]}.$$

To obtain the unconditional density we need to multiply (38) by the density of $Z_1^{(2)}, \dots, Z_n^{(2)}$ to obtain the joint density of W and $Z_1^{(2)}, \dots, Z_n^{(2)}$ and then integrate with respect to the latter set to obtain the marginal density of W. We have

$$(39)\qquad \frac{\beta' A_{22}\beta}{\sigma_{11\cdot 2}} = \frac{\beta' \sum_{\alpha=1}^{n} z_\alpha^{(2)} z_\alpha^{(2)\prime}\beta}{\sigma_{11\cdot 2}} = \sum_{\alpha=1}^{n}\left(\frac{\beta' z_\alpha^{(2)}}{\sqrt{\sigma_{11\cdot 2}}}\right)^2.$$

Since the distribution of $\mathbf{Z}_\alpha^{(2)}$ is $N(\mathbf{0}, \mathbf{\Sigma}_{22})$, the distribution of $\boldsymbol{\beta}'\mathbf{Z}_\alpha^{(2)}/\sqrt{\sigma_{11\cdot 2}}$ is normal with mean zero and variance

$$(40) \qquad \mathscr{E}\left(\frac{\boldsymbol{\beta}'\mathbf{Z}_\alpha^{(2)}}{\sqrt{\mathbf{\Sigma}_{11\cdot 2}}}\right)^2 = \frac{\mathscr{E}\boldsymbol{\beta}'\mathbf{Z}_\alpha^{(2)}\mathbf{Z}_\alpha^{(2)\prime}\boldsymbol{\beta}}{\sigma_{11\cdot 2}}$$

$$= \frac{\boldsymbol{\beta}'\mathbf{\Sigma}_{22}\boldsymbol{\beta}}{\sigma_{11} - \boldsymbol{\beta}'\mathbf{\Sigma}_{22}\boldsymbol{\beta}} = \frac{\boldsymbol{\beta}'\mathbf{\Sigma}_{22}\boldsymbol{\beta}/\sigma_{11}}{1 - \boldsymbol{\beta}'\mathbf{\Sigma}_{22}\boldsymbol{\beta}/\sigma_{11}}$$

$$= \frac{\bar{R}^2}{1 - \bar{R}^2}.$$

Thus $(\boldsymbol{\beta}'A_{22}\boldsymbol{\beta}/\sigma_{11\cdot 2})/[\bar{R}^2/(1-\bar{R}^2)]$ has a χ^2-distribution with n degrees of freedom. Let $\bar{R}^2/(1-\bar{R}^2) = \phi$. Then $\boldsymbol{\beta}'A_{22}\boldsymbol{\beta}/\sigma_{11\cdot 2} = \phi\chi_n^2$. We compute

$$(41) \quad \mathscr{E}e^{-\frac{1}{2}\phi\chi_n^2}\left(\frac{\phi\chi_n^2}{2}\right)^\alpha$$

$$= \frac{\phi^\alpha}{2^\alpha}\int_0^\infty u^\alpha e^{-\frac{1}{2}\phi u}\frac{1}{2^{\frac{1}{2}n}\Gamma\left(\frac{1}{2}n\right)}u^{\frac{1}{2}n-1}e^{-\frac{1}{2}u}\,du$$

$$= \frac{\phi^\alpha}{2^\alpha}\int_0^\infty \frac{1}{2^{\frac{1}{2}n}\Gamma\left(\frac{1}{2}n\right)}u^{\frac{1}{2}n+\alpha-1}e^{-\frac{1}{2}(1+\phi)u}\,du$$

$$= \frac{\phi^\alpha}{(1+\phi)^{\frac{1}{2}n+\alpha}}\frac{\Gamma\left(\frac{1}{2}n+\alpha\right)}{\Gamma\left(\frac{1}{2}n\right)}\int_0^\infty \frac{1}{2^{\frac{1}{2}n+\alpha}\Gamma\left(\frac{1}{2}n+\alpha\right)}v^{\frac{1}{2}n+\alpha-1}e^{-\frac{1}{2}v}\,dv$$

$$= \frac{\phi^\alpha}{(1+\phi)^{\frac{1}{2}n+\alpha}}\frac{\Gamma\left(\frac{1}{2}n+\alpha\right)}{\Gamma\left(\frac{1}{2}n\right)}.$$

Applying this result to (38), we obtain as the density of R^2

$$(42) \quad \frac{(1-R^2)^{\frac{1}{2}(n-p-1)}(1-\bar{R}^2)^{\frac{1}{2}n}}{\Gamma\left[\frac{1}{2}(n-p+1)\right]\Gamma\left(\frac{1}{2}n\right)}\sum_{\mu=0}^\infty \frac{(\bar{R}^2)^\mu (R^2)^{\frac{1}{2}(p-1)+\mu-1}\Gamma^2\left(\frac{1}{2}n+\mu\right)}{\mu!\,\Gamma\left[\frac{1}{2}(p-1)+\mu\right]}.$$

Fisher (1928) found this distribution. It can also be written

$$(43) \quad \frac{\Gamma\left(\frac{1}{2}n\right)(1-\bar{R}^2)^{\frac{1}{2}n}}{\Gamma\left[\frac{1}{2}(n-p+1)\right]\Gamma\left[\frac{1}{2}(p-1)\right]}(R^2)^{\frac{1}{2}(p-3)}(1-R^2)^{\frac{1}{2}(n-p-1)}$$

$$\cdot F\left[\tfrac{1}{2}n, \tfrac{1}{2}n; \tfrac{1}{2}(p-1); R^2\bar{R}^2\right],$$

where F is the hypergeometric function defined in (41) of Section 4.2.

Another form of the density can be obtained when $n-p+1$ is even. We have

$$(44)\quad \sum_{\mu=0}^{\infty} \frac{(R^2\bar{R}^2)^{\mu}}{\mu!} \frac{\Gamma^2(\frac{1}{2}n+\mu)}{\Gamma[\frac{1}{2}(p-1)+\mu]}$$

$$= \sum_{\mu=0}^{\infty} \frac{(R^2\bar{R}^2)^{\mu}}{\mu!} \Gamma(\tfrac{1}{2}n+\mu)\left(\frac{\partial}{\partial t}\right)^{\frac{1}{2}(n-p+1)} t^{\frac{1}{2}n+\mu-1}\bigg|_{t=1}$$

$$= \left(\frac{\partial}{\partial t}\right)^{\frac{1}{2}(n-p+1)} t^{\frac{1}{2}n-1} \sum_{\mu=0}^{\infty} \frac{(t\bar{R}^2R^2)^{\mu}}{\mu!} \frac{\Gamma(\frac{1}{2}n+\mu)}{\Gamma(\frac{1}{2}n)}\bigg|_{t=1} \Gamma(\tfrac{1}{2}n)$$

$$= \Gamma(\tfrac{1}{2}n)\left(\frac{\partial}{\partial t}\right)^{\frac{1}{2}(n-p+1)} t^{\frac{1}{2}n-1}(1-tR^2\bar{R}^2)^{-\frac{1}{2}n}\bigg|_{t=1}.$$

The density is therefore

$$(45)\quad \frac{(1-\bar{R}^2)^{\frac{1}{2}n}(R^2)^{\frac{1}{2}(p-3)}(1-R^2)^{\frac{1}{2}(n-p-1)}}{\Gamma[\frac{1}{2}(n-p+1)]} \cdot \left(\frac{\partial}{\partial t}\right)^{\frac{1}{2}(n-p+1)} t^{\frac{1}{2}n-1}(1-tR^2\bar{R}^2)^{-\frac{1}{2}n}\bigg|_{t=1}.$$

Theorem 4.4.5. *The density of the square of the multiple correlation coefficient, R^2, between X_1 and $X_2, \dots, X_p$ based on a sample of $N=n+1$ is given by* (42) *or* (43) [*or* (45) *in the case of $n-p+1$ even*], *where $\bar{R}^2$ is the corresponding population multiple correlation coefficient.*

The moments of R are

$$(46)\quad \mathscr{E}R^h = \frac{(1-\bar{R}^2)^{\frac{1}{2}n}}{\Gamma[\frac{1}{2}(n-p+1)]\Gamma(\frac{1}{2}n)} \sum_{\mu=0}^{\infty} \frac{(\bar{R}^2)^{\mu}\Gamma^2(\frac{1}{2}n+\mu)}{\Gamma[\frac{1}{2}(p-1)+\mu]\,\mu!}$$

$$\cdot \int_0^1 (1-R^2)^{\frac{1}{2}(n-p+1)-1}(R^2)^{\frac{1}{2}(p+h-1)+\mu-1}\,d(R^2)$$

$$= \frac{(1-\bar{R}^2)^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)} \sum_{\mu=0}^{\infty} \frac{(\bar{R}^2)^{\mu}\Gamma^2(\frac{1}{2}n+\mu)\Gamma[\frac{1}{2}(p+h-1)+\mu]}{\mu!\Gamma[\frac{1}{2}(p-1)+\mu]\Gamma[\frac{1}{2}(n+h)+\mu]}.$$

The sample multiple correlation tends to overestimate the population multiple correlation. The sample multiple correlation is the maximum sample correlation between x_1 and linear combinations of $x^{(2)}$ and hence is greater than the sample correlation between x_1 and $\boldsymbol{\beta}'\boldsymbol{x}^{(2)}$; however, the latter is the simple sample correlation corresponding to the simple population correlation between x_1 and $\boldsymbol{\beta}'\boldsymbol{x}^{(2)}$, which is $\overline{R}$, the population multiple correlation.

Suppose R_1 is the multiple correlation in the first of two samples and $\hat{\boldsymbol{\beta}}_1$ is the estimate of $\boldsymbol{\beta}$; then the simple correlation between x_1 and $\hat{\boldsymbol{\beta}}_1'x^{(2)}$ in the second sample will tend to be less than R_1 and in particular will be less than R_2, the multiple correlation in the second sample. This has been called "the shrinkage of the multiple correlation."

Kramer (1963) and Lee (1972) have given tables of the upper significance points of R. Gajjar (1967), Gurland (1968), Gurland and Milton (1970), Khatri (1966), and Lee (1917b) have suggested approximations to the distributions of $R^2/(1-R^2)$ and obtained large-sample results.

4.4.4. Some Optimal Properties of the Multiple Correlation Test

Theorem 4.4.6. *Given the observations $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N$ from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, of all tests of $\overline{R} = 0$ at a given significance level based on $\bar{\boldsymbol{x}}$ and $A = \sum_{\alpha=1}^N (\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})'$ that are invariant with respect to transformations*

$$\bar{x}_1^* = c\bar{x}_1 + d, \qquad \bar{\boldsymbol{x}}^{(2)*} = \boldsymbol{C}\bar{\boldsymbol{x}}^{(2)} + \boldsymbol{d},$$
$$a_{11}^* = c^2 a_{11}, \qquad \boldsymbol{a}_{(1)}^* = c\boldsymbol{C}\boldsymbol{a}_{(1)}, \qquad \boldsymbol{A}_{22}^* = \boldsymbol{C}\boldsymbol{A}_{22}\boldsymbol{C}', \tag{47}$$

any critical rejection region given by R greater than a constant is uniformly most powerful.

Proof. The multiple correlation coefficient R is invariant under the transformation, and any function of the sufficient statistics that is invariant is a function of R. (See Problem 4.34.) Therefore, any invariant test must be based on R. The Neyman–Pearson fundamental lemma applied to testing the null hypothesis $\overline{R} = 0$ against a specific alternative $\overline{R} = \overline{R}_0 > 0$ tells us the most powerful test at a given level of significance is based on the ratio of the density of R for $\overline{R} = \overline{R}_0$, which is (42) times $2R$ [because (42) is the density of R^2], to the density for $R = 0$, which is (22). The ratio is a positive constant times

$$\sum_{\mu=0}^{\infty} \frac{\left(\overline{R}_0^2\right)^{\mu} \Gamma^2\left(\frac{1}{2}n + \mu\right)}{\mu! \Gamma\left[\frac{1}{2}(p-1) + \mu\right]} R^{p-2+2\mu}. \tag{48}$$

Since (48) is an increasing function of R for $R \geq 0$, the set of R for which (48) is greater than a constant is an interval of R greater than a constant. ∎

Theorem 4.4.7. *On the basis of observations $x_1, \ldots, x_N$ from $N(\mu, \Sigma)$, of all tests of $\bar{R} = 0$ at a given significance level with power depending only on $\bar{R}$, the test with critical region given by R greater than a constant is uniformly most powerful.*

Theorem 4.4.7 follows from Theorem 4.4.6 in the same way that Theorem 5.6.4 follows from Theorem 5.6.1.

4.5. ELLIPTICALLY CONTOURED DISTRIBUTIONS

4.5.1. Observations Elliptically Contoured

Suppose $x_1, \ldots, x_N$ are N independent observations on a random p-vector X with density

$$(1) \qquad |\Lambda|^{-\frac{1}{2}} g\left[(x - \nu)'\Lambda^{-1}(x - \nu)\right].$$

The sample covariance matrix S is an unbiased estimator of the covariance matrix $\Sigma = [\mathscr{E}R^2/p]\Lambda$, where $R^2 = (X - \nu)'\Lambda^{-1}(X - \nu)$ and $\mathscr{E}R^2 < \infty$. An estimator of $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}} = \lambda_{ij}/\sqrt{\lambda_{ii}\lambda_{jj}}$ is $r_{ij} = s_{ij}/\sqrt{s_{ii}s_{jj}}$, $i, j = 1, \ldots, p$. The small-sample distribution of r_{ij} is in general difficult to obtain, but the asymptotic distribution can be obtained from the limiting normal distribution of $\sqrt{N}(S - \Sigma)$ given in (13) of Section 3.6.

First we prove a general theorem on asymptotic distributions of functions of the sample covariance matrix S using Theorems 4.2.3 and 3.6.5. Define

$$(2) \qquad s = \operatorname{vec} S, \qquad \sigma = \operatorname{vec} \Sigma.$$

Theorem 4.5.1. *Let $f(s)$ be a vector-valued function such that each component of $f(s)$ has a nonzero differential at $s = \sigma$. Suppose S is the covariance of a sample from* (1) *such that $\mathscr{E}R^4 < \infty$. Then*

$$(3) \qquad \sqrt{N}\left[f(s) - f(\sigma)\right] = \frac{\partial f(\sigma)}{\partial \sigma'}\sqrt{N}(s - \sigma) + o_p(1)$$

$$\xrightarrow{d} N\left\{0, \frac{\partial f(\sigma)}{\partial \sigma'}\left[2(1+\kappa)(\Sigma \otimes \Sigma) + \kappa\sigma\sigma'\right]\left(\frac{\partial f(\sigma)}{\partial \sigma'}\right)'\right\}.$$

Corollary 4.5.1. *If*

$$(4) \qquad f(cs) = f(s)$$

for all $c > 0$ and all positive definite S and the conditions of Theorem 4.5.1 *hold, then*

$$(5) \qquad \sqrt{N}\left[f(s) - f(\sigma)\right] \xrightarrow{d} N\left[0, 2(1+\kappa)\frac{\partial f(\sigma)}{\partial \sigma'}(\Sigma \otimes \Sigma)\left(\frac{\partial f(\sigma)}{\partial \sigma'}\right)'\right].$$

Proof. From (4) we deduce

$$0 = \frac{\partial f(cs)}{\partial c} = \frac{\partial f(cs)}{\partial s'}\frac{\partial (cs)}{\partial c} = \frac{\partial f(cs)}{\partial s'}s. \tag{6}$$

That is,

$$\frac{\partial f(\sigma)}{\partial \sigma'}\sigma = 0. \qquad \blacksquare \tag{7}$$

The conclusion of Corollary 4.5.1 can be framed as

$$\frac{\sqrt{N}}{\sqrt{1+\kappa}}[f(s) - f(\sigma)] \xrightarrow{d} N\left[0, 2\frac{\partial f(\sigma)}{\partial \sigma'}(\Sigma \otimes \Sigma)\left(\frac{\partial f(\sigma)}{\partial \sigma'}\right)'\right]. \tag{8}$$

The limiting normal distribution in (8) holds in particular when the sample is drawn from the normal distribution. The corollary holds true if κ is replaced by a consistent estimator $\hat{\kappa}$. For example, a consistent estimator of $1+\hat{\kappa}$ given by (16) of Section 3.6 is

$$1 + \hat{\kappa} = \sum_{\alpha=1}^{N}\left[(x_\alpha - \bar{x})'S^{-1}(x_\alpha - \bar{x})\right]^2 \Big/ [Np(p+2)]. \tag{9}$$

A sample correlation such as $f(s) = r_{ij} = s_{ij}/\sqrt{s_{ii}s_{jj}}$ or a set of such correlations is a function of S that is invariant under scale transformations; that is, it satisfies (4).

Corollary 4.5.2. *Under the conditions of Theorem* 4.5.1,

$$\sqrt{\frac{N}{1+\hat{\kappa}}}\,\frac{(r_{ij} - \rho_{ij})}{\sqrt{1 - r_{ij}^2}} \xrightarrow{d} N(0,1). \tag{10}$$

As in the case of the observations normally distributed,

$$\sqrt{\frac{N}{1+\hat{\kappa}}}\left(\frac{1}{2}\log\frac{1+r_{ij}}{1-r_{ij}} - \frac{1}{2}\log\frac{1+\rho_{ij}}{1-\rho_{ij}}\right) \xrightarrow{d} N(0,1). \tag{11}$$

Of course, any improvement of (11) over (10) depends on the distribution samples.

Partial correlations such as $r_{ij.q+1,\dots,p}$, $i,j = 1,\dots,q$, are also invariant functions of S.

Corollary 4.5.3. *Under the conditions of Theorem* 4.5.1,

$$\sqrt{\frac{N}{1+\hat{\kappa}}}\left(r_{ij.q+1,\dots,p} - \rho_{ij.q+1,\dots,p}\right) \xrightarrow{d} N(0,1). \tag{12}$$

Now let us consider the asymptotic distribution of R^2, the square of the multiple correlation, when $\overline{R}^2$, the square of the population multiple correlation, is 0. We use the notation of Section 4.4. $\overline{R}^2 = 0$ is equvialent to $\boldsymbol{\sigma}_{(1)} = \mathbf{0}$. Since the sample and population multiple correlation coefficients between X_1 and $X^{(2)} = (X_2, \dots, X_p)'$ are invariant with respect to linear transformations (47) of Section 4.4, for purposes of studying the distribution of R^2 we can assume $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_p$. In that case $s_{11} \xrightarrow{p} 1$, $\boldsymbol{s}_{(1)} \xrightarrow{p} \mathbf{0}$, and $S_{22} \xrightarrow{p} \boldsymbol{I}_{p-1}$. Furthermore, for $k, i \neq 1$ and $j = l = 1$, Lemma 3.6.1 gives

$$\mathscr{E}\boldsymbol{s}_{(1)}\boldsymbol{s}'_{(1)} = \left(\frac{1}{n} + \frac{\kappa}{N}\right)\boldsymbol{I}_{p-1}. \tag{13}$$

Theorem 4.5.2. *Under the conditions of Theorem* 4.5.1

$$\sqrt{\frac{N}{1+\kappa}}\,\boldsymbol{s}_{(1)} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{I}_{p-1}). \tag{14}$$

Corollary 4.5.4. *Under the conditions of Theorem* 4.5.1

$$\frac{NR^2}{1+\hat{\kappa}} = \frac{N\boldsymbol{s}'_{(1)}S_{22}^{-1}\boldsymbol{s}_{(1)}}{(1+\hat{\kappa})s_{11}} \xrightarrow{d} \chi^2_{p-1}. \tag{15}$$

4.5.2. Elliptically Contoured Matrix Distributions

Now let us turn to the model

$$|\boldsymbol{\Lambda}|^{-N/2} g\left[\operatorname{tr}(\boldsymbol{X} - \boldsymbol{\varepsilon}_N\boldsymbol{\nu}')\boldsymbol{\Lambda}^{-1}(\boldsymbol{X} - \boldsymbol{\varepsilon}_N\boldsymbol{\nu}')'\right] \tag{16}$$

based on the vector spherical model $g(\operatorname{tr} Y'Y)$. The unbiased estimators of $\boldsymbol{\nu}$ and $\boldsymbol{\Sigma} = (\mathscr{E}R^2/p)\boldsymbol{\Lambda}$ are $\bar{\boldsymbol{x}} = (1/N)\boldsymbol{X}'\boldsymbol{\varepsilon}_N$ and $S = (1/n)A$, where $A = (\boldsymbol{X} - \boldsymbol{\varepsilon}_n\bar{\boldsymbol{x}}')'(\boldsymbol{X} - \boldsymbol{\varepsilon}_N\bar{\boldsymbol{x}}')$.

Since

$$(\boldsymbol{X} - \boldsymbol{\varepsilon}_N\boldsymbol{\nu}')'(\boldsymbol{X} - \boldsymbol{\varepsilon}_N\boldsymbol{\nu}') = A + N(\bar{\boldsymbol{x}} - \boldsymbol{\nu})(\bar{\boldsymbol{x}} - \boldsymbol{\nu})', \tag{17}$$

A and $\bar{\boldsymbol{x}}$ are a complete set of sufficient statistics.

The maximum likelihood estimators of $\boldsymbol{\nu}$ and $\boldsymbol{\Lambda}$ are $\hat{\boldsymbol{\nu}} = \bar{\boldsymbol{x}}$ and $\hat{\boldsymbol{\Lambda}} = (p/w_g)A$. The maximum likelihood estimator of $\rho_{ij} = \lambda_{ij}/\sqrt{\lambda_{ii}\lambda_{jj}} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$ is $\hat{\rho}_{ij} = a_{ij}/\sqrt{a_{ii}a_{jj}} = s_{ij}/\sqrt{s_{ii}s_{jj}}$ (Theorem 3.6.4).

The sample correlation r_{ij} is a function $f(\boldsymbol{X})$ that satisfies the conditions (45) and (46) of Theorem 3.6.5 and hence has the same distribution for an arbitrary density $g[\operatorname{tr}(\cdot)]$ as for the normal density $g[\operatorname{tr}(\cdot)] = \text{const } e^{-\frac{1}{2}\operatorname{tr}(\cdot)}$. Similarly, a partial correlation $r_{ij.q+1,\dots,p}$ and a multiple correlation R^2 satisfy the conditions, and the conclusion holds.

Theorem 4.5.3. *When X has the vector elliptical density* (16), *the distributions of r_{ii}, $r_{ij.q+1}$, and R^2 are the distributions derived for normally distributed observations.*

It follows from Theorem 4.5.3 that the asymptotic distributions of r_{ij}, $r_{ij.q+1,\dots,p}$, and R^2 are the same as for sampling from normal distributions.

The class of left spherical matrices Y with densities is the class of $g(Y'Y)$. Let $X = YC' + \varepsilon_N \nu'$, where $C'\Lambda^{-1}C = I$, that is, $\Lambda = CC'$. Then X has the density

$$|C|^{-N} g\left[C^{-1}(X-\varepsilon_N\nu')'(X-\varepsilon_N\nu')(C')^{-1}\right]. \tag{18}$$

We now find a stochastic representation of the matrix Y.

Lemma 4.5.1. *Let $V = (v_1,\dots,v_p)$, where v_i is an N-component vector, $i = 1,\dots,p$. Define recursively $w_1 = v_1$,*

$$w_i = v_i - \sum_{j=1}^{i-1} \frac{v_i' w_j}{w_j' w_j} w_j, \qquad i = 2,\dots,p. \tag{19}$$

Let $u_i = w_i/\|w_i\|$. Then $\|u_i\| = 1$, $i = 1,\dots,p$, and $u_i'u_j = 0$, $i \neq j$. Further,

$$V = UT', \tag{20}$$

where $U = (u_1,\dots,u_p)$; $t_{ii} = \|w_i\|$, $i = 1,\dots,p$; $t_{ij} = v_i'w_j/\|w_j\| = v_i'u_j$, $j = 1,\dots,i-1$, $i = 1,\dots,p$; and $t_{ij} = 0$, $i < j$.

The proof of the lemma is given in the first part of Section 7.2 and as the Gram–Schmidt orthogonalization in the Appendix (Section A.5.1). This lemma generalizes the construction in Section 3.2; see Figure 3.1. See also Figure 7.1.

Note that T is lower triangular, $U'U = I_p$, and $V'V = TT'$. The last equation, $t_{ii} \geq 0$, $i = 1,\dots,p$, and $t_{ij} = 0$, $i < j$, can be solved uniquely for T. Thus T is a function of $V'V$ (and the restrictions).

Let Y $(N \times p)$ have the density $g(Y'Y)$, and let O_N be an orthogonal $N \times N$ matrix. Then $Y^* = O_N Y$ has the density $g(Y^{*\prime}Y^*)$. Hence $Y^* = O_N Y \overset{d}{=} Y$. Let $Y^* = U^* T^{*\prime}$, where $t_{ii}^* > 0$, $i = 1,\dots,p$, and $t_{ij}^* = 0$, $i < j$. From $Y^{*\prime}Y^* = Y'Y$ it follows that $T^*T^{*\prime} = TT'$ and hence $T^* = T$, $Y^* = U^*T$, and $U^* = O_N U \overset{d}{=} U$. Let the space of U $(N \times p)$ such that $U'U = I_p$ be denoted $O(N \times p)$.

Definition 4.5.1. *If U $(N \times p)$ satisfies $U'U = I_p$ and $O_N U \overset{d}{=} U$ for all orthogonal O_N, then U is* uniformly distributed on $O(N \times p)$.

The space of U satisfying $U'U = I_p$ is known as a *Steifel manifold*. The probability measure of Definition 4.5.1 is known as the *Haar invariant distribution*. The property $O_N U \stackrel{d}{=} U$ for all orthogonal O_N defines the (normalized) measure uniquely [Halmos (1956)].

Theorem 4.5.4. *If Y $(N \times p)$ has the density $g(Y'Y)$, then U defined by $Y = UT'$, $U'U = I_p$, $t_{ii} > 0$, $i = 1, \ldots, p$, and $t_{ij} = 0$, $i < j$, is* uniformly distributed on $O(N \times p)$.

The proof of Corollary 7.2.1 shows that for arbitrary $g(\cdot)$ the density of T is

$$\prod_{i=1}^{p} \{C[\tfrac{1}{2}(N+1-i)] t_{ii}^{N-i}\} g(\operatorname{tr} TT'), \tag{21}$$

where $C(\cdot)$ is defined in (8) of Section 2.7.

The *stochastic representation* of Y $(N \times p)$ with density $g(Y'Y)$ is

$$Y = UT', \tag{22}$$

where U $(N \times p)$ is uniformly distributed on $O(N \times p)$ and T is lower triangular with positive diagonal elements and has density (21).

Theorem 4.5.5. *Let $f(X)$ be a vector-valued function of X $(N \times p)$ such that*

$$f(X + \varepsilon_N \nu') = f(X) \tag{23}$$

for all ν and

$$f(XG') = f(X) \tag{24}$$

for all G $(p \times p)$. Then the distribution of $f(X)$ where X has an arbitrary density (18) *is the same as the distribution of $f(X)$ where X has the normal density* (18).

Proof. From (23) we find that $f(X) = f(YC')$, and from (24) we find $f(YC') = f(UT'C') = f(U)$, which is the same for arbitrary and normal densities (18). ■

Corollary 4.5.5. *Let $f(X)$ be a vector-valued function of X $(N \times p)$ with the density* (18), *where $\nu = 0$. Suppose* (24) *holds for all G $(p \times p)$. Then the distribution of $f(X)$ for an arbitrary density* (18) *is the same as the distribution of $f(X)$ when X has the normal density* (18).

The condition (24) of Corollary 4.5.5 is that $f(X)$ is invariant with respect to linear transformations $X \to XG$.

The density (18) can be written as

$$|C|^{-1} g\{C^{-1}[A + N(\bar{x} - \nu)(\bar{x} - \nu)'](C')^{-1}\}, \tag{25}$$

which shows that A and $\bar{x}$ are a complete set of sufficient statistics for $\Lambda = CC'$ and ν.

PROBLEMS

4.1. (Sec. 4.2.1) Sketch

$$k_N(r) = \frac{\Gamma[\frac{1}{2}(N-1)]}{\Gamma(\frac{1}{2}N - 1)\sqrt{\pi}}(1 - r^2)^{\frac{1}{2}(N-4)}$$

for **(a)** $N = 3$, **(b)** $N = 4$, **(c)** $N = 5$, and **(d)** $N = 10$.

4.2. (Sec. 4.2.1) Using the data of Problem 3.1, test the hypothesis that X_1 and X_2 are independent against all alternatives of dependence at significance level 0.01.

4.3. (Sec. 4.2.1) Suppose a sample correlation of 0.65 is observed in a sample of 10. Test the hypothesis of independence against the alternatives of positive correlation at significance level 0.05.

4.4. (Sec. 4.2.2) Suppose a sample correlation of 0.65 is observed in a sample of 20. Test the hypothesis that the population correlation is 0.4 against the alternatives that the population correlation is greater than 0.4 at significance level 0.05.

4.5. (Sec. 4.2.1) Find the significance points for testing $\rho = 0$ at the 0.01 level with $N = 15$ observations against alternatives **(a)** $\rho \neq 0$, **(b)** $\rho > 0$, and **(c)** $\rho < 0$.

4.6. (Sec. 4.2.2) Find significance points for testing $\rho = 0.6$ at the 0.01 level with $N = 20$ observations against alternatives **(a)** $\rho \neq 0.6$, **(b)** $\rho > 0.6$, and **(c)** $\rho < 0.6$.

4.7. (Sec. 4.2.2) Tablulate the power function at $\rho = -1(0.2)1$ for the tests in Problem 4.5. Sketch the graph of each power function.

4.8. (Sec. 4.2.2) Tablulate the power function at $\rho = -1(0.2)1$ for the tests in Problem 4.6. Sketch the graph of each power function.

4.9. (Sec. 4.2.2) Using the data of Problem 3.1, find a (two-sided) confidence interval for ρ_{12} with confidence coefficient 0.99.

4.10. (Sec. 4.2.2) Suppose $N = 10$, $r = 0.795$. Find a one-sided confidence interval for ρ [of the form $(r_0, 1)$] with confidence coefficient 0.95.

4.11. (Sec. 4.2.3) Use Fisher's z to test the hypothesis $\rho = 0.7$ against alternatives $\rho \neq 0.7$ at the 0.05 level with $r = 0.5$ and $N = 50$.

4.12. (Sec. 4.2.3) Use Fisher's z to test the hypothesis $\rho_1 = \rho_2$ against the alternatives $\rho_1 \neq \rho_2$ at the 0.01 level with $r_1 = 0.5, N_1 = 40, r_2 = 0.6, N_2 = 40$.

4.13. (Sec. 4.2.3) Use Fisher's z to estimate ρ based on sample correlations of -0.7 ($N = 30$) and of -0.6 ($N = 40$).

4.14. (Sec. 4.2.3) Use Fisher's z to obtain a confidence interval for ρ with confidence 0.95 based on a sample correlation of 0.65 and a sample size of 25.

4.15. (Sec. 4.2.2). Prove that when $N = 2$ and $\rho = 0$, $\Pr\{r = 1\} = \Pr\{r = -1\} = \frac{1}{2}$.

4.16. (Sec. 4.2) Let $k_N(r, \rho)$ be the density of the sample correlation coefficient r for a given value of ρ and N. Prove that r has a monotone likelihood ratio; that is, show that if $\rho_1 > \rho_2$, then $k_N(r, \rho_1)/k_N(r, \rho_2)$ is monotonically increasing in r. [*Hint:* Using (40), prove that if

$$F\left[\tfrac{1}{2}, \tfrac{1}{2}; n + \tfrac{1}{2}; \tfrac{1}{2}(1 + \rho r)\right] = \sum_{\alpha=0}^{\infty} c_\alpha (1 + \rho r)^\alpha = g(r, \rho)$$

has a monotone ratio, then $k_N(r, \rho)$ does. Show

$$\frac{\partial^2}{\partial \rho\, \partial r} \log g(r, \rho) = \frac{\sum_{\alpha, \beta=0}^{\infty} c_\alpha c_\beta \left[(\alpha - \beta)^2 r\rho + (\alpha + \beta)\right](1 + r\rho)^{\alpha+\beta-2}}{2\left[\sum_{\alpha=0}^{\infty} c_\alpha (1 + r\rho)^\alpha\right]^2};$$

if $(\partial^2/\partial\rho\,\partial r)\log g(r, \rho) > 0$, then $g(r, \rho)$ has a monotone ratio. Show the numerator of the above expression is positive by showing that for each α the sum on β is positive; use the fact that $c_{\alpha+1} < \frac{1}{2} c_\alpha$.]

4.17. (Sec. 4.2) Show that of all tests of ρ_0 against a specific ρ_1 ($> \rho_0$) based on r, the procedures for which $r > c$ implies rejection are the best. [*Hint:* This follows from Problem 4.16.]

4.18. (Sec. 4.2) Show that of all tests of $\rho = \rho_0$ against $\rho > \rho_0$ based on r, a procedure for which $r > c$ implies rejection is uniformly most powerful.

4.19. (Sec. 4.2) Prove r has a monotone likelihood ratio for $r > 0$, $\rho > 0$ by proving $h(r) = k_N(r, \rho_1)/k_N(r, \rho_2)$ is monotonically increasing for $\rho_1 > \rho_2$. Here $h(r)$ is a constant times $(\sum_{\alpha=0}^{\infty} c_\alpha \rho_1^\alpha r^\alpha)/(\sum_{\alpha=0}^{\infty} c_\alpha \rho_2^\alpha r^\alpha)$. In the numerator of $h'(r)$, show that the coefficient of r^β is positive.

4.20. (Sec. 4.2) Prove that if $\mathbf{\Sigma}$ is diagonal, then the sets r_{ij} and a_{ii} are independently distributed. [*Hint:* Use the facts that r_{ij} is invariant under scale transformations and that the density of the observations depends only on the a_{ii}.]

4.21. (Sec. 4.2.1) Prove that if $\rho = 0$

$$\mathscr{E} r^{2m} = \frac{\Gamma[\frac{1}{2}(N-1)]\Gamma(m+\frac{1}{2})}{\sqrt{\pi}\,\Gamma[\frac{1}{2}(N-1)+m]}.$$

4.22. (Sec. 4.2.2) Prove $f_1(\rho)$ and $f_2(\rho)$ are monotonically increasing functions of ρ.

4.23. (Sec. 4.2.2) Prove that the density of the sample correlation r [given by (38)] is

$$\frac{n-1}{\pi}(1-\rho^2)^{\frac{1}{2}n}(1-r^2)^{\frac{1}{2}(n-3)}\int_0^1 \frac{x^{n-1}\,dx}{(1-\rho r x)^n\sqrt{1-x^2}}.$$

[*Hint:* Expand $(1-\rho r x)^{-n}$ in a power series, integrate, and use the duplication formula for the gamma function.]

4.24. (Sec. 4.2) Prove that (39) is the density of r. [*Hint:* From Problem 2.12 show

$$\int_0^\infty \int_0^\infty e^{-\frac{1}{2}(y^2-2xyz+z^2)}\,dy\,dz = \frac{\cos^{-1}(-x)}{\sqrt{1-x^2}}.$$

Then argue

$$\int_0^\infty \int_0^\infty (yz)^{n-1} e^{-\frac{1}{2}(y^2-2xyz+z^2)}\,dy\,dz = \frac{d^{n-1}}{dx^{n-1}}\frac{\cos^{-1}(-x)}{\sqrt{1-x^2}}.$$

Finally show that the integral of (31) with respect to a_{11} $(=y^2)$ and a_{22} $(=z^2)$ is (39).]

4.25. (Sec. 4.2) Prove that (40) is the density of r. [*Hint:* In (31) let $a_{11} = ue^{-v}$ and $a_{22} = ue^{v}$; show that the density of v $(0 \le v < \infty)$ and r $(-1 \le r \le 1)$ is

$$\frac{n-1}{\pi\sqrt{2}}(1-\rho^2)^{\frac{1}{2}n}(1-\rho r)^{-n+\frac{1}{2}}(1-r^2)^{\frac{1}{2}(n-3)}v^{-\frac{1}{2}}(1-v)^{n-1}[1-\tfrac{1}{2}(1+\rho r)v]^{-\frac{1}{2}}.$$

Use the expansion

$$(1-y)^{-\frac{1}{2}} = \sum_{j=0}^{\infty}\frac{\Gamma(j+\frac{1}{2})}{\Gamma(\frac{1}{2})j!}y^j.$$

Show that the integral is (40).]

4.26. (Sec. 4.2) Prove for integer h

$$\mathscr{E}r^{2h+1} = \frac{(1-\rho^2)^{\frac{1}{2}n}}{\sqrt{\pi}\,\Gamma(\frac{1}{2}n)} \sum_{\beta=0}^{\infty} \frac{(2\rho)^{2\beta+1}}{(2\beta+1)!} \frac{\Gamma^2[\frac{1}{2}(n+1)+\beta]\Gamma(h+\beta+\frac{3}{2})}{\Gamma(\frac{1}{2}n+h+\beta+1)},$$

$$\mathscr{E}r^{2h} = \frac{(1-\rho^2)^{\frac{1}{2}n}}{\sqrt{\pi}\,\Gamma(\frac{1}{2}n)} \sum_{\beta=0}^{\infty} \frac{(2\rho)^{2\beta}}{(2\beta)!} \frac{\Gamma^2(\frac{1}{2}n+\beta)\Gamma(h+\beta+\frac{1}{2})}{\Gamma(\frac{1}{2}n+h+\beta)}.$$

4.27. (Sec. 4.2) *The t-distribution.* Prove that if X and Y are independently distributed, X having the distribution $N(0,1)$ and Y having the χ^2-distribution with m degrees of freedom, then $W = X/\sqrt{Y/m}$ has the density

$$\frac{\Gamma[\frac{1}{2}(m+1)]}{\sqrt{m}\,\sqrt{\pi}\,\Gamma(\frac{1}{2}m)}\left(1+\frac{t^2}{m}\right)^{-\frac{1}{2}(m+1)}.$$

[*Hint:* In the joint density of X and Y, let $x = tw^{\frac{1}{2}}m^{-\frac{1}{2}}$ and integrate out w.]

4.28. (Sec. 4.2) Prove

$$\mathscr{E}r = \frac{(1-\rho^2)^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)} \sum_{\beta=0}^{\infty} \frac{\rho^{2\beta+1}\Gamma^2[\frac{1}{2}(n+1)+\beta]}{\beta!\,\Gamma[\frac{1}{2}n+\beta+1]}.$$

[*Hint:* Use Problem 4.26 and the duplication formula for the gamma function.]

4.29. (Sec. 4.2) Show that $\sqrt{n}(r_{ij}-\rho_{ij})$, $(i,j) = (1,2),(1,3),(2,3)$, have a joint limiting distribution with variances $(1-\rho_{ij}^2)^2$ and covariances of r_{ij} and r_{ik}, $j \neq k$ being $\frac{1}{2}(2\rho_{jk} - \rho_{ij}\rho_{jk})(1-\rho_{ij}^2-\rho_{ik}^2-\rho_{jk}^2)+\rho_{jk}^2$.

4.30. (Sec. 4.3.2) Find a confidence interval for $\rho_{13\cdot 2}$ with confidence 0.95 based on $r_{13\cdot 2} = 0.097$ and $N = 20$.

4.31. (Sec. 4.3.2) Use Fisher's z to test the hypothesis $\rho_{12\cdot 34} = 0$ against alternatives $\rho_{12\cdot 34} \neq 0$ at significance level 0.01 with $r_{12\cdot 34} = 0.14$ and $N = 40$.

4.32. (Sec. 4.3) Show that the inequality $r_{12\cdot 3}^2 \leq 1$ is the same as the inequality $|r_{ij}| \geq 0$, where $|r_{ij}|$ denotes the determinant of the 3×3 correlation matrix.

4.33. (Sec. 4.3) *Invariance of the sample partial correlation coefficient.* Prove that $r_{12\cdot 3,\dots,p}$ is invariant under the transformations $x_{i\alpha}^* = a_i x_{i\alpha} + b_i' x_\alpha^{(3)} + c_i$, $a_i > 0$, $i = 1,2$, $x_\alpha^{(3)*} = Cx_\alpha^{(3)} + b$, $\alpha = 1,\dots,N$, where $x_\alpha^{(3)} = (x_{3\alpha},\dots,x_{p\alpha})'$, and that any function of $\bar{x}$ and $\hat{\Sigma}$ that is invariant under these transformations is a function of $r_{12\cdot 3,\dots,p}$.

4.34. (Sec. 4.4) *Invariance of the sample multiple correlation coefficient.* Prove that R is a function of the sufficient statistics $\bar{x}$ and S that is invariant under changes of location and scale of $x_{1\alpha}$ and nonsingular linear transformations of $x_\alpha^{(2)}$ (that is, $x_{1\alpha}^* = cx_{1\alpha} + d$, $x_\alpha^{(2)*} = Cx_\alpha^{(2)} + d$, $\alpha = 1,\dots,N$) and that every function of $\bar{x}$ and S that is invariant is a function of R.

4.35. (Sec. 4.4) Prove that conditional on $Z_{1\alpha} = z_{1\alpha}$, $\alpha = 1, \dots, n$, $R^2/(1 - R^2)$ is distributed like $T^2/(N^* - 1)$, where $T^2 = N^* \bar{x}' S^{-1} \bar{x}$ based on $N^* = n$ observations on a vector X with $p^* = p - 1$ components, with mean vector $(c/\sigma_{11})\boldsymbol{\sigma}'_{(1)}$ $(nc^2 = \Sigma z_{1\alpha}^2)$ and covariance matrix $\boldsymbol{\Sigma}_{22\cdot1} = \boldsymbol{\Sigma}_{22} - (1/\sigma_{11})\boldsymbol{\sigma}_{(1)}\boldsymbol{\sigma}'_{(1)}$. [*Hint:* The conditional distribution of $Z_\alpha^{(2)}$ given $Z_{1\alpha} = z_{1\alpha}$ is $N[(1/\sigma_{11})\boldsymbol{\sigma}_{(1)} z_{1\alpha}, \boldsymbol{\Sigma}_{22\cdot1}]$. There is an $n \times n$ orthogonal matrix $\boldsymbol{B}$ which carries $(z_{11}, \dots, z_{1n})$ into $(c, \dots, c)$ and $(Z_{i1}, \dots, Z_{in})$ into $(Y_{i1}, \dots, Y_{in}$, $i = 2, \dots, p$. Let the new X'_α be $(Y_{2\alpha}, \dots, Y_{p\alpha})$.]

4.36. (Sec. 4.4) Prove that the noncentrality parameter in the distribution in Problem 4.35 is $(a_{11}/\sigma_{11})\bar{R}^2/(1 - \bar{R}^2)$.

4.37. (Sec. 4.4) Find the distribution of $R^2/(1 - R^2)$ by multiplying the density of Problem 4.35 by the density of a_{11} and integrating with respect to a_{11}.

4.38. (Sec. 4.4) Show that the density of r^2 derived from (38) of Section 4.2 is identical with (42) in Section 4.4 for $p = 2$. [*Hint:* Use the duplication formula for the gamma function.]

4.39. (Sec. 4.4) Prove that (30) is the uniformly most powerful test of $\bar{R} = 0$ based on r. [*Hint:* Use the Neyman–Pearson fundamental lemma.]

4.40. (Sec. 4.4) Prove that (47) is the unique unbiased estimator of $\bar{R}^2$ based on R^2.

4.41. The estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in Problem 3.1 are

$$\bar{x} = (\ 185.72 \quad 151.12 \qquad 183.84 \quad 149.24\ )',$$

$$S = \left(\begin{array}{cc:cc} 95.2933 & 52.8683 & 69.6617 & 46.1117 \\ 52.8683 & 54.3600 & 51.3117 & 35.0533 \\ \hdashline 69.6617 & 51.3117 & 100.8067 & 56.5400 \\ 46.1117 & 35.0533 & 56.5400 & 45.0233 \end{array} \right).$$

(a) Find the estimates of the parameters of the conditional distribution of (x_3, x_4) given (x_1, x_2); that is, find $S_{21}S_{11}^{-1}$ and $S_{22\cdot1} = S_{22} - S_{21}S_{11}^{-1}S_{12}$.

(b) Find the partial correlation $r_{34\cdot12}$.

(c) Use Fisher's z to find a confidence interval for $\rho_{34\cdot12}$ with confidence 0.95.

(d) Find the sample multiple correlation coefficients between x_3 and (x_1, x_2) and between x_4 and (x_1, x_2).

(e) Test the hypotheses that x_3 is independent of (x_1, x_2) and x_4 is independent of (x_1, x_2) at significance levels 0.05.

4.42. Let the components of X correspond to scores on tests in arithmetic speed (X_1), arithmetic power (X_2), memory for words (X_3), memory for meaningful symbols (X_4), and memory for meaningless symbols (X_5). The observed correlations in a sample of 140 are [Kelley (1928)]

$$\begin{pmatrix} 1.0000 & 0.4248 & 0.0420 & 0.0215 & 0.0573 \\ 0.4248 & 1.0000 & 0.1487 & 0.2489 & 0.2843 \\ 0.0420 & 0.1487 & 1.0000 & 0.6693 & 0.4662 \\ 0.0215 & 0.2489 & 0.6693 & 1.0000 & 0.6915 \\ 0.0573 & 0.2843 & 0.4662 & 0.6915 & 1.0000 \end{pmatrix}.$$

(a) Find the partial correlation between X_4 and X_5, holding X_3 fixed.

(b) Find the partial correlation between X_1 and X_2, holding X_3, X_4, and X_5 fixed.

(c) Find the multiple correlation between X_1 and the set X_3, X_4, and X_5.

(d) Test the hypothesis at the 1% significance level that arithmetic speed is independent of the three memory scores.

4.43. (Sec. 4.3) Prove that if $\rho_{ij\cdot q+1,\dots,p}=0$, then $\sqrt{N-2-(p-q)}\,r_{ij\cdot q+1,\dots,p}/\sqrt{1-r^2_{ij\cdot q+1,\dots,p}}$ is distributed according to the t-distribution with $N-2-(p-q)$ degrees of freedom.

4.44. (Sec. 4.3) Let $X'=(X_1,X_2,X^{(2)\prime})$ have the distribution $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$. The conditional distribution of X_1 given $X_2=x_2$ and $X^{(2)}=x^{(2)}$ is

$$N\left[\mu_1+\gamma_2(x_2-\mu_2)+\boldsymbol{\gamma}'(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)}),\sigma_{11\cdot 2,\dots,p}\right],$$

where

$$\begin{pmatrix} \sigma_{22} & \boldsymbol{\sigma}'_{(2)} \\ \boldsymbol{\sigma}_{(2)} & \boldsymbol{\Sigma}_{22} \end{pmatrix}\begin{pmatrix} \gamma_2 \\ \boldsymbol{\gamma} \end{pmatrix}=\begin{pmatrix} \sigma_{12} \\ \boldsymbol{\sigma}_{(1)} \end{pmatrix}.$$

The estimators of γ_2 and $\boldsymbol{\gamma}$ are defined by

$$\begin{pmatrix} a_{22} & \boldsymbol{a}'_{(2)} \\ \boldsymbol{a}_{(2)} & A_{22} \end{pmatrix}\begin{pmatrix} c_2 \\ \boldsymbol{c} \end{pmatrix}=\begin{pmatrix} a_{12} \\ \boldsymbol{a}_{(1)} \end{pmatrix}.$$

Show $c_2=a_{12\cdot 3,\dots,p}/a_{22\cdot 3,\dots,p}$. [*Hint:* Solve for $\boldsymbol{c}$ in terms of c_2 and the a's, and substitute.]

4.45. (Sec. 4.3) In the notation of Problem 4.44, prove

$$\begin{aligned} a_{11\cdot 2,\dots,p} &= a_{11}-\boldsymbol{a}'_{(1)}A_{22}^{-1}\boldsymbol{a}_{(1)}-c_2^2\left(a_{22}-\boldsymbol{a}'_{(2)}A_{22}^{-1}\boldsymbol{a}_{(2)}\right) \\ &= a_{11\cdot 3,\dots,p}-c_2^2a_{22\cdot 3,\dots,p}. \end{aligned}$$

Hint: Use

$$a_{11\cdot 2,\ldots,p} = a_{11} - (c_2 \quad c')\begin{pmatrix} a_{22} & a'_{(2)} \\ a_{(2)} & A_{22} \end{pmatrix}\begin{pmatrix} c_2 \\ c \end{pmatrix}.$$

4.46. (Sec. 4.3) Prove that $1/a_{22\cdot 3,\ldots,p}$ is the element in the upper left-hand corner of

$$\begin{pmatrix} a_{22} & a'_{(2)} \\ a_{(2)} & A_{22} \end{pmatrix}^{-1}.$$

4.47. (Sec. 4.3) Using the results in Problems 4.43–4.46, prove that the test for $\rho_{12\cdot 3,\ldots,p} = 0$ is equivalent to the usual t-test for $\gamma_2 = 0$.

4.48. *Missing observations.* Let $X = (Y' \quad Z')'$, where Y has p components and Z has q components, be distributed according to $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_z \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yz} \\ \boldsymbol{\Sigma}_{zy} & \boldsymbol{\Sigma}_{zz} \end{pmatrix}.$$

Let M observations be made on X, and $N - M$ additional observations be made on Y. Find the maximum likelihood estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. [Anderson (1957).] [*Hint:* Express the likelihood function in terms of the marginal density of Y and the conditional density of Z given Y.]

4.49. Suppose X is distributed according to $N(\mathbf{0}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}.$$

Show that on the basis of one observation, $x' = (x_1, x_2, x_3)$, we can obtain a confidence interval for ρ (with confidence coefficient $1 - \alpha$) by using as endpoints of the interval the solutions in t of

$$\left[x_2^2 + \chi_3^2(\alpha)\right]t^2 - 2(x_1x_2 + x_2x_3)t + x_1^2 + x_2^2 + x_3^2 - \chi_3^2(\alpha) = 0,$$

where $\chi_3^2(\alpha)$ is the significance point of the χ^2-distribution with three degrees of freedom at significance level α.