CHAPTER 8

Testing the General Linear Hypothesis; Multivariate Analysis of Variance

8.1. INTRODUCTION

In this chapter we generalize the univariate least squares theory (i.e., regression analysis) and the analysis of variance to vector variates. The algebra of the multivariate case is essentially the same as that of the univariate case. This leads to distribution theory that is analogous to that of the univariate case and to test criteria that are analogs of F-statistics. In fact, given a univariate test, we shall be able to write down immediately a corresponding multivariate test. Since the analysis of variance based on the model of fixed effects can be obtained from least squares theory, we obtain directly a theory of multivariate analysis of variance. However, in the multivariate case there is more latitude in the choice of tests of significance.

In univariate least squares we consider scalar dependent variates $x_1, \dots, x_N$ drawn from populations with expected values $\boldsymbol{\beta}'z_1, \dots, \boldsymbol{\beta}'z_N$, respectively, where $\boldsymbol{\beta}$ is a column vector of q components and each of the z_α is a column vector of q known components. Under the assumption that the variances in the populations are the same, the least squares estimator of $\boldsymbol{\beta}'$ is

$$b' = \left(\sum_{\alpha=1}^{N} x_\alpha z'_\alpha \right) \left(\sum_{\alpha=1}^{N} z_\alpha z'_\alpha \right)^{-1}. \tag{1}$$

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

If the populations are normal, the vector is the maximum likelihood estimator of $\boldsymbol{\beta}$. The unbiased estimator of the common variance σ^2 is

$$s^2 = \sum_{\alpha=1}^{N} (x_\alpha - b'z_\alpha)^2/(N-q), \tag{2}$$

and under the assumption of normality, the maximum likelihood estimator of σ^2 is $\hat{\sigma}^2 = (N-q)s^2/N$.

In the multivariate case $\boldsymbol{x}_\alpha$ is a vector, $\boldsymbol{\beta}'$ is replaced by a matrix $\mathbf{B}$, and σ^2 is replaced by a covariance matrix $\boldsymbol{\Sigma}$. The estimators of $\mathbf{B}$ and $\boldsymbol{\Sigma}$, given in Section 8.2, are matric analogs of (1) and (2).

To test a hypothesis concerning $\boldsymbol{\beta}$, say the hypothesis $\boldsymbol{\beta} = \mathbf{0}$, we use an F-test. A criterion equivalent to the F-ratio is

$$\frac{1}{[q/(N-q)]F+1} = \frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}, \tag{3}$$

where $\hat{\sigma}_0^2$ is the maximum likelihood estimator of σ^2 under the null hypothesis. We shall find that the likelihood ratio criterion for the corresponding multivariate hypothesis, say $\mathbf{B} = \mathbf{0}$, is the above with the variances replaced by generalized variances. The distribution of the likelihood ratio criterion under the null hypothesis is characterized, the moments are found, and some specific distributions obtained. Satisfactory approximations are given as well as tables of significance points (Appendix B).

The hypothesis testing problem is invariant under several groups of linear transformations. Other invariant criteria are treated, including the Lawley–Hotelling trace, the Bartlett–Nanda–Pillai trace, and the Roy maximum root criteria. Some comparison of power is made.

Confidence regions or simultaneous confidence intervals for elements of $\mathbf{B}$ can be based on the likelihood ratio test, the Lawley–Hotelling trace test, and the Roy maximum root test. Procedures are given explicitly for several problems of the analysis of variance. Optimal properties of admissibility, unbiasedness, and monotonicity of power functions are studied. Finally, the theory and methods are extended to elliptically contoured distributions.

8.2. ESTIMATORS OF PARAMETERS IN MULTIVARIATE LINEAR REGRESSION

8.2.1. Maximum Likelihood Estimators; Least Squares Estimators

Suppose $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N$ are a set of N independent observations, $\boldsymbol{x}_\alpha$ being drawn from $N(\mathbf{B}z_\alpha, \boldsymbol{\Sigma})$. Ordinarily the vectors z_α (with q components) are known ectors, and the $p \times p$ matrix Σ and the $p \times q$ matrix $\boldsymbol{\beta}$ are unknown. We issume $N \geq p+q$ and the rank of

$$Z = (z_1, \dots, z_N) \tag{1}$$

s q. We shall estimate Σ and $\boldsymbol{\beta}$ by the method of maximum likelihood. The ikelihood function is

$$L = (2\pi)^{-\frac{1}{2}Np} |\Sigma^*|^{-\frac{1}{2}N} \exp\left[-\tfrac{1}{2} \sum_{\alpha=1}^{N} (x_\alpha - \boldsymbol{\beta}^* z_\alpha)' \Sigma^{*-1} (x_\alpha - \boldsymbol{\beta}^* z_\alpha)\right]. \tag{2}$$

In (2) the elements of Σ^* and $\boldsymbol{\beta}^*$ are indeterminates. The method of maximum likelihood specifies the estimators of Σ and $\boldsymbol{\beta}$ based on the given sample $x_1, z_1, \dots, x_N, z_N$ as the Σ^* and $\boldsymbol{\beta}^*$ that maximize (2). It is convenient to use the following lemma.

Lemma 8.2.1. *Let*

$$B = \sum_{\alpha=1}^{N} x_\alpha z'_\alpha \left(\sum_{\alpha=1}^{N} z_\alpha z'_\alpha \right)^{-1}. \tag{3}$$

Then for any $p \times q$ matrix F

$$\sum_{\alpha=1}^{N} (x_\alpha - F z_\alpha)(x_\alpha - F z_\alpha)' = \sum_{\alpha=1}^{N} (x_\alpha - B z_\alpha)(x_\alpha - B z_\alpha)' + (B - F) \sum_{\alpha=1}^{N} z_\alpha z'_\alpha (B - F)'. \tag{4}$$

Proof. The left-hand side of (4) is

$$\sum_{\alpha=1}^{N} [(x_\alpha - B z_\alpha) + (B - F) z_\alpha][(x_\alpha - B z_\alpha) + (B - F) z_\alpha]', \tag{5}$$

which is equal to the right-hand side of (4) because

$$\sum_{\alpha=1}^{N} z_\alpha (x_\alpha - B z_\alpha)' = 0 \tag{6}$$

by virtue of (3). ■

The exponential in L is $-\frac{1}{2}$ times

(7)
$$\operatorname{tr} \Sigma^{*-1} \sum_{\alpha=1}^{N} (x_\alpha - \boldsymbol{\beta}^* z_\alpha)(x_\alpha - \boldsymbol{\beta}^* z_\alpha)' = \operatorname{tr} \Sigma^{*-1} \sum_{\alpha=1}^{N} (x_\alpha - B z_\alpha)(x_\alpha - B z_\alpha)' + \operatorname{tr} \Sigma^{*-1} (B - \boldsymbol{\beta}^*) A (B - \boldsymbol{\beta}^*)',$$

where

$$A = \sum_{\alpha=1}^{N} z_\alpha z_\alpha'. \tag{8}$$

The likelihood is maximized with respect to $\boldsymbol{\beta}^*$ by minimizing the last term in (7).

Lemma 8.2.2. *If A and G are positive definite, $\operatorname{tr} FAF'G > 0$ for $F \neq 0$.*

Proof. Let $A = HH'$, $G = KK'$. Then

$$\begin{aligned} \operatorname{tr} FAF'G &= \operatorname{tr} FHH'F'KK' = \operatorname{tr} K'FHH'F'K \\ &= \operatorname{tr}(K'FH)(K'FH)' > 0 \end{aligned} \tag{9}$$

for $F \neq 0$ because then $K'FH \neq 0$ since H and K are nonsingular. ■

It follows from (7) and the lemma that L is maximized with respect to $\boldsymbol{\beta}^*$ by $\boldsymbol{\beta}^* = B$, that is,

$$\hat{\boldsymbol{\beta}} = CA^{-1}, \tag{10}$$

where

$$C = \sum_{\alpha=1}^{N} x_\alpha z_\alpha'. \tag{11}$$

Then by Lemma 3.2.2, L is maximized with respect to Σ^* at

$$\hat{\Sigma} = \frac{1}{N} \sum_{\alpha=1}^{N} (x_\alpha - \hat{\boldsymbol{\beta}} z_\alpha)(x_\alpha - \hat{\boldsymbol{\beta}} z_\alpha)'. \tag{12}$$

This is the multivariate analog of $\hat{\sigma}^2 = (N-q)s^2/N$ defined by (2) of Section 8.1.

Theorem 8.2.1. *If x_α is an observation from $N(\boldsymbol{\beta} z_\alpha, \Sigma)$, $\alpha = 1, \dots, N$, with $(z_1, \dots, z_N)$ of rank q, the maximum likelihood estimator of $\boldsymbol{\beta}$ is given by* (10), *where $C = \sum_\alpha x_\alpha z_\alpha'$ and $A = \sum_\alpha z_\alpha z_\alpha'$. The maximum likelihood estimator of Σ is given by* (12).

A useful algebraic result follows from (12) and (4) with $F=0$:

$$N\hat{\Sigma} = \sum_{\alpha=1}^{N} x_\alpha x'_\alpha - \hat{\beta} A \hat{\beta}' = \sum_{\alpha=1}^{N} x_\alpha x'_\alpha - CA^{-1}C'. \tag{13}$$

Now let us consider a geometric interpretation of the estimation procedure. Let the ith row of $(x_1, \ldots, x_N)$ be x_i^* (with N components) and the ith row of $(z_1, \ldots, z_N)$ be z_i^* (with N components). Then $\sum_j \hat{\beta}_{ij} z_j^*$, being a linear combination of the vectors $z_1^*, \ldots, z_q^*$, is a vector in the q-space spanned by $z_1^*, \ldots, z_q^*$, and is in fact, of all such vectors, the one nearest to x_i^*; hence, it is the projection of x_i^* on the q-space. Thus $x_i^* - \sum_j \hat{\beta}_{ij} z_j^*$ is the vector orthogonal to the q-space going from the projection of x_i^* on the q-space to x_i^*. Translate this vector so that one endpoint is at the origin. Then the set of p vectors $x_1^* - \sum_j \hat{\beta}_{1j} z_j^*, \ldots, x_p^* - \sum_j \hat{\beta}_{pj} z_j^*$ is a set of vectors emanating from the origin. $N\hat{\sigma}_{ii} = (x_i^* - \sum_j \hat{\beta}_{ij} z_j^*)(x_i^* - \sum_j \hat{\beta}_{ij} z_j^*)'$ is the square of the length of the ith such vector, and $N\hat{\sigma}_{ij} = (x_i^* - \sum_h \hat{\beta}_{ih} z_h^*)(x_j^* - \sum_g \hat{\beta}_{jg} z_g^*)'$ is the product of the length of the ith vector, the length of the jth vector, and the cosine of the angle between them.

The equations defining the maximum likelihood estimator of β, namely, $AB' = C'$, consist of p sets of q linear equations in q unknowns. Each set can be solved by the method of pivotal condensation or successive elimination (Section A.5 of the Appendix). The forward solutions are the same (except the right-hand sides) for all sets. Use of (13) to compute $N\hat{\Sigma}$ involves an efficient computation of $\hat{\beta} A \hat{\beta}'$.

Let $X_\alpha = (x_{1\alpha}, \ldots, x_{p\alpha})'$, $\mathbf{B} = (b_1, \ldots, b_p)'$, and $\beta = (\beta_1, \ldots, \beta_p)'$. Then $\mathscr{E} x_{i\alpha} = \beta_i' z_\alpha$ and b_i is the least squares estimator of β_i. If G is a positive definite matrix, then $\operatorname{tr} G \sum_{\alpha=1}^{n} (x_\alpha - Fz_\alpha)(x_\alpha - Fz_\alpha)'$ is minimized by $F = \mathbf{B}$. This is another sense in which $\mathbf{B}$ is the least squares estimator.

8.2.2. Distribution of $\hat{\beta}$ and $\hat{\Sigma}$

Now let us find the joint distribution of $\hat{\beta}_{ig}$ $(i = 1, \ldots, p,\ g = 1, \ldots, q)$. The joint distribution is normal since the $\hat{\beta}_{ig}$ are linear combinations of the $X_{i\alpha}$. From (10) we see that

$$\begin{aligned} \mathscr{E}\hat{\beta} &= \mathscr{E} \sum_{\alpha=1}^{N} X_\alpha z'_\alpha A^{-1} \\ &= \sum_{\alpha=1}^{N} \beta z_\alpha z'_\alpha A^{-1} = \hat{\beta} A A^{-1} \\ &= \beta. \end{aligned} \tag{14}$$

Thus $\hat{\boldsymbol\beta}$ is an *unbiased estimator* of $\boldsymbol\beta$. The covariance between $\hat{\boldsymbol\beta}'_i$ and $\hat{\boldsymbol\beta}'_j$, two rows of $\hat{\boldsymbol\beta}$, is

(15)

$$\begin{aligned}
\mathscr{E}(\hat{\boldsymbol\beta}_i - \boldsymbol\beta_i)(\hat{\boldsymbol\beta}_j - \boldsymbol\beta_j)' &= A^{-1}\mathscr{E}\sum_{\alpha=1}^{N}(X_{i\alpha} - \mathscr{E}X_{i\alpha})z_\alpha \sum_{\gamma=1}^{N}(X_{j\gamma} - \mathscr{E}X_{j\gamma})z'_\gamma A^{-1} \\
&= A^{-1}\sum_{\alpha,\gamma=1}^{N}\mathscr{E}(X_{i\alpha} - \mathscr{E}X_{i\alpha})(X_{j\gamma} - \mathscr{E}X_{j\gamma})z_\alpha z'_\gamma A^{-1} \\
&= A^{-1}\sum_{\alpha,\gamma=1}^{N}\delta_{\alpha\gamma}\sigma_{ij}z_\alpha z'_\gamma A^{-1} \\
&= A^{-1}\sum_{\alpha=1}^{N}\sigma_{ij}z_\alpha z'_\alpha A^{-1} \\
&= \sigma_{ij}A^{-1}AA^{-1} \\
&= \sigma_{ij}A^{-1}.
\end{aligned}$$

To summarize, the vector of pq components $(\hat{\boldsymbol\beta}'_1, \dots, \hat{\boldsymbol\beta}'_p)' = \operatorname{vec}\hat{\boldsymbol\beta}'$ is normally distributed with mean $(\boldsymbol\beta'_1, \dots, \boldsymbol\beta'_p)' = \operatorname{vec}\boldsymbol\beta'$ and covariance matrix

$$(16)\qquad \begin{pmatrix} \sigma_{11}A^{-1} & \sigma_{12}A^{-1} & \cdots & \sigma_{1p}A^{-1} \\ \sigma_{21}A^{-1} & \sigma_{22}A^{-1} & \cdots & \sigma_{2p}A^{-1} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1}A^{-1} & \sigma_{p2}A^{-1} & \cdots & \sigma_{pp}A^{-1} \end{pmatrix}.$$

The matrix (16) is the Kronecker (or direct) product of the matrices $\boldsymbol\Sigma$ and A^{-1}, denoted by $\boldsymbol\Sigma \otimes A^{-1}$.

From Theorem 4.3.3 it follows that $N\hat{\boldsymbol\Sigma} = \sum_{\alpha=1}^{N} x_\alpha x'_\alpha - \hat{\boldsymbol\beta}A\hat{\boldsymbol\beta}'$ is distributed according to $W(\boldsymbol\Sigma, N-q)$. From this we see that an unbiased estimator of $\boldsymbol\Sigma$ is $S = [N/(N-q)]\hat{\boldsymbol\Sigma}$.

Theorem 8.2.2. *The maximum likelihood estimator $\hat{\boldsymbol\beta}$ based on a set of N observations, the αth from $N(\boldsymbol\beta z_\alpha, \boldsymbol\Sigma)$, is normally distributed with mean $\boldsymbol\beta$, and the covariance matrix of the ith and jth rows of $\hat{\boldsymbol\beta}$ is $\sigma_{ij}A^{-1}$, where $A = \sum_\alpha z_\alpha z'_\alpha$. The maximum likelihood estimator $\hat{\boldsymbol\Sigma}$ multiplied by N is independently distributed according to $W(\boldsymbol\Sigma, N-q)$, where q is the number of components of z_α.*

The density then can be written [by virtue of (4)]

$$(17)\qquad \frac{1}{(2\pi)^{\frac{1}{2}pN}|\Sigma|^{\frac{1}{2}N}}\exp\left(-\tfrac{1}{2}\operatorname{tr}\left\{\Sigma^{-1}\left[(\hat{\boldsymbol\beta}-\boldsymbol\beta)A(\hat{\boldsymbol\beta}-\boldsymbol\beta)'+N\hat{\Sigma}\right]\right\}\right).$$

This proves the following:

Corollary 8.2.1. *$\hat{\boldsymbol\beta}$ and $\hat\Sigma$ form a sufficient set of statistics for $\boldsymbol\beta$ and Σ.*

A useful theorem is the following.

Theorem 8.2.3. *Let X_α be distributed according to $N(\boldsymbol\beta z_\alpha, \Sigma)$, $\alpha = 1,\dots,N$, and suppose $X_1,\dots,X_N$ are independent.*

(a) *If $w_\alpha = Hz_\alpha$ and $\Gamma = \boldsymbol\beta H^{-1}$, then X_α is distributed according to $N(\Gamma w_\alpha, \Sigma)$.*
(b) *The maximum likelihood estimator of Γ based on observations x_α on X_α, $\alpha = 1,\dots,N$, is $\hat\Gamma = \hat{\boldsymbol\beta}H^{-1}$, where $\hat{\boldsymbol\beta}$ is the maximum likelihood estimator of $\boldsymbol\beta$.*
(c) *$\hat\Gamma(\Sigma_\alpha w_\alpha w_\alpha')\hat\Gamma' = \hat{\boldsymbol\beta}A\hat{\boldsymbol\beta}'$, where $A = \Sigma_\alpha z_\alpha z_\alpha'$, and the maximum likelihood estimator of $N\Sigma$ is $N\hat\Sigma = \Sigma_\alpha x_\alpha x_\alpha' - \hat\Gamma(\Sigma_\alpha w_\alpha w_\alpha')\hat\Gamma' = \Sigma_\alpha x_\alpha x_\alpha' - \hat{\boldsymbol\beta}A\hat{\boldsymbol\beta}'$.*
(d) *$\hat\Gamma$ and $\hat\Sigma$ are independently distributed.*
(e) *$\hat\Gamma$ is normally distributed with mean Γ and the covariance matrix of the ith and jth rows of $\hat\Gamma$ is $\sigma_{ij}(HAH')^{-1} = \sigma_{ij}H'^{-1}A^{-1}H^{-1}$.*

The proof is left to the reader.

An estimator F is a *linear estimator* of β_{ig} if $F = \sum_{\alpha=1}^N f_\alpha' x_\alpha$. It is a *linear unbiased estimator* of β_{ig} if

$$(18)\qquad \beta_{ig} = \mathscr{E}F = \mathscr{E}\sum_{\alpha=1}^N f_\alpha' x_\alpha = \sum_{\alpha=1}^N f_\alpha'\boldsymbol\beta z_\alpha = \sum_{\alpha=1}^N\sum_{j=1}^p\sum_{h=1}^q f_{j\alpha}\beta_{jh}z_{h\alpha}$$

is an identity in $\boldsymbol\beta$, that is, if

$$(19)\qquad \begin{aligned}\sum_{\alpha=1}^N f_{j\alpha}z_{h\alpha} &= 1, && j=i,\quad h=g,\\ &= 0, && \text{otherwise}.\end{aligned}$$

A linear unbiased estimator is *best* if it has minimum variance over all linear unbiased estimators; that is, if $\mathscr{E}(F-\beta_{ig})^2 \le \mathscr{E}(G-\beta_{ig})^2$ for $G = \sum_{\alpha=1}^N g_\alpha' x_\alpha$ and $\mathscr{E}G = \beta_{ig}$.

Theorem 8.2.4. *The least squares estimator is the best linear unbiased estimator of* β_{ig}.

Proof. Let $\tilde{\beta}_{ig} = \sum_{\alpha=1}^{N} \sum_{j=1}^{p} f_{j\alpha} x_{j\alpha}$ be an arbitrary unbiased estimator of β_{ig}, and let $\hat{\beta}_{ig} = \sum_{\alpha=1}^{N} \sum_{h=1}^{q} x_{i\alpha} z_{h\alpha} a^{hg}$ be the least squares estimator, where $A = \sum_{\alpha=1}^{N} z_\alpha z'_\alpha$. Then

(20)

$$\begin{aligned} \mathscr{E}\left(\tilde{\beta}_{ig} - \beta_{ig}\right)^2 &= \mathscr{E}\left[\hat{\beta}_{ig} - \beta_{ig} + \left(\tilde{\beta}_{ig} - \hat{\beta}_{ig}\right)\right]^2 \\ &= \mathscr{E}\left(\hat{\beta}_{ig} - \beta_{ig}\right)^2 + 2\,\mathscr{E}\left(\hat{\beta}_{ig} - \beta_{ig}\right)\left(\tilde{\beta}_{ig} - \hat{\beta}_{ig}\right) + \mathscr{E}\left(\tilde{\beta}_{ig} - \hat{\beta}_{ig}\right)^2. \end{aligned}$$

Because $\tilde{\beta}_{ig}$ and $\hat{\beta}_{ig}$ are unbiased, $\tilde{\beta}_{ig} - \beta_{ig} = \sum_{\alpha=1}^{N} \sum_{j=1}^{p} f_{j\alpha} u_{j\alpha}$, $\hat{\beta}_{ig} - \beta_{ig} = \sum_{\alpha=1}^{N} \sum_{h=1}^{q} u_{i\alpha} z_{h\alpha} a^{hg}$, and

$$\tilde{\beta}_{ig} - \hat{\beta}_{ig} = \sum_{\alpha=1}^{N} \sum_{j=1}^{p} \left(f_{j\alpha}^{ig} - \delta_{ij} \sum_{h=1}^{q} z_{h\alpha} a^{hg} \right) u_{j\alpha}, \tag{21}$$

where $\delta_{ii} = 1$ and $\delta_{ij} = 0$, $i \neq j$. Then

$$\begin{aligned} &\mathscr{E}\left(\hat{\beta}_{ig} - \beta_{ig}\right)\left(\tilde{\beta}_{ig} - \hat{\beta}_{ig}\right) \\ &\quad = \mathscr{E} \sum_{\alpha,\gamma=1}^{N} \sum_{h=1}^{q} z_{h\alpha} a^{hg} u_{i\alpha} \sum_{j=1}^{p} \left(f_{j\gamma} - \delta_{ij} \sum_{h'=1}^{q} z_{h'\gamma} a^{h'g} \right) u_{j\gamma} \\ &\quad = \sum_{\alpha=1}^{N} \sum_{h=1}^{q} \sum_{j=1}^{p} z_{h\alpha} a^{hg} \left(f_{j\alpha} - \delta_{ij} \sum_{h'=1}^{q} z_{h'\alpha} a^{h'g} \right) \sigma_{ij} \\ &\quad = \sigma_{ii} a^{gg} - \sigma_{ii} \sum_{h=1}^{q} \sum_{h'=1}^{q} a_{hh'} a^{hg} a^{h'g} \\ &\quad = 0. \end{aligned} \tag{22}$$

Then (20) implies $\mathscr{E}(\tilde{\beta}_{ig} - \beta_{ig})^2 \geq \mathscr{E}(\hat{\beta}_{ig} - \beta_{ig})^2$. ■

8.3. LIKELIHOOD RATIO CRITERIA FOR TESTING LINEAR HYPOTHESES ABOUT REGRESSION COEFFICIENTS

8.3.1. Likelihood Ratio Criteria

Suppose we partition

$$\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1 \quad \boldsymbol{\beta}_2 \right) \tag{1}$$

so that $\boldsymbol{\beta}_1$ has q_1 columns and $\boldsymbol{\beta}_2$ has q_2 columns. We shall derive the likelihood ratio criterion for testing the hypothesis

$$H: \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^*, \tag{2}$$

where $\boldsymbol{\beta}_1^*$ is a given matrix. The maximum of the likelihood function L for the sample $x_1, \dots, x_N$ is

$$\max_{\boldsymbol{\beta}, \boldsymbol{\Sigma}} L = (2\pi)^{-\frac{1}{2}pN} |\hat{\boldsymbol{\Sigma}}_\Omega|^{-\frac{1}{2}N} e^{-\frac{1}{2}pN}, \tag{3}$$

where $\hat{\boldsymbol{\Sigma}}_\Omega$ is given by (12) or (13) of Section 8.2.

To find the maximum of the likelihood function for the parameters restricted to ω defined by (2) we let

$$y_\alpha = x_\alpha - \boldsymbol{\beta}_1^* z_\alpha^{(1)}, \qquad \alpha = 1, \dots, N, \tag{4}$$

where

$$z_\alpha = \begin{pmatrix} z_\alpha^{(1)} \\ z_\alpha^{(2)} \end{pmatrix}, \qquad \alpha = 1, \dots, N, \tag{5}$$

is partitioned in a manner corresponding to the partitioning of $\boldsymbol{\beta}$. Then y_α can be considered as an observation from $N(\boldsymbol{\beta}_2 z_\alpha^{(2)}, \boldsymbol{\Sigma})$. The estimator of $\boldsymbol{\beta}_2$ is obtained by the procedure of Section 8.2 as

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{2\omega} &= \sum_{\alpha=1}^N y_\alpha z_\alpha^{(2)\prime} A_{22}^{-1} = \sum_{\alpha=1}^N \left(x_\alpha - \boldsymbol{\beta}_1^* z_\alpha^{(1)}\right) z_\alpha^{(2)\prime} A_{22}^{-1} \\ &= (C_2 - \boldsymbol{\beta}_1^* A_{12}) A_{22}^{-1} \end{aligned} \tag{6}$$

with C and A partitioned in the manner corresponding to the partitioning of $\boldsymbol{\beta}$ and z_α,

$$C = \begin{pmatrix} C_1 & C_2 \end{pmatrix}, \tag{7}$$

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}. \tag{8}$$

The estimator of $\boldsymbol{\Sigma}$ is given by

$$\begin{aligned} N\hat{\boldsymbol{\Sigma}}_\omega &= \sum_{\alpha=1}^N \left(y_\alpha - \hat{\boldsymbol{\beta}}_{2\omega} z_\alpha^{(2)}\right)\left(y_\alpha - \hat{\boldsymbol{\beta}}_{2\omega} z_\alpha^{(2)}\right)' \\ &= \sum_{\alpha=1}^N y_\alpha y_\alpha' - \hat{\boldsymbol{\beta}}_{2\omega} A_{22} \hat{\boldsymbol{\beta}}_{2\omega}' \\ &= \sum_{\alpha=1}^N \left(x_\alpha - \boldsymbol{\beta}_1^* z_\alpha^{(1)}\right)\left(x_\alpha - \boldsymbol{\beta}_1^* z_\alpha^{(1)}\right)' - \hat{\boldsymbol{\beta}}_{2\omega} A_{22} \hat{\boldsymbol{\beta}}_{2\omega}'. \end{aligned} \tag{9}$$

Thus the maximum of the likelihood function over ω is

$$\max_{\boldsymbol{\beta}_2, \boldsymbol{\Sigma}} L = (2\pi)^{-\frac{1}{2}pN} |\hat{\boldsymbol{\Sigma}}_\omega|^{-\frac{1}{2}N} e^{-\frac{1}{2}pN}. \tag{10}$$

The likelihood ratio criterion for testing H is (10) divided by (3), namely,

$$\lambda = \frac{|\hat{\boldsymbol{\Sigma}}_\Omega|^{\frac{1}{2}N}}{|\hat{\boldsymbol{\Sigma}}_\omega|^{\frac{1}{2}N}}. \tag{11}$$

In testing H, one rejects the hypothesis if $\lambda < \lambda_0$, where λ_0 is a suitably chosen number.

A special case of this problem led to Hotelling's T^2-criterion. If $q = q_1 = 1$ $(q_2 = 0)$, $z_\alpha = 1$, $\alpha = 1, \dots, N$, and $\boldsymbol{\beta} = \boldsymbol{\beta}_1 = \boldsymbol{\mu}$, then the T^2-criterion for testing the hypothesis $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ is a monotonic function of (11) for $\boldsymbol{\beta}_1^* = \boldsymbol{\mu}_0$.

The hypothesis $\boldsymbol{\mu} = \mathbf{0}$ and the T^2-statistic are invariant with respect to the transformations $X^* = DX$ and $x_\alpha^* = Dx_\alpha$, $\alpha = 1, \dots, N$, for nonsingular D. Similarly, in this problem the null hypothesis $\boldsymbol{\beta}_1 = \mathbf{0}$ and the likelihood ratio criterion for testing it are invariant with respect to nonsingular linear transformations.

Theorem 8.3.1. *The likelihood ratio criterion* (11) *for testing the null hypothesis $\boldsymbol{\beta}_1 = \mathbf{0}$ is invariant with respect to transformations $x_\alpha^* = Dx_\alpha$, $\alpha = 1, \dots, N$, for nonsingular D.*

Proof. The estimators in terms of x_α^* are

$$\hat{\boldsymbol{\beta}}^* = DCA^{-1} = D\hat{\boldsymbol{\beta}}, \tag{12}$$

$$\hat{\boldsymbol{\Sigma}}_\Omega^* = \frac{1}{N}\sum_{\alpha=1}^N (Dx_\alpha - D\hat{\boldsymbol{\beta}}z_\alpha)(Dx_\alpha - D\hat{\boldsymbol{\beta}}z_\alpha)' = D\hat{\boldsymbol{\Sigma}}_\Omega D', \tag{13}$$

$$\hat{\boldsymbol{\beta}}_{2\omega}^* = DC_2A_{22}^{-1} = D\hat{\boldsymbol{\beta}}_{2\omega}, \tag{14}$$

$$\hat{\boldsymbol{\Sigma}}_\omega^* = \frac{1}{N}\sum_{\alpha=1}^N (Dx_\alpha - D\hat{\boldsymbol{\beta}}_{2\omega}z_\alpha^{(2)})(Dx_\alpha - D\hat{\boldsymbol{\beta}}_{2\omega}z_\alpha^{(2)})' = D\hat{\boldsymbol{\Sigma}}_\omega D'. \tag{15}$$

■

8.3.2. Geometric Interpretation

An insight into the algebra developed here can be given in terms of a geometric interpretation. It will be convenient to use the following lemma:

Lemma 8.3.1.

$$\hat{\boldsymbol{\beta}}_{2\omega} - \hat{\boldsymbol{\beta}}_{2\Omega} = (\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)A_{12}A_{22}^{-1}. \tag{16}$$

Proof. The normal equation $\hat{\boldsymbol{\beta}}_\Omega A = C$ is written in partitioned form

$$(17) \qquad \left(\hat{\boldsymbol{\beta}}_{1\Omega} A_{11} + \hat{\boldsymbol{\beta}}_{2\Omega} A_{21}, \hat{\boldsymbol{\beta}}_{1\Omega} A_{12} + \hat{\boldsymbol{\beta}}_{2\Omega} A_{22}\right) = (C_1, C_2).$$

Thus $\hat{\boldsymbol{\beta}}_{2\Omega} = C_2 A_{22}^{-1} - \hat{\boldsymbol{\beta}}_{1\Omega} A_{12} A_{22}^{-1}$. The lemma follows by comparison with (6). ■

We can now write

$$\begin{aligned}(18) \qquad X - \boldsymbol{\beta} Z &= \left(X - \hat{\boldsymbol{\beta}}_\Omega Z\right) + \left(\hat{\boldsymbol{\beta}}_{2\Omega} - \boldsymbol{\beta}_2\right) Z_2 + \left(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*\right) Z_1 \\ &= \left(X - \hat{\boldsymbol{\beta}}_\Omega Z\right) + \left(\hat{\boldsymbol{\beta}}_{2\omega} - \boldsymbol{\beta}_2\right) Z_2 \\ &\quad - \left(\hat{\boldsymbol{\beta}}_{2\omega} - \hat{\boldsymbol{\beta}}_{2\Omega}\right) Z_2 + \left(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*\right) Z_1 \\ &= \left(X - \hat{\boldsymbol{\beta}}_\Omega Z\right) + \left(\hat{\boldsymbol{\beta}}_{2\omega} - \boldsymbol{\beta}_2\right) Z_2 \\ &\quad + \left(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*\right)\left(Z_1 - A_{12} A_{22}^{-1} Z_2\right)\end{aligned}$$

as an identity; here $X = (x_1, \dots, x_N)$, $Z_1 = (z_1^{(1)}, \dots, z_N^{(1)})$, and $Z_2 = (z_1^{(2)}, \dots, z_N^{(2)})$. The rows of $Z = (Z_1', Z_2')'$ span a q-dimensional subspace in N-space. Each row of $\boldsymbol{\beta} Z$ is a vector in the q-space, and hence each row of $X - \boldsymbol{\beta} Z$ is a vector from a vector in the q-space to the corresponding row vector of X. Each row vector of $X - \boldsymbol{\beta} Z$ is expressed above as the sum of three row vectors. The first matrix on the right of (18) has as its ith row a vector orthogonal to the q-space and leading to the ith row vector of X (as shown in the preceding section). The row vectors of $(\hat{\boldsymbol{\beta}}_{2\omega} - \boldsymbol{\beta}_2) Z_2$ are vectors in the q_2-space spanned by the rows of Z_2 (since they are linear combinations of the rows of Z_2). The row vectors of $(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)(Z_1 - A_{12} A_{22}^{-1} Z_2)$ are vectors in the q_1-space of $Z_1 - A_{12} A_{22}^{-1} Z_2$, and this space is in the q-space of Z, but orthogonal to the q_2-space of Z_2 [since $(Z_1 - A_{12} A_{22}^{-1} Z_2) Z_2' = 0$]. Thus each row of $X - \boldsymbol{\beta} Z$ is indicated in Figure 8.1 as the sum of three orthogonal vectors: one vector is in the space orthogonal to Z, one is in the space of Z_2, and one is in the subspace of Z that is orthogonal to Z_2.

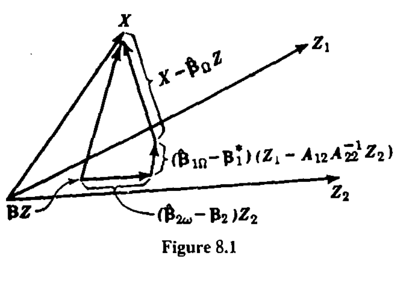

Figure 8.1

From the orthogonality relations we have

$$
\begin{aligned}
(19)\quad (X-\boldsymbol{\beta}Z)(X-\boldsymbol{\beta}Z)'
&= (X-\hat{\boldsymbol{\beta}}_\Omega Z)(X-\hat{\boldsymbol{\beta}}_\Omega Z)' + (\hat{\boldsymbol{\beta}}_{2\omega}-\boldsymbol{\beta}_2)Z_2Z_2'(\hat{\boldsymbol{\beta}}_{2\omega}-\boldsymbol{\beta}_2)' \\
&\quad + (\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)(Z_1-A_{12}A_{22}^{-1}Z_2)(Z_1-A_{12}A_{22}^{-1}Z_2)'(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)' \\
&= N\hat{\Sigma}_\Omega + (\hat{\boldsymbol{\beta}}_{2\omega}-\boldsymbol{\beta}_2)A_{22}(\hat{\boldsymbol{\beta}}_{2\omega}-\boldsymbol{\beta}_2)' \\
&\quad + (\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)(A_{11}-A_{12}A_{22}^{-1}A_{21})(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)'.
\end{aligned}
$$

If we subtract $(\hat{\boldsymbol{\beta}}_{2\omega}-\boldsymbol{\beta}_2)Z_2$ from both sides of (18), we have

$$
(20)\quad X-\boldsymbol{\beta}_1^*Z_1-\hat{\boldsymbol{\beta}}_{2\omega}Z_2 = (X-\hat{\boldsymbol{\beta}}_\Omega Z) + (\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)(Z_1-A_{12}A_{22}^{-1}Z_2).
$$

From this we obtain

$$
\begin{aligned}
(21)\quad N\hat{\Sigma}_\omega &= (X-\boldsymbol{\beta}_1^*Z_1-\hat{\boldsymbol{\beta}}_{2\omega}Z_2)(X-\boldsymbol{\beta}_1^*Z_1-\hat{\boldsymbol{\beta}}_{2\omega}Z_2)' \\
&= (X-\hat{\boldsymbol{\beta}}_\Omega Z)(X-\hat{\boldsymbol{\beta}}_\Omega Z)' \\
&\quad + (\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)(Z_1-A_{12}A_{22}^{-1}Z_2)(Z_1-A_{12}A_{22}^{-1}Z_2)'(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)' \\
&= N\hat{\Sigma}_\Omega + (\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)(A_{11}-A_{12}A_{22}^{-1}A_{21})(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)'.
\end{aligned}
$$

The determinant $|\hat{\Sigma}_\Omega| = (1/N^p)|(X-\hat{\boldsymbol{\beta}}_\Omega Z)(X-\hat{\boldsymbol{\beta}}_\Omega Z)'|$ is proportional to the volume squared of the parallelotope spanned by the row vectors of $X-\hat{\boldsymbol{\beta}}_\Omega Z$ (translated to the origin). The determinant $|\hat{\Sigma}_\omega| = (1/N^p)|(X-\boldsymbol{\beta}_1^*Z_1-\hat{\boldsymbol{\beta}}_{2\omega}Z_2)(X-\boldsymbol{\beta}_1^*Z_1-\hat{\boldsymbol{\beta}}_{2\omega}Z_2)'|$ is proportional to the volume squared of the parallelotope spanned by the row vectors of $X-\boldsymbol{\beta}_1^*Z_1-\hat{\boldsymbol{\beta}}_{2\omega}Z_2$ (translated to the origin); each of these vectors is the part of the vector of $X-\boldsymbol{\beta}_1^*Z_1$ that is orthogonal to Z_2. Thus the test based on the likelihood ratio criterion depends on the ratio of volumes of parallelotopes. One parallelotope involves vectors orthogonal to Z, and the other involves vectors orthogonal to Z_2.

From (15) we see that the density of $x_1,\dots,x_N$ can be written as

$$
\begin{aligned}
(22)\quad \frac{1}{(2\pi)^{\frac{1}{2}pN}|\Sigma|^{\frac{1}{2}N}}\exp\Big(-\tfrac{1}{2}\operatorname{tr}\Big\{\Sigma^{-1}\Big[N\hat{\Sigma} &+ (\hat{\boldsymbol{\beta}}_{2\omega}-\boldsymbol{\beta}_2)A_{22}(\hat{\boldsymbol{\beta}}_{2\omega}-\boldsymbol{\beta}_2)' \\
&+ (\hat{\boldsymbol{\beta}}_{1\omega}-\boldsymbol{\beta}_1^*)(A_{11}-A_{12}A_{22}^{-1}A_{21})(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)'\Big]\Big\}\Big).
\end{aligned}
$$

Thus, $\hat{\Sigma}$, $\hat{\boldsymbol{\beta}}_{1\Omega}$, and $\hat{\boldsymbol{\beta}}_{2\omega}$ form a sufficient set of statistics for Σ, $\boldsymbol{\beta}_1$, and $\boldsymbol{\beta}_2$.

Wilks (1932) first gave the likelihood ratio criterion for testing the equality of mean vectors from several populations (Section 8.8). Wilks (1934) and Bartlett (1934) extended its use to regression coefficients.

8.3.3. The Canonical Form

In studying the distributions of criteria it will be convenient to put the distribution of the observations in canonical form. This amounts to picking a coordinate system in the N-dimensional space so that the first q_1 coordinate axes are in the space of $\mathbf{Z}$ that is orthogonal to $\mathbf{Z}_2$, the next q_2 coordinate axes are in the space of $\mathbf{Z}_2$, and the last n $(=N-q)$ coordinate axes are orthogonal to the $\mathbf{Z}$-space.

Let $\boldsymbol{P}_2$ be a $q_2 \times q_2$ matrix such that

$$I = P_2 A_{22} P_2' = (P_2 Z_2)(P_2 Z_2)', \tag{23}$$

and let $\boldsymbol{P}_1$ be a $q_1 \times q_1$ matrix such that $(A_{11\cdot 2} = A_{11} - A_{12}A_{22}^{-1}A_{21})$

$$I = P_1 A_{11\cdot 2} P_1' = \left[P_1\left(Z_1 - A_{12}A_{22}^{-1}Z_2\right)\right]\left[P_1\left(Z_1 - A_{12}A_{22}^{-1}Z_2\right)\right]'. \tag{24}$$

Then define the $N \times N$ orthogonal matrix $\boldsymbol{Q}$ as

$$Q = \begin{pmatrix} Q_1 \\ Q_2 \\ Q_3 \end{pmatrix} = \begin{pmatrix} P_1\left(Z_1 - A_{12}A_{22}^{-1}Z_2\right) \\ P_2 Z_2 \\ Q_3 \end{pmatrix}, \tag{25}$$

where $\boldsymbol{Q}_3$ is any $n \times N$ matrix making $\boldsymbol{Q}$ orthogonal. Then the columns of

$$W = \begin{pmatrix} W_1 & W_2 & W_3 \end{pmatrix} = XQ' = X\begin{pmatrix} Q_1' & Q_2' & Q_3' \end{pmatrix} \tag{26}$$

are independently normally distributed with covariance matrix $\boldsymbol{\Sigma}$ (Theorem 3.3.1). Then

$$\begin{aligned} \mathscr{E}W_1 &= \mathscr{E}XQ_1' = (\boldsymbol{\beta}_1 Z_1 + \boldsymbol{\beta}_2 Z_2)\left(Z_1 - A_{12}A_{22}^{-1}Z_2\right)'P_1' \\ &= \boldsymbol{\beta}_1 A_{11\cdot 2} P_1' = \boldsymbol{\beta}_1 P_1^{-1}, \end{aligned} \tag{27}$$

$$\begin{aligned} \mathscr{E}W_2 &= \mathscr{E}XQ_2' = (\boldsymbol{\beta}_1 Z_1 + \boldsymbol{\beta}_2 Z_2)Z_2'P_2' \\ &= (\boldsymbol{\beta}_1 A_{12} + \boldsymbol{\beta}_2 A_{22})P_2', \end{aligned} \tag{28}$$

$$\mathscr{E}W_3 = \mathscr{E}XQ_3' = \boldsymbol{\beta} Z Q_3' = \mathbf{0}. \tag{29}$$

Let

$$\Gamma_1 = (\gamma_1, \ldots, \gamma_{q_1}) = \boldsymbol{\beta}_1 A_{11\cdot 2} P_1' = \boldsymbol{\beta}_1 P_1^{-1}, \tag{30}$$

$$\Gamma_2 = (\gamma_{q_1+1}, \ldots, \gamma_q) = (\boldsymbol{\beta}_1 A_{12} + \boldsymbol{\beta}_2 A_{22}) P_2', \tag{31}$$

$$W = \begin{pmatrix} W_1 & W_2 & W_3 \end{pmatrix} = (w_1, \ldots, w_{q_1}, w_{q_1+1}, \ldots, w_q, w_{q+1}, \ldots, w_N). \tag{32}$$

Then $w_1, \ldots, w_N$ are independently normally distributed with covariance matrix $\boldsymbol{\Sigma}$ and $\mathscr{E} w_\alpha = \gamma_\alpha$, $\alpha = 1, \ldots, q$, and $\mathscr{E} w_\alpha = \mathbf{0}$, $\alpha = q+1, \ldots, N$.

The hypothesis $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^*$ can be transformed to $\boldsymbol{\beta}_1 = \mathbf{0}$ by subtraction, that is, by letting $x_\alpha - \boldsymbol{\beta}_1^* z_\alpha^{(1)} = y_\alpha$, as in Section 8.3.1. In canonical form then, the hypothesis is $\Gamma_1 = \mathbf{0}$. We can study problems in the canonical form, if we wish, and transform solutions back to terms of X and Z.

In (17), which is the partitioned form of $\hat{\boldsymbol{\beta}}_\Omega A = C$, eliminate $\hat{\boldsymbol{\beta}}_{2\Omega}$ to obtain

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{1\Omega}(A_{11} - A_{12} A_{22}^{-1} A_{21}) &= C_1 - C_2 A_{22}^{-1} A_{21} \\ &= X(Z_1' - Z_2' A_{22}^{-1} A_{21}) \\ &= W_1 P_1'^{-1}; \end{aligned} \tag{33}$$

that is, $W_1 = \hat{\boldsymbol{\beta}}_{1\Omega} A_{11\cdot 2} P_1' = \hat{\boldsymbol{\beta}}_{1\Omega} P_1^{-1}$ and $\Gamma_1 = \boldsymbol{\beta}_1 P_1^{-1}$. Similarly, from (6) we obtain

$$\hat{\boldsymbol{\beta}}_{2\omega} A_{22} + \boldsymbol{\beta}_1^* A_{12} = C_2 = X Z_2' = W_2 P_2'^{-1}; \tag{34}$$

that is, $W_2 = (\hat{\boldsymbol{\beta}}_{2\omega} A_{22} + \boldsymbol{\beta}_1^* A_{12}) P_2' = \hat{\boldsymbol{\beta}}_{2\omega} P_2^{-1} + \boldsymbol{\beta}_1^* A_{12} P_2^{-1}$ and $\Gamma_2 = \boldsymbol{\beta}_2 P_2^{-1} + \boldsymbol{\beta}_1 A_{12} P_2^{-1}$.

8.4. THE DISTRIBUTION OF THE LIKELIHOOD RATIO CRITERION WHEN THE HYPOTHESIS IS TRUE

8.4.1. Characterization of the Distribution

The likelihood ratio criterion is the $\frac{1}{2}N$th power of

$$U = \lambda^{2/N} = \frac{|\hat{\boldsymbol{\Sigma}}_\Omega|}{|\hat{\boldsymbol{\Sigma}}_\omega|} = \frac{|N\hat{\boldsymbol{\Sigma}}_\Omega|}{\left|N\hat{\boldsymbol{\Sigma}}_\Omega + (\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*) A_{11\cdot 2} (\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)'\right|}, \tag{1}$$

where $A_{11\cdot 2} = A_{11} - A_{12} A_{22}^{-1} A_{21}$. We shall study the distribution and the moments of U when $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^*$. It has been shown in Section 8.2 that $N\hat{\boldsymbol{\Sigma}}_\Omega$ is distributed according to $W(\boldsymbol{\Sigma}, n)$, where $n = N - q$, and the elements of $\hat{\boldsymbol{\beta}}_\Omega - \boldsymbol{\beta}$ have a joint normal distribution independent of $N\hat{\boldsymbol{\Sigma}}_\Omega$.

From (33) of Section 8.3, we have

$$(2)\qquad (\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)A_{11\cdot 2}(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)' = (W_1 - \boldsymbol{\Gamma}_1)P_1 A_{11\cdot 2} P_1'(W_1 - \boldsymbol{\Gamma}_1)' = (W_1 - \boldsymbol{\Gamma}_1)(W_1 - \boldsymbol{\Gamma}_1)',$$

by (24) of Section 8.3; the columns of $W_1 - \boldsymbol{\Gamma}_1$ are independently distributed, each according to $N(\mathbf{0}, \boldsymbol{\Sigma})$.

Lemma 8.4.1. $(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)A_{11\cdot 2}(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)'$ *is distributed according to* $W(\boldsymbol{\Sigma}, q_1)$.

Lemma 8.4.2. *The criterion U has the distribution of*

$$(3)\qquad U = \frac{|\boldsymbol{G}|}{|\boldsymbol{G} + \boldsymbol{H}|},$$

where $\boldsymbol{G}$ is distributed according to $W(\boldsymbol{\Sigma}, n)$, $\boldsymbol{H}$ is distributed according to $W(\boldsymbol{\Sigma}, m)$, where $m = q_1$, and $\boldsymbol{G}$ and $\boldsymbol{H}$ are independent.

Let

$$(4)\qquad \boldsymbol{G} = N\hat{\boldsymbol{\Sigma}}_\Omega = XX' - XZ'(ZZ')^{-1}ZX',$$

$$(5)\qquad \boldsymbol{G} + \boldsymbol{H} = N\hat{\boldsymbol{\Sigma}}_\Omega + (\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)A_{11\cdot 2}(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)' = N\hat{\boldsymbol{\Sigma}}_\omega = YY' - YZ_2'(Z_2Z_2')^{-1}Z_2Y',$$

where $Y = X - \boldsymbol{\beta}_1^* Z_1 = X - (\boldsymbol{\beta}_1^*\ \ \mathbf{0})Z$. Then

$$(6)\qquad \boldsymbol{G} = YY' - YZ'(ZZ')^{-1}ZY'.$$

We shall denote this criterion as $U_{p,m,n}$, where p is the dimensionality, $m = q_1$ is the number of columns of $\boldsymbol{\beta}_1$, and $n = N - q$ is the number of degrees of freedom of $\boldsymbol{G}$.

We now proceed to characterize the distribution of U as the product of beta variables (Section 5.2). Write the criterion U as

$$(7)\qquad U = V_1 V_2 \cdots V_p,$$

where $V_1 = g_{11}/(g_{11} + h_{11})$,

$$(8)\qquad V_i = \frac{|\boldsymbol{G}_i|}{|\boldsymbol{G}_{i-1}|} \Bigg/ \frac{|\boldsymbol{G}_i + \boldsymbol{H}_i|}{|\boldsymbol{G}_{i-1} + \boldsymbol{H}_{i-1}|}, \qquad i = 2, \dots, p,$$

and G_i and H_i are the submatrices of G and H, respectively, of the first i rows and columns. Correspondingly, let $y_\alpha^{(i)}$ consist of the first i components of $y_\alpha = x_\alpha - \boldsymbol{\beta}_1^* z_\alpha^{(1)}$, $\alpha = 1, \dots, N$. We shall show that V_i is the length squared of the vector from $y_i^* = (y_{i1}, \dots, y_{iN})$ to its projection on Z and $Y_{i-1} = (y_1^{(i-1)}, \dots, y_N^{(i-1)})$ divided by the length squared of the vector from y_i^* to its projection on Z_2 and Y_{i-1}.

Lemma 8.4.3. *Let y be an N-component row vector and U an $r \times N$ matrix. Then the sum of squares of the residuals of y from its regression on U is*

$$\frac{\begin{vmatrix} yy' & yU' \\ Uy' & UU' \end{vmatrix}}{|UU'|}. \tag{9}$$

Proof. By Corollary A.3.1 of the Appendix, (9) is $yy' - yU'(UU')^{-1}Uy'$, which is the sum of squares of residuals as indicated in (13) of Section 8.2. ∎

Lemma 8.4.4. *V_i defined by (8) is the ratio of the sum of squares of the residuals of $y_{i1}, \dots, y_{iN}$ from their regression on $y_1^{(i-1)}, \dots, y_N^{(i-1)}$ and Z to the sum of squares of residuals of $y_{i1}, \dots, y_{iN}$ from their regression on $y_1^{(i-1)}, \dots, y_N^{(i-1)}$ and Z_2.*

Proof. The numerator of V_i can be written [from (13) of Section 8.2]

$$\frac{|G_i|}{|G_{i-1}|} = \frac{|Y_i Y_i' - Y_i Z'(ZZ')^{-1} Z Y_i'|}{|Y_{i-1} Y_{i-1}' - Y_{i-1} Z'(ZZ')^{-1} Z Y_{i-1}'|} \tag{10}$$

$$= \frac{\begin{vmatrix} Y_i Y_i' & Y_i Z' \\ Z Y_i' & ZZ' \end{vmatrix} \Big/ |ZZ'|}{\begin{vmatrix} Y_{i-1} Y_{i-1}' & Y_{i-1} Z' \\ Z Y_{i-1}' & ZZ' \end{vmatrix} \Big/ |ZZ'|}$$

$$= \frac{\begin{vmatrix} Y_{i-1} Y_{i-1}' & Y_i y_i^{*\prime} & Y_{i-1} Z' \\ y_i^* Y_{i-1}' & y_i^* y_i^{*\prime} & y_i^* Z' \\ Z Y_{i-1}' & Z y_i^{*\prime} & ZZ' \end{vmatrix}}{\begin{vmatrix} Y_{i-1} Y_{i-1}' & Y_{i-1} Z' \\ Z Y_{i-1}' & ZZ' \end{vmatrix}}$$

$$= \frac{\begin{vmatrix} y_i^* y_i^{*\prime} & y_i^* \begin{bmatrix} Y_{i-1}' & Z' \end{bmatrix} \\ \begin{bmatrix} Y_{i-1} \\ Z \end{bmatrix} y_i^* & \begin{bmatrix} Y_{i-1} \\ Z \end{bmatrix} \begin{bmatrix} Y_{i-1}' & Z' \end{bmatrix} \end{vmatrix}}{\left| \begin{bmatrix} Y_{i-1} \\ Z \end{bmatrix} \begin{bmatrix} Y_{i-1}' & Z' \end{bmatrix} \right|}$$

$$= y_i^* y_i^{*\prime} - y_i^* \begin{pmatrix} Y_{i-1}' & Z' \end{pmatrix} \begin{bmatrix} Y_{i-1} Y_{i-1}' & Y_{i-1} Z' \\ Z Y_{i-1}' & Z Z' \end{bmatrix}^{-1} \begin{pmatrix} Y_{i-1} \\ Z \end{pmatrix} y_i^{*\prime}$$

by Corollary A.3.1. Application of Lemma 8.4.3 shows that the right-hand side of (10) is the sum of squares of the residuals of y_i^* on Y_{i-1} and Z. The denominator is evaluated similarly with Z replaced by Z_2. ■

The ratio V_i is the $2/N$th power of the likelihood ratio criterion for testing the hypothesis that the regression of $y_i^* = x_i^* - \beta_{i1}^* Z_1$ on Z_1 is $\mathbf{0}$ (in the presence of regression on Y_{i-1} and Z_2); here β_{i1}^* is the ith row of β_1^*. For $i = 1$, g_{11} is the sum of squares of the residuals of $y_1^* = (y_{11}, \dots, y_{1N})$ from its regression on Z, and $g_{11} + h_{11}$ is the sum of squares of the residuals from Z_2. The ratio $V_1 = g_{11}/(g_{11} + h_{11})$, which is approximate to test the hypothesis that regression of y_1^* on Z_1 is $\mathbf{0}$, is distributed as $\chi_n^2/(\chi_n^2 + \chi_m^2)$ (by Lemma 8.4.2) and has the beta distribution $\beta(v; \frac{1}{2}n, \frac{1}{2}m)$. (See Section 5.2, for example.) Thus V_i has the beta density

$$\beta\left[v; \tfrac{1}{2}(n+1-i), \tfrac{1}{2}m\right] = \frac{\Gamma\left[\frac{1}{2}(n+m+1-i)\right]}{\Gamma\left[\frac{1}{2}(n+1-i)\right]\Gamma\left(\frac{1}{2}m\right)} v^{\frac{1}{2}(n+1-i)-1}(1-v)^{\frac{1}{2}m-1}, \tag{11}$$

for $0 \le v \le 1$ and 0 for v outside this interval. Since this distribution does not depend on Y_{i-1}, we see that the ratio V_i is independent of Y_{i-1}, and hence independent of $| V_1, \dots, V_{i-1}$. Then $V_1, \dots, V_p$ are independent.

Theorem 8.4.1. *The distribution of U defined by* (3) *is the distribution of the product $\prod_{i=1}^p V_i$, where $V_1, \dots, V_p$ are independent and V_i has the density* (11).

The cdf of U can be found by integrating the joint density of $V_1, \dots, V_p$ over the range

$$\prod_{i=1}^p V_i \le u. \tag{12}$$

We shall now show that for given $N-q_2$ the indices p and q_1 can be interchanged; that is, the distributions of $U_{p,q_1,N-q_2-q_1} = U_{p,m,n}$ and of $U_{q_1,p,N-q_2-p} = U_{m,p,n+m-p}$ are the same. The joint density of G and W_1 defined in Section 8.2 when $\Sigma = I$ and $\boldsymbol{\beta}_1 = \mathbf{0}$ is

$$\text{(13)} \qquad \frac{|G|^{\frac{1}{2}(n-p-1)} e^{-\frac{1}{2}\operatorname{tr} G - \frac{1}{2}\operatorname{tr} W_1 W_1'}}{2^{\frac{1}{2}np} \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma\left[\frac{1}{2}(n+1-i)\right] (2\pi)^{\frac{1}{2}mp}}.$$

Let $G + W_1 W_1' = J = CC'$ and let $W_1 = CU$. Then

$$\text{(14)} \qquad U_{p,m,n} = \frac{|G|}{|G + W_1 W_1'|} = \frac{|CC' - CUU'C'|}{|CC'|} = |I_p - UU'|$$

$$= \begin{vmatrix} I_p & U \\ U' & I_m \end{vmatrix} = \begin{vmatrix} I_m & U' \\ U & I_p \end{vmatrix} = |I_m - U'U|;$$

the fourth and sixth equalities follow from Theorem A.3.2 of the Appendix, and thε fifth from permutation of rows and columns. Since the Jacobian of $W_1 = CU$ is mode$|C|^m = |J|^{\frac{1}{2}m}$, the joint density of J and U is

$$\text{(15)} \qquad \frac{|J|^{\frac{1}{2}(n+m-p-1)} e^{-\frac{1}{2}\operatorname{tr} J}}{2^{\frac{1}{2}(n+m)p} \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma\left[\frac{1}{2}(n+m+1-i)\right]}$$

$$\cdot \prod_{i=1}^{p} \left\{ \frac{\Gamma\left[\frac{1}{2}(n+m+1-i)\right]}{\Gamma\left[\frac{1}{2}(n+1-i)\right]} \right\} \frac{|I_p - UU'|^{\frac{1}{2}(n-p-1)}}{\pi^{\frac{1}{2}mp}}$$

for J and $I_p - UU'$ positive definite, and 0 otherwise. Thus J and U are independently distributed; the density of J is the first term in (15), namely, $w(J|I_p, n+m)$, and the density of U is the second term, namely, of the form

$$\text{(16)} \qquad K|I_p - UU'|^{\frac{1}{2}(n-p-1)}$$

for $I_p - UU'$ positive definite, and 0 otherwise. Let $U_* = U'$, $p^* = m$, $m^* = p$, and $n^* = n + m - p$. Then the density of U_* is

$$\text{(17)} \qquad K|I_p - U_*'U_*|^{\frac{1}{2}(n-p-1)}$$

for $I_p - U_*'U_*$ positive definite, and 0 otherwise. By (14), $|I_p - U_*'U_*| = |I_m - U_*U_*'|$, and hence the density of U_* is

$$\text{(18)} \qquad K|I_{p^*} - U_*U_*'|^{\frac{1}{2}(n^*-p^*-1)},$$

which is of the form of (16) with p replaced by $p^* = m$, m replaced by $m^* = p$, and $n - p - 1$ replaced by $n^* - p^* - 1 = n - p - 1$. Finally we note that $U_{p,m,n}$ given by (14) is $|I_m - U_* U_*'| = U_{m,p,n+m-p}$.

Theorem 8.4.2. *When the hypothesis is true, the distribution of $U_{p,q_1,N-q_1-q_2}$ is the same as that of $U_{q_1,p,N-p-q_2}$ (i.e., that of $U_{p,m,n}$ is that of $U_{m,p,n+m-p}$).*

8.4.2. Moments

Since (11) is a density and hence integrates to 1, by change of notation

$$\int_0^1 v^{a-1}(1-v)^{b-1}\,dv = B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}. \tag{19}$$

From this fact we see that the hth moment of V_i is

$$\begin{aligned} \mathscr{E}V_i^h &= \int_0^1 \frac{\Gamma\left[\frac{1}{2}(n+m+1-i)\right]}{\Gamma\left[\frac{1}{2}(n+1-i)\right]\Gamma\left(\frac{1}{2}m\right)} v^{\frac{1}{2}(n+1-i)+h-1}(1-v)^{\frac{1}{2}m-1}\,dv \\ &= \frac{\Gamma\left[\frac{1}{2}(n+1-i)+h\right]\Gamma\left[\frac{1}{2}(n+m+1-i)\right]}{\Gamma\left[\frac{1}{2}(n+1-i)\right]\Gamma\left[\frac{1}{2}(n+m+1-i)+h\right]}. \end{aligned} \tag{20}$$

Since $V_1, \dots, V_p$ are independent, $\mathscr{E}U^h = \mathscr{E}\prod_{i=1}^p V_i^h = \prod_{i=1}^p \mathscr{E}V_i^h$. We obtain the following theorem:

Theorem 8.4.3. *The hth moment of U [if $h > -\frac{1}{2}(n+1-p)$] is*

$$\begin{aligned} \mathscr{E}U^h &= \prod_{i=1}^p \frac{\Gamma\left[\frac{1}{2}(n+1-i)+h\right]\Gamma\left[\frac{1}{2}(n+m+1-i)\right]}{\Gamma\left[\frac{1}{2}(n+1-i)\right]\Gamma\left[\frac{1}{2}(n+m+1-i)+h\right]} \\ &= \prod_{i=1}^p \frac{\Gamma\left[\frac{1}{2}(N-q_1-q_2+1-i)+h\right]\Gamma\left[\frac{1}{2}(N-q_2+1-i)\right]}{\Gamma\left[\frac{1}{2}(N-q_1-q_2+1-i)\right]\Gamma\left[\frac{1}{2}(N-q_2+1-i)+h\right]}. \end{aligned} \tag{21}$$

In the first expression p can be replaced by m, m by p, and n by $n+m-p$.

Suppose p is even, that is, $p = 2r$. We use the duplication formula

$$\Gamma\left(\alpha+\tfrac{1}{2}\right)\Gamma(\alpha+1) = \frac{\sqrt{\pi}\,\Gamma(2\alpha+1)}{2^{2\alpha}}. \tag{22}$$

Then the hth moment of $U_{2r,m,n}$ is

(23)
$$\mathscr{E}U_{2r,m,n}^h = \prod_{j=1}^{r}\left\{\frac{\Gamma[\frac{1}{2}(m+n+2)-j]}{\Gamma[\frac{1}{2}(m+n+2)-j+h]}\,\frac{\Gamma[\frac{1}{2}(m+n+1)-j]}{\Gamma[\frac{1}{2}(m+n+1)-j+h]}\right.$$
$$\left.\cdot\frac{\Gamma[\frac{1}{2}(n+2)-j+h]\Gamma[\frac{1}{2}(n+1)-j+h]}{\Gamma[\frac{1}{2}(n+2)-j]\Gamma[\frac{1}{2}(n+1)-j]}\right\}$$
$$= \prod_{j=1}^{r}\left\{\frac{\Gamma(m+n+1-2j)\Gamma(n+1-2j+2h)}{\Gamma(m+n-1-2j+2h)\Gamma(n+1-2j)}\right\}.$$

It is clear from the definition of the beta function that (23) is

$$\prod_{j=1}^{r}\left\{\int_0^1 \frac{\Gamma(m+n+1-2j)}{\Gamma(n+1-2j)\Gamma(m)} y^{(n+1-2j)+2h-1}(1-y)^{m-1}\,dy\right\} \tag{24}$$
$$= \prod_{j=1}^{r} \mathscr{E}Y_j^{2h} = \mathscr{E}\left(\prod_{j=1}^{r} Y_j^2\right)^h,$$

where the Y_j are independent and Y_j has density $\beta(y;n+1-2j,m)$.

Suppose p is odd; that is, $p=2s+1$. Then

$$\mathscr{E}U_{2s+1,m,n}^h = \mathscr{E}\left(\prod_{i=1}^{s} Z_i^2 Z_{s+1}\right)^h, \tag{25}$$

where the Z_i are independent and Z_i has density $\beta(z;n+1-2i,m)$ for $i=1,\dots,s$ and Z_{s+1} is distributed with density $\beta[z;(n+1-p)/2,m/2]$.

Theorem 8.4.4. *$U_{2r,m,n}$ is distributed as $\prod_{i=1}^{r} Y_i^2$, where $Y_1,\dots,Y_r$ are independent and Y_i has density $\beta(y;n+1-2i,m)$; $U_{2s+1,m,n}$ is distributed as $\prod_{i=1}^{s} Z_i^2 Z_{s+1}$, where the Z_i, $i=1,\dots,s$, are independent and Z_i has density $\beta(z;n+1-2i,m)$, and Z_{s+1} is independently distributed with density $\beta[z;\frac{1}{2}(n+1-p),\frac{1}{2}m]$.*

8.4.3. Some Special Distributions

$p=1$

From the preceding characterization we see that the density of $U_{1,m,n}$ is

$$\frac{\Gamma[\frac{1}{2}(n+m)]}{\Gamma(\frac{1}{2}n)\Gamma(\frac{1}{2}m)} u^{\frac{1}{2}n-1}(1-u)^{\frac{1}{2}m-1}. \tag{26}$$

Another way of writing $U_{1,m,n}$ is

$$U_{1,m,n} = \frac{1}{1+\sum_{i=1}^{m} Y_i^2/g_{11}} = \frac{1}{1+(m/n)F_{m,n}}, \tag{27}$$

where g_{11} is the one element of $G = N\hat{\Sigma}_\Omega$ and $F_{m,n}$ is an F-statistic. Thus

$$\frac{1-U_{1,m,n}}{U_{1,m,n}} \cdot \frac{n}{m} = F_{m,n}. \tag{28}$$

Theorem 8.4.5. *The distribution of* $[(1-U_{1,m,n})/U_{1,m,n}] \cdot n/m$ *is the* F*-distribution with* m *and* n *degrees of freedom; the distribution of* $[(1-U_{p,1,n}/U_{p,1,n}] \cdot (n+1-p)/p$ *is the* F*-distribution with* p *and* $n+1-p$ *degrees of freedom.*

***p* = 2**

From Theorem 8.4.4, we see that the density of $\sqrt{U_{2,m,n}}$ is

$$\frac{\Gamma(n+m-1)}{\Gamma(n-1)\Gamma(m)} x^{n-2}(1-x)^{m-1}, \tag{29}$$

and thus the density of $U_{2,m,n}$ is

$$\frac{\Gamma(n+m-1)}{2\Gamma(n-1)\Gamma(m)} u^{\frac{1}{2}(n-3)}\left(1-\sqrt{u}\right)^{m-1}. \tag{30}$$

From (29) it follows that

$$\frac{1-\sqrt{U_{2,m,n}}}{\sqrt{U_{2,m,n}}} \cdot \frac{n-1}{m} = F_{2m,2(n-1)}. \tag{31}$$

Theorem 8.4.6. *The distribution of* $[(1-\sqrt{U_{2,m,n}})/\sqrt{U_{2,m,n}}] \cdot (n-1)/m$ *is the* F*-distribution with* $2m$ *and* $2(n-1)$ *degrees of freedom; the distribution of* $[(1-\sqrt{U_{p,2,n}})/\sqrt{U_{p,2,n}}] \cdot (n+1-p)/p$ *is the* F*-distribution with* $2p$ *and* $2(n+1-p)$ *degrees of freedom.*

***p* Even**

Wald and Brookner (1941) gave a method for finding the distribution of $U_{p,m,n}$ for p or m even. We shall present the method of Schatzoff (1966a). It will be convenient first to consider $U_{p,m,n}$ for $m = 2r$. We can write the event $\prod_{i=1}^{p} V_i \le u$ as

$$Y_1 + \cdots + Y_p \ge -\log u, \tag{32}$$

where $Y_1, \dots, Y_p$ are independent and $Y_i = -\log V_i$ has the density

$$(33)\quad K_i e^{-\frac{1}{2}(n+1-i)y}(1-e^{-y})^{r-1} = K_i \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} e^{-[\frac{1}{2}(n+1-i)+j]y}$$

for $0 \le y < \infty$ and 0 otherwise, and

$$(34)\qquad K_i = \frac{\Gamma[\frac{1}{2}(n+1-i)+r]}{\Gamma[\frac{1}{2}(n+1-i)]\Gamma(r)} = \frac{1}{(r-1)!}\prod_{j=0}^{r-1} \frac{n+1-i+2j}{2}.$$

The joint density of $Y_1, \dots, Y_p$ is then a linear combination of terms $\exp[-\sum_{i=1}^p a_i y_i]$. The density of $W_j = \sum_{i=1}^j Y_i$ can be obtained inductively from the density of $W_{j-1} = \sum_{i=1}^{j-1} Y_i$ and Y_j, $j = 2, \dots, p$, which is a linear combination of terms $w_{j-1}^k e^{cw_{j-1}+a_j y_j}$. The density of W_j consists of linear combinations of

$$
\begin{aligned}
(35)\qquad e^{a_j w_j}\int_0^{w_j} w^k e^{(c-a_j)w}\,dw &= e^{a_j w_j}\cdot\frac{w_j^{k+1}}{k+1} && \text{if } a_j = c,\\
&= e^{cw_j}\sum_{h=0}^{k}(-1)^h \frac{k!}{(k-h)!}\frac{w_j^{k-h}}{(c-a_j)^{h+1}}\\
&\quad + (-1)^{k+1} e^{a_j w_j}\frac{k!}{(c-a_j)^{k+1}} && \text{if } a_j \ne c.
\end{aligned}
$$

The evaluation involves integration by parts.

Theorem 8.4.7. *If p is even or if m is even, the density of $U_{p,m,n}$ can be expressed as a linear combination of terms $(-\log u)^k u^l$, where k is an integer and l is a half integer.*

From (35) we see that the cumulative distribution function of $-\log U$ is a linear combination of terms $w^k e^{-lw}$ and hence the cumulative distribution function of U is a linear combination of terms $(-\log u)^k u^l$. The values of k and l and the coefficients depend on p, m, and n. They can be obtained by inductively carrying out the procedure leading to Theorem 8.4.7. Pillai and Gupta (1969) used Theorem 8.4.3 for obtaining distributions.

An alternative approach is to use Theorem 8.4.4. The complement to the cumulative distribution function $U_{2r,m,n}$ is

(36)

$$\Pr\{U_{2r,m,n} \geq u\} = \Pr\left\{\prod_{i=1}^{r} Y_i > \sqrt{u}\right\}$$
$$= \int_{\sqrt{u}}^{1} \int_{\frac{\sqrt{u}}{y_1}}^{1} \cdots \int_{\frac{\sqrt{u}}{\prod_{i=1}^{r-1} y_i}}^{1} \prod_{i=1}^{r} \beta(y_i|n+1-2i, m)\, dy_r \cdots dy_2\, dy_1.$$

In the density, $(1-y_i)^{m-1}$ can be expanded by the binomial theorem. Then all integrations are expressed as integrations of powers of the variables.

As an example, consider $r = 2$. The density of Y_1 and Y_2 is

$$Cy_1^{n-2} y_2^{n-4} (1-y_1)^{m-1} (1-y_2)^{m-1} \tag{37}$$
$$= C \sum_{i,j=0}^{m-1} \frac{[(m-1)!]^2 (-1)^{i+j}}{(m-i-1)!(m-j-1)!i!j!} y_1^{n-2+i} y_2^{n-4+j},$$

where

$$C = \frac{\Gamma(n+m-1)\Gamma(n+m-3)}{\Gamma(n-1)\Gamma(n-3)\Gamma^2(m)}. \tag{38}$$

The complement to the cdf of $U_{4,m,n}$ is

$$\Pr\{U_{4,m,n} \geq u\} = C \sum_{i,j=0}^{m-1} \frac{[(m-1)!]^2(-1)^{i+j}}{(m-i-1)!(m-j-1)!i!j!} \tag{39}$$
$$\cdot \int_{\sqrt{u}}^{1} \int_{\frac{\sqrt{u}}{y_1}}^{1} y_1^{n-2+i} y_2^{n-4+j}\, dy_2\, dy_1$$
$$= C \sum_{i,j=0}^{m-1} \frac{[(m-1)!]^2(-1)^{i+j}}{(m-i-1)!(m-j-1)!i!j!(n-3+j)}$$
$$\cdot \int_{\sqrt{u}}^{1} \left[y_1^{n-2+i} - u^{\frac{1}{2}(n-3+j)} y_1^{1+i-j} \right] dy_1.$$

The last step of the integration yields powers of $\sqrt{u}$ and products of powers of $\sqrt{u}$ and $\log u$ (for $1+i-j = -1$).

Particular Values

Wilks (1935) gives explicitly the distributions of U for $p=1$, $p=2$, $p=3$ with $m=3$; $p=3$ with $m=4$; and $p=4$ with $m=4$. Wilks's formula for $p=3$ with $m=4$ appears to be incorrect; see the first edition of this book. Consul (1966) gives many distributions for special cases. See also Mathai (1971).

8.4.4. The Likelihood Ratio Procedure

Let $u_{p,m,n}(\alpha)$ be the α significance point for $U_{p,m,n}$; that is,

$$\Pr\{U_{p,m,n} \le u_{p,m,n}(\alpha) \mid H \text{ true}\} = \alpha. \tag{40}$$

It is shown in Section 8.5 that $-[n-\frac{1}{2}(p-m+1)]\log U_{p,m,n}$ has a limiting χ^2-distribution with pm degrees of freedom. Let $\chi^2_{pm}(\alpha)$ denote the α significance point of χ^2_{pm}, and let

$$C_{p,m,n-p+1}(\alpha) = \frac{-[n-\frac{1}{2}(p-m+1)]\log u_{p,m,n}(\alpha)}{\chi^2_{pm}(\alpha)}. \tag{41}$$

Table B.1 [from Pearson and Hartley (1972)] gives value of $C_{p,m,M}(\alpha)$ for $\alpha=0.10$ and 0.05, $p=1(1)10$, various even values of m, and $M=n-p+1$ $=1(1)10(2)20, 24, 30, 40, 60, 120$.

To test a null hypothesis one computes $U_{p,m,n}$ and rejects the null hypothesis at significance level α if

$$-[n-\tfrac{1}{2}(p-m+1)]\log U_{p,m,n} > C_{p,m,n-p+1}(\alpha)\chi^2_{pm}(\alpha). \tag{42}$$

Since $C_{p,m,n}(\alpha)>1$, the hypothesis is accepted if the left-hand side of (42) is less than $\chi^2_{pm}(\alpha)$.

The purpose of tabulating $C_{p,m,M}(\alpha)$ is that linear interpolation is reasonably accurate because the entries decrease monotonically and smoothly to 1 as M increases. Schatzoff (1966a) has recommended interpolation for odd p by using adjacent even values of p and displays some examples. The table also indicates how accurate the χ^2-approximation is. The table has been extended by Pillai and Gupta (1969).

8.4.5. A Step-down Procedure

The criterion U has been expressed in (7) as the product of independent beta variables $V_1, V_2, \dots, V_p$. The ratio V_i is a least squares criterion for testing the null hypothesis that in the regression of $x_i^* - \boldsymbol{\beta}_{i1}^*\mathbf{Z}_1$ on $\mathbf{Z}=(\mathbf{Z}_1'\ \mathbf{Z}_2')'$ and

X_{i-1} the coefficient of Z_1 is $\mathbf{0}$. The null hypothesis that the regression of X on Z_1 is $\boldsymbol{\beta}_1^*$, which is equivalent to the hypothesis that the regression of $X - \boldsymbol{\beta}_1^* Z_1$ on Z_1 is $\mathbf{0}$, is composed of the hypotheses that the regression of $x_i^* - \beta_{i1}^* Z_1$ on Z_1 is $\mathbf{0}$, $i = 1, \dots, p$. Hence the null hypothesis $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^*$ can be tested by use of $V_1, \dots, V_p$.

Since V_i has the beta density (11) under the hypothesis $\beta_{i1} = \beta_{i1}^*$,

$$\frac{1 - V_i}{V_i} \frac{n - i + 1}{m} \tag{43}$$

has the F-distribution with m and $n - i + 1$ degrees of freedom. The step-down testing procedure is to compare (43) for $i = 1$ with the significance point $F_{m,n}(\varepsilon_1)$; if (43) for $i = 1$ is larger, reject the null hypothesis that the regression of $x_1^* - \beta_{11}^* Z_1$ on Z_1 is $\mathbf{0}$ and hence reject the null hypothesis that $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^*$. If this first component null hypothesis is accepted, compare (43) for $i = 2$ with $F_{m,n-1}(\varepsilon_2)$. In sequence, the component null hypotheses are tested. If one is rejected, the sequence is stopped and the hypothesis $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^*$ is rejected. If all component null hypotheses are accepted, the composite hypothesis is accepted. When the hypothesis $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^*$ is true, the probability of accepting it is $\prod_{i=1}^p (1 - \varepsilon_i)$. Hence the significance level of the step-down test is $1 - \prod_{i=1}^p (1 - \varepsilon_i)$.

In the step-down procedure the investigator usually has a choice of the ordering of the variables† (i.e., the numbering of the components of X) and a selection of component significance levels. It seems reasonable to order the variables in descending order of importance. The choice of significance levels will affect the power. If ε_i is a very small number, it will take a correspondingly large deviation from the ith null hypothesis to lead to rejection. In the absence of any other reason, the component significance levels can be taken equal. This procedure, of course, is not invariant with respect to linear transformation of the dependent vector variable. However, before carrying out a step-down procedure, a linear transformation can be used to determine the p variables.

The factors can be grouped. For example, group $x_1, \dots, x_k$ into one set and $x_{k+1}, \dots, x_p$ into another set. Then $U_{k,m,n} = \prod_{i=1}^k V_i$ can be used to test the null hypothesis that the first k rows of $\boldsymbol{\beta}_1$ are the first k rows of $\boldsymbol{\beta}_1^*$. Subsequently $\prod_{i=k+1}^p V_i$ is used to test the hypothesis that the last $p - k$ rows of $\boldsymbol{\beta}_1$ are those of $\boldsymbol{\beta}_1^*$; this latter criterion has the distribution under the null hypothesis of $U_{p-k,m,n-k}$.

†In some cases the ordering of variables may be imposed; for example, x_1 might be an observation at the first time point, x_2 at the second time point, and so on.

The investigator may test the null hypothesis $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^*$ by the likelihood ratio procedure. If the hypothesis is rejected, he may look at the factors $V_1, \dots, V_p$ to try to determine which rows of $\boldsymbol{\beta}_1$ might be different from $\boldsymbol{\beta}_1^*$.

The factors can also be used to obtain confidence regions for $\boldsymbol{\beta}_{11}, \dots, \boldsymbol{\beta}_{p1}$. Let $v_i(\varepsilon_i)$ be defined by

$$\frac{1 - v_i(\varepsilon_i)}{v_i(\varepsilon_i)} \cdot \frac{n-i+1}{m} = F_{m, n-i+1}(\varepsilon_i). \tag{44}$$

Then a confidence region for $\boldsymbol{\beta}_{i1}$ of confidence $1 - \varepsilon_i$ is

$$\frac{\begin{vmatrix} x_i^* x_i^{*\prime} & x_i^* X_{i-1}' & x_i^* Z' \\ X_{i-1} x_i^{*\prime} & X_{i-1} X_{i-1}' & X_{i-1} Z' \\ Z x_i^{*\prime} & Z X_{i-1}' & ZZ' \end{vmatrix}}{\begin{vmatrix} (x_i^* - \overline{\boldsymbol{\beta}}_{i1} Z_1)(x_i^* - \overline{\boldsymbol{\beta}}_{i1} Z_1)' & (x_i^* - \overline{\boldsymbol{\beta}}_{i1} Z_1) X_{i-1}' & (x_i^* - \overline{\boldsymbol{\beta}}_{i1} Z_1) Z_2' \\ X_{i-1}(x_i^* - \overline{\boldsymbol{\beta}}_{i1} Z_1)' & X_{i-1} X_{i-1}' & X_{i-1} Z_2' \\ Z_2 (x_i^* - \overline{\boldsymbol{\beta}}_{i1} Z_1)' & Z_2 X_{i-1}' & Z_2 Z_2' \end{vmatrix}} \tag{45}$$

$$\cdot \frac{\begin{vmatrix} X_{i-1} X_{i-1}' & Z_{i-1} Z_2^* \\ Z_2 X_{i-1}' & Z_2 X_2' \end{vmatrix}}{\begin{vmatrix} X_{i-1} X_{i-1}' & X_{i-1} Z' \\ Z X_{i-1}' & ZZ' \end{vmatrix}} \ge v_i(\varepsilon_i).$$

8.5. AN ASYMPTOTIC EXPANSION OF THE DISTRIBUTION OF THE LIKELIHOOD RATIO CRITERION

8.5.1. General Theory of Asymptotic Expansions

In this section we develop a large-sample distribution theory for the criterion studied in this chapter. First we develop a general asymptotic expansion of the distribution of a random variable whose moments are certain functions of gamma functions [Box (1949)]. Then we apply it to the case of the likelihood ratio criterion for the linear hypothesis.

We consider a random variable W $(0 \le W \le 1)$ with hth moment†

$$\mathscr{E} W^h = K \left(\frac{\prod_{j=1}^b y_j^{y_j}}{\prod_{k=1}^a x_k^{x_k}} \right)^h \frac{\prod_{k=1}^a \Gamma[x_k(1+h) + \xi_k]}{\prod_{j=1}^b \Gamma[y_j(1+h) + \eta_j]}, \qquad h = 0, 1, \dots, \tag{1}$$

†In all cases where we apply this result, the parameters x_k, ξ_k, y_j, and η_j will be such that there is a distribution with such moments.

where K is a constant such that $\mathscr{E}W^0 = 1$ and

$$\sum_{k=1}^{a} x_k = \sum_{j=1}^{b} y_j. \tag{2}$$

It will be observed that the hth moment of $\lambda = U_{p,q_1,n}^{\frac{1}{2}N}$ is of this form where $x_k = \frac{1}{2}N = y_j$, $\xi_k = \frac{1}{2}(-q+1-k)$, $\eta_j = \frac{1}{2}(-q_2+1-j)$, $a = b = p$. We treat a more general case here because applications later in this book require it.

If we let

$$M = -2\log W, \tag{3}$$

the characteristic function of ρM $(0 \le \rho < 1)$ is

$$\begin{aligned} \phi(t) &= \mathscr{E}e^{it\rho M} \\ &= \mathscr{E}W^{-2it\rho} \\ &= K\left(\frac{\prod_{j=1}^{b} y_j^{y_j}}{\prod_{k=1}^{a} x_k^{x_k}}\right)^{-2it\rho} \frac{\prod_{k=1}^{a}\Gamma[x_k(1-2it\rho)+\xi_k]}{\prod_{j=1}^{b}\Gamma[y_j(1-2it\rho)+\eta_j]}. \end{aligned} \tag{4}$$

Here ρ is arbitrary; later it will depend on N. If $a = b$, $x_k = y_k$, $\xi_k \le \eta_k$, then (1) is the hth moment of the product of powers of variables with beta distributions, and then (1) holds for all h for which the gamma functions exist. In this case (4) is valid for all real t. We shall assume here that (4) holds for all real t, and in each case where we apply the result we shall verify this assumption.

Let

$$\Phi(t) = \log\phi(t) = g(t) - g(0), \tag{5}$$

where

$$\begin{aligned} g(t) = 2it\rho\left(\sum_{k=1}^{a} x_k \log x_k - \sum_{j=1}^{b} y_j \log y_j\right) \\ + \sum_{k=1}^{a} \log\Gamma[\rho x_k(1-2it) + \beta_k + \xi_k] \\ - \sum_{j=1}^{b} \log\Gamma[\rho y_j(1-2it) + \varepsilon_j + \eta_j], \end{aligned}$$

where $\beta_k = (1-\rho)x_k$ and $\varepsilon_j = (1-\rho)y_j$. The form $g(t) - g(0)$ makes $\Phi(0) = 0$, which agrees with the fact that K is such that $\phi(0) = 1$. We make use of an expansion formula for the gamma function [Barnes (1899), p. 64] which is asymptotic in x for bounded h:

$$\log\Gamma(x+h)=\log\sqrt{2\pi}+\left(x+h-\tfrac{1}{2}\right)\log x-x - \sum_{r=1}^{m}(-1)^r\frac{B_{r+1}(h)}{r(r+1)x^r}+R_{m+1}(x), \tag{6}$$

where† $R_{m+1}(x)=O(x^{-(m+1)})$ and $B_r(h)$ is the Bernoulli polynomial of degree r and order unity defined by‡

$$\frac{\tau e^{h\tau}}{e^\tau-1}=\sum_{r=0}^{\infty}\frac{\tau^r}{r!}B_r(h). \tag{7}$$

The first three polynomials are $[B_0(h)=1]$

$$\begin{aligned} B_1(h)&=h-\tfrac{1}{2},\\ B_2(h)&=h^2-h+\tfrac{1}{6},\\ B_3(h)&=h^3-\tfrac{3}{2}h^2+\tfrac{1}{2}h. \end{aligned} \tag{8}$$

Taking $x=\rho x_k(1-2it)$, $\rho y_j(1-2it)$ and $h=\beta_k+\xi_k$, $\varepsilon_j+\eta_j$ in turn, we obtain

$$\Phi(t)=Q-g(0)-\tfrac{1}{2}f\log(1-2it) + \sum_{r=1}^{m}\omega_r(1-2it)^{-r}+\sum_{k=1}^{a}O\left(x_k^{-(m+1)}\right)+\sum_{j=1}^{b}O\left(y_j^{-(m+1)}\right), \tag{9}$$

where

$$f=-2\left\{\sum_k\xi_k-\sum_j\eta_j-\tfrac{1}{2}(a-b)\right\}, \tag{10}$$

$$\omega_r=\frac{(-1)^{r+1}}{r(r+1)}\left\{\sum_k\frac{B_{r+1}(\beta_k+\xi_k)}{(\rho x_k)^r}-\sum_j\frac{B_{r+1}(\varepsilon_j+\eta_j)}{(\rho y_j)^r}\right\}, \tag{11}$$

$$Q=\tfrac{1}{2}(a-b)\log 2\pi-\tfrac{1}{2}f\log\rho + \sum_k\left(x_k+\xi_k-\tfrac{1}{2}\right)\log x_k-\sum_j\left(y_j+\eta_j-\tfrac{1}{2}\right)\log y_j. \tag{12}$$

†$R_{m+1}(x)=O(x^{-(m+1)})$ means $|x^{m+1}R_{m+1}(x)|$ is bounded as $|x|\to\infty$.

‡This definition differs slightly from that of Whittaker and Watson [(1943), p. 126], who expand $\tau(e^{h\tau}-1)/(e^\tau-1)$. If $B_r^*(h)$ is this second type of polynomial, $B_1(h)=B_1^*(h)-\frac{1}{2}$, $B_{2r}(h)=B_{2r}^*(h)+(-1)^{r+1}B_r$, where B_r is the rth Bernoulli number, and $B_{2r+1}(h)=B_{2r+1}^*(h)$.

One resulting form for $\phi(t)$ (which we shall not use here) is

$$(13)\qquad \phi(t) = e^{\Phi(t)} = e^{Q-g(0)}(1-2it)^{-\frac{1}{2}f}\sum_{v=0}^{m} a_v(1-2it)^{-v} + R^*_{m+1},$$

where $\sum_{v=0}^{m} a_v z^{-v}$ is the sum of the first $m+1$ terms in the series expansion of $\exp(-\sum_{r=0}^{m}\omega_r z^{-r})$, and R^*_{m+1} is a remainder term. Alternatively,

$$(14)\qquad \Phi(t) = -\tfrac{1}{2}f\log(1-2it) + \sum_{r=1}^{m}\omega_r\left[(1-2it)^{-r}-1\right] + R'_{m+1},$$

where

$$(15)\qquad R'_{m+1} = \sum_k O\left(x_k^{-(m+1)}\right) + \sum_j O\left(y_j^{-(m+1)}\right).$$

In (14) we have expanded $g(0)$ in the same way we expanded $g(t)$ and have collected similar terms.

Then

$$\begin{aligned}
(16)\qquad \phi(t) &= e^{\Phi(t)} \\
&= (1-2it)^{-\frac{1}{2}f}\exp\left[\sum_{r=1}^{m}\omega_r(1-2it)^{-r} - \sum_{r=1}^{m}\omega_r + R'_{m+1}\right] \\
&= (1-2it)^{-\frac{1}{2}f}\left\{\prod_{r=1}^{m}\left[1+\omega_r(1-2it)^{-r} + \frac{1}{2!}\omega_r^2(1-2it)^{-2r}\cdots\right]\right. \\
&\qquad \left.\times\prod_{r=1}^{m}\left(1-\omega_r+\frac{1}{2!}\omega_r^2-\cdots\right) + R''_{m+1}\right\} \\
&= (1-2it)^{-\frac{1}{2}f}\left[1+T_1(t)+T_2(t)+\cdots+T_m(t)+R'''_{m+1}\right],
\end{aligned}$$

where $T_r(t)$ is the term in the expansion with terms $\omega_1^{s_1}\cdots\omega_r^{s_r}$, $\sum is_i = r$; for example,

$$(17)\qquad T_1(t) = \omega_1\left[(1-2it)^{-1}-1\right],$$

$$(18)\qquad T_2(t) = \omega_2\left[(1-2it)^{-2}-1\right] + \tfrac{1}{2}\omega_1^2\left[(1-2it)^{-2}-2(1-2it)^{-1}+1\right].$$

In most applications, we will have $x_k = c_k\theta$ and $y_j = d_j\theta$, where c_k and d_j will be constant and θ will vary (i.e., will grow with the sample size). In this case if ρ is chosen so $(1-\rho)x_k$ and $(1-\rho)y_j$ have limits, then R'''_{m+1} is $O(\theta^{-(m+1)})$. We collect in (16) all terms $\omega_1^{s_1}\cdots\omega_r^{s_r}$, $\sum is_i = r$, because these terms are $O(\theta^{-r})$.

It will be observed that $T_r(t)$ is a polynomial of degree r in $(1-2it)^{-1}$ and each term of $(1-2it)^{-\frac{1}{2}f}T_r(t)$ is a constant times $(1-2it)^{-\frac{1}{2}v}$ for an integral v. We know that $(1-2it)^{-\frac{1}{2}v}$ is the characteristic function of the χ^2-density with v degrees of freedom; that is,

$$g_v(z) = \frac{1}{2^{\frac{1}{2}v}\Gamma\left(\frac{1}{2}v\right)} z^{\frac{1}{2}v-1} e^{-\frac{1}{2}z} \tag{19}$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi}(1-2it)^{-\frac{1}{2}v} e^{-itz}\,dt.$$

Let

$$\begin{aligned} S_r(z) &= \int_{-\infty}^{\infty} \frac{1}{2\pi}(1-2it)^{-\frac{1}{2}f} T_r(t) e^{-itz}\,dt, \\ R_{m+1}^{\text{iv}} &= \int_{-\infty}^{\infty} \frac{1}{2\pi}(1-2it)^{-\frac{1}{2}f} R_{m+1}''' e^{-itz}\,dt. \end{aligned} \tag{20}$$

Then the density of ρM is

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{1}{2\pi}\phi(t) e^{-itz}\,dt &= \sum_{r=0}^{m} S_r(z) + R_{m+1}^{\text{iv}} \\ &= g_f(z) + \omega_1\left[g_{f+2}(z) - g_f(z)\right] \\ &\quad + \Big\{\omega_2\left[g_{f+4}(z) - g_f(z)\right] \\ &\qquad + \frac{\omega_1^2}{2}\left[g_{f+4}(z) - 2g_{f+2}(z) + g_f(z)\right]\Big\} \\ &\quad + \cdots + S_m(z) + R_{m+1}^{\text{iv}}. \end{aligned} \tag{21}$$

Let

$$\begin{aligned} U_r(z_0) &= \int_0^{z_0} S_r(z)\,dz, \\ R_{m+1}^{\text{v}} &= \int_0^{z_0} R_{m+1}^{\text{iv}}\,dz. \end{aligned} \tag{22}$$

The cdf of M is written in terms of the cdf of ρM, which is the integral of the density, namely,

$$
\begin{aligned}
(23)\quad \Pr\{M \le M_0\} &= \Pr\{\rho M \le \rho M_0\} \\
&= \sum_{r=0}^{m} U_r(\rho M_0) + R_{m+1}^{\mathrm{v}} \\
&= \Pr\{\chi_f^2 \le \rho M_0\} + \omega_0\left(\Pr\{\chi_{f+2}^2 \le \rho M_0\} - \Pr\{\chi_f^2 \le \rho M_0\}\right) \\
&\quad + \left[\omega_2\left(\Pr\{\chi_{f+4}^2 \le \rho M_0\} - \Pr\{\chi_f^2 \le \rho M_0\}\right) + \frac{\omega_1^2}{2}\left(\Pr\{\chi_{f+4}^2 \le \rho M_0\}\right.\right. \\
&\qquad \left.\left. -2\Pr\{\chi_{f+2}^2 \le \rho M_0\} + \Pr\{\chi_f^2 \le \rho M_0\}\right)\right] \\
&\quad + \cdots + U_m(\rho M_0) + R_{m+1}^{\mathrm{v}}.
\end{aligned}
$$

The remainder R_{m+1}^{v} is $O(\theta^{-(m+1)})$; this last statement can be verified by following the remainder terms along. (In fact, to make the proof rigorous one needs to verify that each remainder is of the proper order in a uniform sense.)

In many cases it is desirable to choose ρ so that $\omega_1 = 0$. In such a case using only the first term of (23) gives an error of order θ^{-2}.

Further details of the expansion can be found in Box's paper (1949).

Theorem 8.5.1. *Suppose that $\mathscr{E}W^h$ is given by* (1) *for all purely imaginary h, with* (2) *holding. Then the* cdf *of $-2\rho\log W$ is given by* (23). *The error, R_{m+1}^{v}, is $O(\theta^{-(m+1)})$ if $x_k \ge c_k\theta$, $y_j \ge d_j\theta$ $(c_k > 0,\ d_j > 0)$, and if $(1-\rho)x_k$, $(1-\rho)y_j$ have limits, where ρ may depend on θ.*

Box also considers approximating the distribution of $-2\rho\log W$ by an F-distribution. He finds that the error in this approximation can be made to be of order θ^{-3}.

8.5.2. Asymptotic Distribution of the Likelihood Ratio Criterion

We now apply Theorem 8.5.1 to the distribution of $-2\log\lambda$, the likelihood ratio criterion developed in Section 8.3. We let $W = \lambda$. The hth moment of λ is

$$
(24)\qquad \mathscr{E}\lambda^h = K\frac{\prod_{k=1}^{p}\Gamma\left[\tfrac{1}{2}(N-q+1-k+Nh)\right]}{\prod_{j=1}^{p}\Gamma\left[\tfrac{1}{2}(N-q_2+1-j+Nh)\right]},
$$

and this holds for all h for which the gamma functions exist, including purely imaginary h. We let $a = b = p$,

$$
(25) \qquad \begin{aligned} x_k &= \tfrac{1}{2}N, & \xi_k &= \tfrac{1}{2}(-q+1-k), & \beta_k &= \tfrac{1}{2}(1-\rho)N, \\ y_j &= \tfrac{1}{2}N, & \eta_j &= \tfrac{1}{2}(-q_2+1-j), & \varepsilon_j &= \tfrac{1}{2}(1-\rho)N. \end{aligned}
$$

We observe that

$$
(26) \quad \begin{aligned} 2\omega_1 &= \sum_{k=1}^{p} \left\{ \frac{\left\{\frac{1}{2}[(1-\rho)N-q+1-k]\right\}^2 - \frac{1}{2}[(1-\rho)N-q+1-k]}{\frac{1}{2}\rho N} \right. \\ &\qquad \left. - \frac{\left\{\frac{1}{2}[(1-\rho)N-q_2+1-k]\right\}^2 - \frac{1}{2}[(1-\rho)N-q_2+1-k]}{\frac{1}{2}\rho N} \right\} \\ &= \frac{2}{\rho N} \sum_{k=1}^{p} \left[\frac{-2[(1-\rho)N-q_2+1-k]q_1 + q_1^2}{4} + \frac{q_1}{2} \right] \\ &= \frac{pq_1}{2\rho N}\left[-2(1-\rho)N + 2q_2 - 2 + (p+1) + q_1 + 2\right]. \end{aligned}
$$

To make this zero, we require that

$$
(27) \qquad \rho = \frac{N - q_2 - \frac{1}{2}(p+q_1+1)}{N}.
$$

Then

$$
(28) \quad \begin{aligned} &\Pr\left\{-2\frac{k}{N}\log\lambda \le z\right\} \\ &\quad = \Pr\left\{-k\log U_{p,q_1,N-q} \le z\right\} \\ &\quad = \Pr\left\{\chi^2_{pq_1} \le z\right\} \\ &\qquad + \frac{\gamma_2}{k^2}\left(\Pr\left\{\chi^2_{pq_1+4} \le z\right\} - \Pr\left\{\chi^2_{pq_1} \le z\right\}\right) \\ &\qquad + \frac{1}{k^4}\left[\gamma_4\left(\Pr\left\{\chi^2_{pq_1+8} \le z\right\} - \Pr\left\{\chi^2_{pq_1} \le z\right\}\right)\right. \\ &\qquad\qquad \left. - \gamma_2^2\left(\Pr\left\{\chi^2_{pq_1+4} \le z\right\} - \Pr\left\{\chi^2_{pq_1} \le z\right\}\right)\right] + R_5^{\mathrm{v}}, \end{aligned}
$$

where

$$(29)\qquad k = \rho N = N - q_2 - \tfrac{1}{2}(p + q_1 + 1) = n - \tfrac{1}{2}(p - q_1 + 1),$$

$$(30)\qquad \gamma_2 = \frac{pq_1(p^2 + q_1^2 - 5)}{48},$$

$$(31)\qquad \gamma_4 = \frac{\gamma_2^2}{2} + \frac{pq_1}{1920}\left[3p^4 + 3q_1^4 + 10p^2q_1^2 - 50(p^2 + q_1^2) + 159\right].$$

Since $\lambda = U_{p,q_1,n}^{\frac{1}{2}N}$, where $n = N - q$, (28) gives $\Pr\{-k \log U_{p,q_1,n} \le z\}$.

Theorem 8.5.2. *The* cdf *of* $-k \log U_{p,q_1,n}$ *is given by* (28) *with* $k = n - \frac{1}{2}(p - q_1 + 1)$, *and* γ_2 *and* γ_4 *given by* (30) *and* (31), *respectively. The remainder term is* $O(N^{-6})$.

The coefficient $k = n - \frac{1}{2}(p - q_1 + 1)$ is known as the *Bartlett correction.* If the first term of (28) is used, the error is of the order N^{-2}; if the second, N^{-4}; and if the third,† N^{-6}. The second term is always negative and is numerically maximum for $z = \sqrt{(pq_1 + 2)(pq_1)}$ $(= pq_1 + 1$, approximately). For $p \ge 3, q_1 \ge 3$, we have $\gamma_2/k^2 \le [(p^2 + q_1^2)/k]^2/96$, and the contribution of the second term lies between $-0.005[(p^2 + q_1^2)/k]^2$ and 0. For $p \ge 3$, $q_1 \ge 3$, we have $\gamma_4 \le \gamma_2^2$, and the contribution of the third term is numerically less than $(\gamma_2/k^2)^2$. A rough rule that may be followed is that use of the first term is accurate to three decimal places if $p^2 + q_1^2 \le k/3$.

As an example of the calculation, consider the case of $p = 3$, $q_1 = 6$, $N - q_2 = 24$, and $z = 26.0$ (the 10% significance point χ_{18}^2). In this case $\gamma_2/k^2 = 0.048$ and the second term is -0.007: $\gamma_4/k^4 = 0.0015$ and the third term is -0.0001. Thus the probability of $-19 \log U_{3,6,18} \le 26.0$ is 0.893 to three decimal places.

Since

$$(32)\qquad -\left[n - \tfrac{1}{2}(p - m + 1)\right]\log u_{p,m,n}(\alpha) = C_{p,m,n-p+1}(\alpha)\chi_{pm}^2(\alpha),$$

the proportional error in approximating the left-hand side by $\chi_{pm}^2(\alpha)$ is $C_{p,m,n-p+1} - 1$. The proportional error increases slowly with p and m.

8.5.3. A Normal Approximation

Mudholkar and Trivedi (1980), (1981) developed a normal approximation to the distribution of $-\log U_{p,m,n}$ which is asymptotic as p and/or $m \to \infty$. It is related to the Wilson–Hilferty normal approximation for the χ^2-distribution.

†Box has shown that the term of order N^{-5} is 0 and gives the coefficients to be used in the term of order N^{-6}.

First, we give the background of the approximation. Suppose $\{Y_k\}$ is a sequence of nonnegative random variables such that $(Y_k - \mu_k)/\sigma_k \xrightarrow{\mathscr{L}} N(0,1)$ as $k \to \infty$, where $\mathscr{E} Y_k = \mu_k$ and $\mathscr{V}(Y_k) = \sigma_k^2$. Suppose also that $\mu_k \to \infty$ and σ_k^2/μ_k is bounded as $k \to \infty$. Let $Z_k = (Y_k/\mu_k)^h$. Then

$$(33) \qquad \frac{Z_k - 1}{h(\sigma_k/\mu_k)} = \frac{\mu_k(Z_k - 1)}{h\sigma_k} \xrightarrow{\mathscr{L}} N(0,1)$$

by Theorem 4.2.3. The approach to normality may be accelerated by choosing h to make the distribution of Z_k nearly symmetric as measured by its third cumulant. The normal distribution is to be used as an approximation and is justified by its accuracy in practice. However, it will be convenient to develop the ideas in terms of limits, although rigor is not necessary.

By a Taylor expansion we express the hth moment of Y_k/μ_k as

$$(34) \quad \begin{aligned} \mathscr{E} Z_k &= \mathscr{E}\left(\frac{Y_k}{\mu_k}\right)^h \\ &= 1 + \frac{h(h-1)}{2}\frac{\sigma_k^2}{\mu_k} \\ &\quad + \frac{h(h-1)(h-2)}{24}\frac{4\phi_k - 3(h-3)\left(\sigma_k^2/\mu_k\right)^2}{\mu_k^2} + O(\mu_k^{-3}), \end{aligned}$$

where $\phi_k = \mathscr{E}(Y_k - \mu_k)^3/\mu_k$, assumed bounded. The rth moment of Z_k is expressed by replacment of h by rh in (34). The central moments of Z_k are

(35)

$$\mathscr{E}(Z_k - 1)^2 = h^2\frac{\sigma_k^2/\mu_k}{\mu_k} + \frac{h^2(h-1)}{2}\frac{2\phi_k + (3h-5)\left(\sigma_k^2/\mu_k\right)^2}{\mu_k^2} + O(\mu_k^{-3}),$$

$$(36) \qquad \mathscr{E}(Z_k - 1)^3 = h^3\frac{\phi_k + 3(h-1)\left(\sigma_k^2/\mu_k\right)^2}{\mu_k^2} + O(\mu_k^{-3}).$$

To make the third moment approximately 0 we take h to be

$$(37) \qquad h_0 = 1 - \frac{\mathscr{E}(Y_k - \mu_k)^3\mu_k}{3\sigma_k^4}.$$

Then $Z_k = (Y_k/\mu_k)^{h_0}$ is treated as normally distributed with mean and variance given by (34) and (35), respectively, with $h = h_0$.

Now we consider $-\log U_{p,m,n} = -\sum_{i=1}^{p} \log V_i$, where $V_1, \dots, V_p$ are independent and V_i has the density $\beta(x;(n+1-i)/2, m/2)$, $i = 1, \dots, p$. As $n \to \infty$ and $m \to \infty$, $-\log V_i$ tends to normality. If V has the density $\beta(x; a/2, b/2)$, the moment generating function of $-\log V$ is

$$\mathscr{E} e^{-t \log V} = \frac{\Gamma[(a+b)/2]\Gamma(a/2-t)}{\Gamma(a/2)\Gamma[(a+b)/2-t]}. \tag{38}$$

Its logarithm is the cumulant generating function. Differentiation of the last yields as the rth cumulant of V

$$C_r = (-1)^r \left[\psi^{(r-1)}\left(\frac{a}{2}\right) - \psi^{(r-1)}\left(\frac{a+b}{2}\right)\right], \qquad r = 1, 2, \dots, \tag{39}$$

where $\psi(w) = d \log \Gamma(w)/dw$. [See Abramovitz and Stegun (1972), p. 258, for example.] From $\Gamma(w+1) = w\Gamma(w)$ we obtain the recursion relation $\psi(w+1) = \psi(w) + 1/w$. This yields for $s = 0$ and l an integer

$$\psi^{(s)}(w+l) - \psi^{(s)}(w) = (-1)^{s+1} s! \sum_{j=0}^{l-1} \frac{1}{(w+j)^{s+1}}. \tag{40}$$

The validity of (40) for $s = 1, 2, \dots$ is verified by differentiation. [The expression for $\psi'(Z)$ in the first line of page 223 of Mudholkar and Trivedi (1981) is incorrect.] Thus for $b = 2l$

$$C_r = (r-1)! \sum_{j=0}^{l-1} \frac{1}{(a/2+j)^r}. \tag{41}$$

From these results we obtain as the rth cumulant of $-\log U_{p,2l,n}$

$$\kappa_r(-\log U_{p,2l,n}) = 2^r (r-1)! \sum_{i=1}^{p} \sum_{j=0}^{l-1} \frac{1}{(n-i+1-2j)^r}. \tag{42}$$

As $l \to \infty$ the series diverges for $r = 1$ and converges for $r = 2, 3$, and hence $\kappa_r/\kappa_1 \to 0$, $r = 2, 3$. The same is true as $p \to \infty$ (if n/p approaches a positive constant).

Given n, p, and l, the first three cumulants are calculated from (42). Then h_0 is determined from (37), and $(-\log U_{p,2l,n})^{h_0}$ is treated as approximately normally distributed with mean and variance calculated from (34) and (35) for $h = h_0$.

Mudholkar and Trivedi (1980) calculated the error of approximation for significance levels of 0.01 and 0.05 for n from 4 to 66, $p = 3, 7$, and

$q = 2, 6, 10$. The maximum error is less than 0.0007; in most cases the error is considerably less. The error for the χ^2-approximation is much larger, especially for small values of n.

In case of m odd the rth cumulant can be approximated by

$$2^r(r-1)!\sum_{i=1}^{p}\left[\sum_{j=0}^{\frac{1}{2}(m-3)}\frac{1}{(n-i+1-2j)^r}+\frac{1}{2}\frac{1}{(n-i+m)^r}\right]. \tag{43}$$

Davis (1933, 1935) gave tables of $\psi(w)$ and its derivatives.

8.5.4. An F-Approximation

Rao (1951) has used the expansion of Section 8.5.2 to develop an expansion of the distribution of another function of $U_{p,m,n}$ in terms of beta distributions. The constants can be adjusted so that the term after the leading one is of order m^{-4}. A good approximation is to consider

$$\frac{1-U^{1/s}}{U^{1/s}}\cdot\frac{ks-r}{pm} \tag{44}$$

as F with pm and $ks-r$ degrees of freedom, where

$$s=\sqrt{\frac{p^2m^2-4}{p^2+m^2-5}}, \qquad r=\frac{pm}{2}-1, \tag{45}$$

and k is $n-\frac{1}{2}(p-m-1)$. For $p=1$ or 2 or $m=1$ or 2 the F-distribution is exactly as given in Section 8.4. If $ks-r$ is not an integer, interpolation between two integer values can be used. For smaller values of m this approximation is more accurate than the χ^2-approximation.

8.6. OTHER CRITERIA FOR TESTING THE LINEAR HYPOTHESIS

8.6.1. Functions of Roots

Thus far the only test of the linear hypothesis we have considered is the likelihood ratio test. In this section we consider other test procedures.

Let $\hat{\boldsymbol{\Sigma}}_\Omega$, $\hat{\boldsymbol{\beta}}_{1\Omega}$, and $\hat{\boldsymbol{\beta}}_{2\omega}$ be the estimates of the parameters in $N(\boldsymbol{\beta}z, \boldsymbol{\Sigma})$, based on a sample of N observations. These are a sufficient set of statistics, and we shall base test procedures on them. As was shown in Section 8.3, if the hypothesis is $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^*$, one can reformulate the hypothesis as $\boldsymbol{\beta}_1 = \mathbf{0}$ (by replacing x_α by $x_\alpha - \boldsymbol{\beta}_1^* z_\alpha^{(1)}$). Moreover,

$$(1)\qquad \begin{aligned} \boldsymbol{\beta} z_\alpha &= \boldsymbol{\beta}_1 z_\alpha^{(1)} + \boldsymbol{\beta}_2 z_\alpha^{(2)} \\ &= \boldsymbol{\beta}_1\left(z_\alpha^{(1)} - A_{12}A_{22}^{-1}z_\alpha^{(2)}\right) + \left(\boldsymbol{\beta}_2 + \boldsymbol{\beta}_1 A_{12}A_{22}^{-1}\right)z_\alpha^{(2)} \\ &= \boldsymbol{\beta}_1 z_\alpha^{*(1)} + \boldsymbol{\beta}_2^* z_\alpha^{(2)}, \end{aligned}$$

where $\sum_\alpha z_\alpha^{*(1)} z_\alpha^{(2)\prime} = \mathbf{0}$ and $\sum_\alpha z_\alpha^{*(1)} z_\alpha^{*(1)\prime} = A_{11\cdot 2}$. Then $\hat{\boldsymbol{\beta}}_1 = \hat{\boldsymbol{\beta}}_{1\Omega}$ and $\hat{\boldsymbol{\beta}}_2^* = \hat{\boldsymbol{\beta}}_{2\omega}$.

We shall use the principle of invariance to reduce the set of tests to be considered. First, if we make the transformation $X_\alpha^* = X_\alpha + \Gamma z_\alpha^{(2)}$, we leave the null hypothesis invariant, since $\mathscr{E} X_\alpha^* = \boldsymbol{\beta}_1 z_\alpha^{*(1)} + (\boldsymbol{\beta}_2^* + \Gamma)z_\alpha^{(2)}$ and $\boldsymbol{\beta}_2^* + \Gamma$ is unspecified. The only invariants of the sufficient statistics are $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\beta}}_1$ (since for each $\hat{\boldsymbol{\beta}}_2^*$, there is a Γ that transforms it to $\mathbf{0}$, that is, $-\hat{\boldsymbol{\beta}}_2^*$).

Second, the null hypothesis is invariant under the transformation $z_\alpha^{**(1)} = Cz_\alpha^{*(1)}$ (C nonsingular); the transformation carries $\boldsymbol{\beta}_1$ to $\boldsymbol{\beta}_1 C^{-1}$. Under this transformation $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\beta}}_1 A_{11\cdot 2}\hat{\boldsymbol{\beta}}_1'$ are invariant; we consider $A_{11\cdot 2}$ as information relevant to inference. However, these are the only invariants. For consider a function of $\hat{\boldsymbol{\beta}}_1$ and $A_{11\cdot 2}$, say $f(\hat{\boldsymbol{\beta}}_1, A_{11\cdot 2})$. Then there is a C^* that carries this into $f(\hat{\boldsymbol{\beta}}_1 C^{*-1}, I)$, and a further orthogonal transformation carries this into $f(T, I)$, where $t_{iv} = 0$, $i < v$, $t_{ii} \geq 0$. (If each row of T is considered a vector in q_1-space, the rotation of coordinate axes can be done so the first vector is along the first coordinate axis, the second vector is in the plane determined by the first two coordinate axes, and so forth). But T is a function of $TT' = \hat{\boldsymbol{\beta}}_1 A_{11\cdot 2}\hat{\boldsymbol{\beta}}_1'$; that is, the elements of T are uniquely determined by this equation and the preceding restrictions. Thus our tests will depend on $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\beta}}_1 A_{11\cdot 2}\hat{\boldsymbol{\beta}}_1'$. Let $N\hat{\boldsymbol{\Sigma}} = G$ and $\hat{\boldsymbol{\beta}}_1 A_{11\cdot 2}\hat{\boldsymbol{\beta}}_1' = H$.

Third, the null hypothesis is invariant when x_α is replaced by Kx_α, for $\boldsymbol{\Sigma}$ and $\boldsymbol{\beta}_2^*$ are unspecified. This transforms G to KGK' and H to KHK'. The only invariants of G and H under such transformations are the roots of

$$(2)\qquad |H - lG| = 0.$$

It is clear the roots are invariant, for

$$(3)\qquad \begin{aligned} 0 &= |KHK' - lKGK'| \\ &= |K(H - lG)K'| \\ &= |K| \cdot |H - lG| \cdot |K'|. \end{aligned}$$

On the other hand, these are the only invariants, for given G and H there is a K such that $KGK' = I$ and

$$(4) \qquad KHK' = L = \begin{pmatrix} l_1 & 0 & \cdots & 0 \\ 0 & l_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & l_p \end{pmatrix},$$

where $l_1 \geq \cdots l_p$ are the roots of (2). (See Theorem A.2.2 of the Appendix.)

Theorem 8.6.1. *Let x_α be an observation from $N(\boldsymbol{\beta}_1 z_\alpha^{*(1)} + \boldsymbol{\beta}_2^* z_\alpha^{(2)}, \Sigma)$, where $\sum_\alpha z_\alpha^{*(1)} z_\alpha^{(2)\prime} = \mathbf{0}$ and $\sum_\alpha z_\alpha^{*(1)} z_\alpha^{*(1)\prime} = A_{11\cdot 2}$. The only functions of the sufficient statistics and $A_{11\cdot 2}$ invariant under the transformations $x_\alpha^* = x_\alpha + \Gamma z_\alpha^{(2)}$, $z_\alpha^{**(1)} = C z_\alpha^{*(1)}$, and $x_\alpha^* = K x_\alpha$ are the roots of (2), where $G = N\hat{\Sigma}$ and $H = \hat{\boldsymbol{\beta}}_1 A_{11\cdot 2} \hat{\boldsymbol{\beta}}_1'$.*

The likelihood ratio criterion is a function of

$$(5) \qquad \begin{aligned} U &= \frac{|G|}{|G+H|} = \frac{|KGK'|}{|KGK' + KHK'|} = \frac{|I|}{|I+L|} \\ &= \prod_{i=1}^{p} (1 + l_i)^{-1}, \end{aligned}$$

which is clearly invariant under the transformations.

Intuitively it would appear that good tests should reject the null hypothesis when the roots in some sense are large, for if $\boldsymbol{\beta}_1$ is very different from $\mathbf{0}$, then $\hat{\boldsymbol{\beta}}_1$ will tend to be large and so will H. Some other criteria that have been suggested are (a) $\sum l_i$, (b) $\sum l_i/(1 + l_i)$, (c) $\max l_i$, and (d) $\min l_i$. In each case we reject the null hypothesis if the criterion exceeds some specified number.

8.6.2. The Lawley–Hotelling Trace Criterion

Let K be the matrix such that $KGK' = I$ $[G = K^{-1}(K')^{-1}$, or $G^{-1} = K'K]$ and so (4) holds. Then the sum of the roots can be written

$$(6) \qquad \begin{aligned} \sum_{i=1}^{p} l_i &= \operatorname{tr} L = \operatorname{tr} KHK' \\ &= \operatorname{tr} HK'K = \operatorname{tr} HG^{-1}. \end{aligned}$$

This criterion was suggested by Lawley (1938), Bartlett (1939), and Hotelling (1947), (1951). The test procedure is to reject the hypothesis if (6) is greater than a constant depending on p, m, and n.

The general distribution† of $\operatorname{tr} \boldsymbol{H}\boldsymbol{G}^{-1}$ cannot be characterized as easily as that of $U_{p,m,n}$. In the case of $p=2$, Hotelling (1951) obtained an explicit expression for the distribution of $\operatorname{tr} \boldsymbol{H}\boldsymbol{G}^{-1} = l_1 + l_2$. A slightly different form of this distribution is obtained from the density of the two roots l_1 and l_2 in Chapter 13. It is

$$(7)\quad \Pr\{\operatorname{tr} \boldsymbol{H}\boldsymbol{G}^{-1} \le w\} = I_{w/(2+w)}(m-1, n-1)$$
$$-\frac{\sqrt{\pi}\,\Gamma\left[\frac{1}{2}(m+n-1)\right]}{\Gamma\left(\frac{1}{2}m\right)\Gamma\left(\frac{1}{2}n\right)}(1+w)^{-\frac{1}{2}(n-1)} I_{w^2/(2+w)^2}\left[\tfrac{1}{2}(m-1), \tfrac{1}{2}(n-1)\right],$$

where $I_x(a,b)$ is the *incomplete beta function*, that is, the integral of $\beta(y;a,b)$ from 0 to x.

Constantine (1966) expressed the density of $\operatorname{tr} \boldsymbol{H}\boldsymbol{G}^{-1}$ as an infinite series in generalized Laguerre polynomials and as an infinite series in zonal polynomials; these series, however, converge only for $\operatorname{tr} \boldsymbol{H}\boldsymbol{G}^{-1} < 1$. Davis (1968) showed that the analytic continuation of these series satisfies a system of linear homogeneous differential equations of order p. Davis (1970a, 1970b) used a solution to compute tables as given in Appendix B.

Under the null hypothesis, $\boldsymbol{G}$ is distributed as $\sum_{\alpha=1}^{n} \boldsymbol{Z}_\alpha \boldsymbol{Z}'_\alpha$ $(n = N - q)$ and $\boldsymbol{H}$ is distributed as $\sum_{\upsilon=1}^{q_1} \boldsymbol{Y}_\upsilon \boldsymbol{Y}'_\upsilon$, where the $\boldsymbol{Z}_\alpha$ and $\boldsymbol{Y}_\upsilon$ are independent, each with distribution $N(\boldsymbol{0}, \boldsymbol{\Sigma})$. Since the roots are invariant under the previously specified linear transformation, we can choose $\boldsymbol{K}$ so that $\boldsymbol{K}\boldsymbol{\Sigma}\boldsymbol{K}' = \boldsymbol{I}$ and let $\boldsymbol{G}^* = \boldsymbol{K}\boldsymbol{G}\boldsymbol{K}'$ $[= \sum(\boldsymbol{K}\boldsymbol{Z}_\alpha)(\boldsymbol{K}\boldsymbol{Z}_\alpha)']$ and $\boldsymbol{H}^* = \boldsymbol{K}\boldsymbol{H}\boldsymbol{K}'$. This is equivalent to assuming at the outset that $\boldsymbol{\Sigma} = \boldsymbol{I}$.

Now

$$(8)\qquad \operatorname*{plim}_{N\to\infty} \frac{1}{N}\boldsymbol{G} = \operatorname*{plim}_{n\to\infty} \frac{n}{n+q}\,\frac{1}{n}\sum_{\alpha=1}^{n} \boldsymbol{Z}_\alpha \boldsymbol{Z}'_\alpha = \boldsymbol{I}.$$

This result follows applying the (weak) law of large numbers to each element of $(1/n)\boldsymbol{G}$,

$$(9)\qquad \operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{\alpha=1}^{n} Z_{i\alpha} Z_{j\alpha} = \mathscr{E} Z_{i\alpha} Z_{j\alpha} = \delta_{ij}.$$

Theorem 8.6.2. *Let $f(\boldsymbol{H})$ be a function whose discontinuities form a set of probability zero when $\boldsymbol{H}$ is distributed as $\sum_{\upsilon=1}^{q_1} \boldsymbol{Y}_\upsilon \boldsymbol{Y}'_\upsilon$ with the $\boldsymbol{Y}_\upsilon$ independent, each with distribution $N(\boldsymbol{0}, \boldsymbol{I})$. Then the limiting distribution of $f(N\boldsymbol{H}\boldsymbol{G}^{-1})$ is the distribution of $f(\boldsymbol{H})$.*

†Lawley (1938) purported to derive the exact distribution, but the result is in error.

Proof. This is a straightforward application of a general theorem [for example, Theorem 2 of Chernoff (1956)] to the effect that if the cdf of $\boldsymbol{X}_n$ converges to that of $\boldsymbol{X}$ (at every continuity point of the latter) and if $g(\boldsymbol{x})$ is a function whose discontinuities form a set of probability 0 according to the distribution of $\boldsymbol{X}$, then the cdf of $g(\boldsymbol{X}_n)$ converges to that of $g(\boldsymbol{X})$. In our case $\boldsymbol{X}_n$ consists of the components of $\boldsymbol{H}$ and $\boldsymbol{G}$, and $\boldsymbol{X}$ consists of the components of $\boldsymbol{H}$ and $\boldsymbol{I}$. ■

Corollary 8.6.1. *The limiting distribution of $N \operatorname{tr} \boldsymbol{HG}^{-1}$ or $n \operatorname{tr} \boldsymbol{HG}^{-1}$ is the χ^2-distribution with pq_1 degrees of freedom.*

This follows from Theorem 8.6.2, because

$$\operatorname{tr} \boldsymbol{H} = \sum_{i=1}^{p} h_{ii} = \sum_{i=1}^{p} \sum_{\upsilon=1}^{q_1} Y_{i\upsilon}^2. \tag{10}$$

Ito (1956), (1960) developed asymptotic formulas, and Fujikoshi (1973) extended them. Let $w_{p,m,n}(\alpha)$ be the α significance point of $\operatorname{tr} \boldsymbol{HG}^{-1}$; that is,

$$\Pr\left\{\operatorname{tr} \boldsymbol{HG}^{-1} \geq w_{p,m,n}(\alpha)\right\} = \alpha, \tag{11}$$

and let $\chi_k^2(\alpha)$ be the α-significance point of the χ^2-distribution with k degrees of freedom. Then

$$nw_{p,m,n}(\alpha) = \chi_{pm}^2(\alpha) + \frac{1}{2n}\left[\frac{p+m+1}{pm+2}\chi_{pm}^4(\alpha) + (p-m+1)\chi_{pm}^2(\alpha)\right] + O(n^{-2}). \tag{12}$$

Ito also gives the term of order n^{-2}. See also Muirhead (1970). Davis (1970a), (1970b) evaluated the accuracy of the approximation (12). Ito also found

$$\Pr\{n \operatorname{tr} \boldsymbol{HG}^{-1} \leq z\} = G_{pm}(z) - \frac{1}{2n}\left[\frac{p+m+1}{pm+2}z^2 + (p-m+1)g_{pm}(z)\right] + O(n^{-2}), \tag{13}$$

where $G_k(z) = \Pr\{\chi_k^2 \leq z\}$ and $g_k(z) = (d/dz)G_k(z)$. Pillai (1956) suggested another approximation to $nw_{p,m,n}(\alpha)$, and Pillai and Samson (1959) gave moments of $\operatorname{tr} \boldsymbol{HG}^{-1}$. Pillai and Young (1971) and Krishnaiah and Chang (1972) evaluated the Laplace transform of $\operatorname{tr} \boldsymbol{HG}^{-1}$ and showed how to invert the transform. Khatri and Pillai (1966) suggest an approximate distribution based on moments. Pillai and Young (1971) suggest approximate distributions based on the first three moments.

Tables of the significance points are given by Grubbs (1954) for $p = 2$ and by Davis (1970a) for $p = 3$ and 4, Davis (1970b) for $p = 5$, and Davis (1980) for $p = 6(1)10$; approximate significance points have been given by Pillai (1960). Davis's tables are reproduced in Table B.2.

8.6.3. The Bartlett–Nanda–Pillai Trace Criterion

Another criterion, proposed by Bartlett (1939), Nanda (1950), and Pillai (1955), is

$$(14)\qquad \begin{aligned} V &= \sum_{i=1}^{p} \frac{l_i}{1+l_i} = \operatorname{tr} L(I+L)^{-1} \\ &= \operatorname{tr} KHK'(KGK' + KHK')^{-1} \\ &= \operatorname{tr} HK'[K(G+H)K']^{-1}K \\ &= \operatorname{tr} H(G+H)^{-1}, \end{aligned}$$

where as before K is such that $KGK' = I$ and (4) holds. In terms of the roots $f_i = l_i/(1+l_i)$, $i = 1, \dots, p$, of

$$(15)\qquad |H - f(H+G)| = 0,$$

the criterion is $\sum_{i=1}^{p} f_i$. In principle, the cdf, density, and moments under the null hypothesis can be found from the density of the roots (Sec. 13.2.3),

$$(16)\qquad C\prod_{i=1}^{p} f_i^{\frac{1}{2}(m-p-1)} \prod_{i=1}^{p} (1-f_i)^{\frac{1}{2}(n-p-1)} \prod_{i<j} (f_i - f_j),$$

where

$$(17)\qquad C = \frac{\pi^{\frac{1}{2}p^2}\Gamma_p\left[\frac{1}{2}(m+n)\right]}{\Gamma_p\left(\frac{1}{2}n\right)\Gamma_p\left(\frac{1}{2}m\right)\Gamma_p\left(\frac{1}{2}p\right)}$$

for $1 > f_1 > \cdots > f_p > 0$, and 0 otherwise. If $m - p$ and $n - p$ are odd, the density is a polynomial in $f_1, \dots, f_p$. Then the density and cdf of the sum of the roots are polynomials.

Many authors have written about the moments, Laplace transforms, densities, and cdfs, using various approaches. Nanda (1950) derived the distribution for $p = 2, 3, 4$ and $m = p + 1$. Pillai (1954), (1956), (1960) and Pillai and

Mijares (1959) calculated the first four moments of V and proposed approximating the distribution by a beta distribution based on the first four moments. Pillai and Jayachandran (1970) show how to evaluate the moment generating function as a weighted sum of determinants whose elements are incomplete gamma functions; they derive exact densities for some special cases and use them for a table of significance points. Krishnaiah and Chang (1972) express the distributions as linear combinations of inverse Laplace transforms of the products of certain double integrals and further develop this technique for finding the distribution. Davis (1972b) showed that the distribution satisfies a differential equation and showed the nature of the solution. Khatri and Pillai (1968) obtained the (nonnull) distributions in series forms. The characteristic function (under the null hypothesis) was given by James (1964). Pillai and Jayachandran (1967) found the nonnull distribution for $p = 2$ and computed power functions. For an extensive bibliography see Krishnaiah (1978).

We now turn to the asymptotic theory. It follows from Theorem 8.6.2 that nV or NV has a limiting χ^2-distribution with pm degrees of freedom.

Let $v_{p,m,n}(\alpha)$ be defined by

$$\Pr\left\{\operatorname{tr} \boldsymbol{H}(\boldsymbol{H}+\boldsymbol{G})^{-1} \geq v_{p,m,n}(\alpha)\right\} = \alpha. \tag{18}$$

Then Davis (1970a), (1970b), Fujikoshi (1973), and Rothenberg (1977) have shown that

$$nv_{p,m,n}(\alpha) = \chi^2_{pm}(\alpha) + \frac{1}{2n}\left[-\frac{p+m+1}{pm+2}\chi^4_{pm}(\alpha) + (p-m+1)\chi^2_{pm}(\alpha)\right] + O(n^{-2}). \tag{19}$$

Since we can write (for the likelihood ratio test)

$$nu_{p,m,n}(\alpha) = \chi^2_{pm}(\alpha) + \frac{1}{2n}(p-m+1)\chi^2_{pm}(\alpha) + O(n^{-2}), \tag{20}$$

we have the comparison

$$nw_{p,m,n}(\alpha) = nu_{p,m,n}(\alpha) + \frac{1}{2n}\cdot\frac{p+m+1}{pm+2}\chi^4_{pm}(\alpha) + O(n^{-2}), \tag{21}$$

$$nv_{p,m,n}(\alpha) = nu_{p,m,n}(\alpha) + \frac{1}{2n}\cdot\frac{p+m+1}{pm+2}\chi^4_{pm}(\alpha) + O(n^{-2}). \tag{22}$$

An asymptotic expansion [Muirhead (1970), Fujikoshi (1973)] is

$$\text{(23)}\quad \Pr\{nV \le z\} = G_{pn}(z) + \frac{pm}{4n}\big[(m-p-1)G_{pm}(z) + 2(p+1)G_{pm+2}(z) - (p+m+1)G_{pm+4}(z)\big] + O(n^{-2}).$$

Higher-order terms are given by Muirhead and Fujikoshi.

Tables. Pillai (1960) tabulated 1% and 5% significance points of V for $p = 2(1)8$ based on fitting Pearson curves (i.e., beta distributions with adjusted ranges) to the first four moments. Mijares (1964) extended the tables to $p = 50$. Table B.3 of some significance points of $(n+m)V/m = \operatorname{tr}(1/m)\boldsymbol{H}\{[1/(n+m)](\boldsymbol{G}+\boldsymbol{H})\}^{-1}$ is from *Concise Statistical Tables*, and was computed on the same basis as Pillai's. Schuurman, Krishnaiah, and Chattopodhyay (1975) gave exact significance points of V for $p = 2(1)5$; a more extensive table is in their technical report (ARL 73-0008). A comparison of some values with those of *Concise Statistical Tables* (Appendix B) shows a maximum difference of 3 in the third decimal place.

8.6.4. The Roy Maximum Root Criterion

Any characteristic root of $\boldsymbol{HG}^{-1}$ can be used as a test criterion. Roy (1953) proposed l_1, the maximum characteristic root of $\boldsymbol{HG}^{-1}$, on the basis of his union–intersection principle. The test procedure is to reject the null hypothesis if l_1 is greater than a certain number, or equivalently, if $f_1 = l_1/(1+l_1) = R$ is greater than a number $r_{p,m,n}(\alpha)$ which satisfies

$$\text{(24)}\qquad \Pr\{R \ge r_{p,m,n}(\alpha)\} = \alpha.$$

The density of the roots $f_1, \dots, f_p$ for $p \le m$ under the null hypothesis is given in (16). The cdf of $R = f_1$, $\Pr\{f_1 \le f^*\}$, can be obtained from the joint density by integration over the range $0 \le f_p \le \cdots \le f_1 \le f^*$. If $m-p$ and $n-p$ are both odd, the density of $f_1, \dots, f_p$ is a polynomial; then the cdf of f_1 is a polynomial in f^* and the density of f_1 is a polynomial. The only difficulty in carrying out the integration is keeping track of the different terms.

Roy [(1945), (1957), Appendix 9] developed a method of integration that results in a cdf that is a linear combination of products of univariate beta densities and beta cdfs. The cdf of f_1 for $p = 2$ is

$$\text{(25)}\quad \Pr\{f_1 \le f\} = I_f(m-1, n-1) - \frac{\sqrt{\pi}\,\Gamma\big[\tfrac{1}{2}(m+n-1)\big]}{\Gamma(\tfrac{1}{2}m)\Gamma(\tfrac{1}{2}n)} f^{\frac{1}{2}(m-1)}(1-f)^{\frac{1}{2}(n-1)} I_f\big[\tfrac{1}{2}(m-1), \tfrac{1}{2}(n-1)\big].$$

This is derived in Section 13.5. Roy (1957), Chapter 8, gives the cdfs for $p = 3$ and 4 also.

By Theorem 8.6.2 the limiting distribution of the largest characteristic root of nHG^{-1}, NHG^{-1}, $nH(H+G)^{-1}$, or $NH(H+G)^{-1}$ is the distribution of the largest characteristic root of H having the distribution $W(I, m)$. The densities of the roots of H are given in Section 13.3. In principle, the marginal density of the largest root can be obtained from the joint density by integration, but in actual fact the integration is more difficult than that for the density of the roots of HG^{-1} or $H(H+G)^{-1}$.

The literature on this subject is too extensive to summarize here. Nanda (1948) obtained the distribution for $p = 2$, 3, 4, and 5. Pillai (1954), (1956), (1965), (1967) treated the distribution under the null hypothesis. Other results were obtained by Sugiyama and Fukutomi (1966) and Sugiyama (1967). Pillai (1967) derived an appropriate distribution as a linear combination of incomplete beta functions. Davis (1972a) showed that the density of a single ordered root satisfies a differential equation and (1972b) derived a recurrence relation for it. Hayakawa (1967), Khatri and Pillai (1968), Pillai and Sugiyama (1969), and Khatri (1972) treated the noncentral case. See Krishnaiah (1978) for more references.

Tables. Tables of the percentage points have been calculated by Nanda (1951) and Foster and Rees (1957) for $p = 2$, Foster (1957) for $p = 3$, Foster (1958) for $p = 4$, and Pillai (1960) for $p = 2(1)6$ on the basis of an approximation. [See also Pillai (1956), (1960), (1964), (1965), (1967).] Heck (1960) presented charts of the significance points for $p = 2(1)6$. Table B.4 of significance points of nl_1/m is from *Concise Statistical Tables*, based on the approximation by Pillai (1967).

8.6.5. Comparison of Powers

The four tests that have been given most consideration are those based on Wilks's U, the Lawley–Hotelling W, the Bartlett–Nanda–Pillai V, and Roy's R. To guide in the choice of one of these four, we would like to compare power functions. The first three have been compared by Rothenberg on the basis of the asymptotic expansions of their distributions in the nonnull case.

Let $\nu_1^N, \dots, \nu_p^N$ be the roots of

$$|(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^*)A_{11\cdot2}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^*)' - \nu\boldsymbol{\Sigma}| = 0. \tag{26}$$

The distribution of

$$\operatorname{tr}(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)A_{11\cdot2}(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)'\boldsymbol{\Sigma}^{-1} \tag{27}$$

is the noncentral χ^2-distribution with pm degrees of freedom and noncentrality parameter $\sum_{i=1}^{p} \nu_i^N$. As $N \to \infty$, the quantity $(1/n)G$ or $(1/N)G$ approaches $\boldsymbol{\Sigma}$ with probability one. If we let $N \to \infty$ and $A_{11\cdot 2}$ is unbounded, the noncentrality parameter grows indefinitely and the power approaches 1. It is more informative to consider a sequence of alternatives such that the powers of the different tests are different. Suppose $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^N$ is a sequence of matrices such that as $N \to \infty$, $(\boldsymbol{\beta}_1^N - \boldsymbol{\beta}_1^*)A_{11\cdot 2}(\boldsymbol{\beta}_1^N - \boldsymbol{\beta}_1^*)'$ approaches a limit and hence $\nu_1^N, \dots, \nu_p^N$ approach some limiting values $\nu_1, \dots, \nu_p$, respectively. Then the limiting distribution of $N \operatorname{tr} HG^{-1}$, $n \operatorname{tr} HG^{-1}$, $N \operatorname{tr} H(H+G)^{-1}$, and $n \operatorname{tr} H(H+G)^{-1}$ is the noncentral χ^2-distribution with pm degrees of freedom and noncentrality parameter $\sum_{i=1}^{p} \nu_i$. Similarly for $-N \log U$ and $-n \log U$.

Rothenberg (1977) has shown under the above conditions that

$$
\begin{aligned}
(28)\quad \Pr\{U \le u_{p,m,n}(\alpha)\} = 1 - G_{pm}\left[\chi^2_{pm}(\alpha)\,\middle|\, \sum_{i=1}^{p} \nu_i\right]
&- \frac{1}{2n}\left\{(p+m+1)\sum_{i=1}^{p} \nu_i g_{pm+4}\left[\chi^2_{pm}(\alpha)\right]\right. \\
&\left. + \sum_{i=1}^{p} \nu_i^2 g_{pm+6}\left[\chi^2_{pm}(\alpha)\right]\right\} + o\left(\frac{1}{n}\right),
\end{aligned}
$$

$$
\begin{aligned}
(29)\quad &\Pr\{\operatorname{tr} HG^{-1} \ge w_{p,m,n}(\alpha)\} \\
&= 1 - G_{pm}\left[\chi^2_{pm}(\alpha)\,\middle|\, \sum_{i=1}^{p} \nu_i\right] \\
&\quad - \frac{1}{2n}\left\{(p+m+1)\sum_{i=1}^{p} \nu_i g_{pm+4}\left[\chi^2_{pm}(\alpha)\right]\right. \\
&\quad + \sum_{i=1}^{p} \nu_i^2 g_{pm+6}\left[\chi^2_{pm}(\alpha)\right] \\
&\quad \left. - \left[\sum_{i=1}^{p} \nu_i^2 - \frac{p+m+1}{pm+2}\left(\sum_{i=1}^{p} \nu_i\right)^2\right] g_{pm+8}\left[\chi^2_{pm}(\alpha)\right]\right\} + o\left(\frac{1}{n}\right),
\end{aligned}
$$

$$
\begin{aligned}
(30)\quad & \Pr\left\{\operatorname{tr} \boldsymbol{H}(\boldsymbol{H}+\boldsymbol{G})^{-1} \geq v_{p,m,n}(\alpha)\right\} \\
&= 1 - G_{pm}\left[\chi^2_{pm}(\alpha) \,\middle|\, \sum_{i=1}^{p} \nu_i\right] \\
&\quad - \frac{1}{2n}\Bigg\{(p+m+1)\sum_{i=1}^{p} \nu_i g_{pm+4}\left[\chi^2_{pm}(\alpha)\right] \\
&\quad + \sum_{i=1}^{p} \nu_i^2 g_{pm+6}\left[\chi^2_{pm}(\alpha)\right] \\
&\quad + \left[\sum_{i=1}^{p} \nu_i^2 - \frac{p+m+1}{pm+2}\left(\sum_{i=1}^{p}\nu_i\right)^2\right] g_{pm+8}\left[\chi^2_{pm}(\alpha)\right]\Bigg\} + o\left(\frac{1}{n}\right),
\end{aligned}
$$

where $G_f(x|y)$ is the noncentral χ^2-distribution with f degrees of freedom and noncentrality parameter y, and $g_f(x)$ is the (central) χ^2-density with f degrees of freedom. The leading terms are the noncentral χ^2-distribution; the power functions of the three tests agree to this order. The power functions of the two trace tests differ from that of the likelihood ratio test by $\pm g_{pm+8}[\chi^2_{pm}(\alpha)]/(2n)$ times

$$
(31)\quad \sum_{i=1}^{p} \nu_i^2 - \frac{p+m+1}{pm+2}\left(\sum_{i=1}^{p}\nu_i\right)^2 = \sum_{i=1}^{p}(\nu_i - \bar{\nu})^2 - \frac{p(p-1)(p+2)}{pm+2}\bar{\nu}^2,
$$

where $\bar{\nu} = \sum_{i=1}^{p} \nu_i / p$. This is positive if

$$
(32)\qquad \frac{\sigma_\nu}{\bar{\nu}} > \sqrt{\frac{(p-1)(p+2)}{pm+2}},
$$

where $\sigma_\nu^2 = \sum_{i=1}^{p}(\nu_i - \bar{\nu})^2/p$ is the (population) variance of $\nu_1 \dots, \nu_p$; the left-hand side of (32) is the coefficient of variation. If the ν_i's are relatively variable in the sense that (32) holds, the power of the Lawley–Hotelling trace test is greater than that of the likelihood ratio test, which in turn is greater than that of the Bartlett–Nanda–Pillai trace test (to order $1/n$); if the inequality (32) is reversed, the ordering of power is reversed.

The differences between the powers decrease as n increases for fixed $\nu_1, \dots, \nu_p$. (However, this comparison is not very meaningful, because increasing n decreases $\boldsymbol{\beta}_1^N - \boldsymbol{\beta}_1^*$ and increases $Z'Z$.)

A number of numerical comparisons have been made. Schatzoff (1966b) and Olson (1974) have used Monte Carlo methods; Mikhail (1965), Pillai and Jayachandran (1967), and Lee (1971a) have used asymptotic expansions of distributions. All of these results agree with Rothenberg's. Among these three procedures, the Bartlett–Nanda–Pillai trace test is to be preferred if the roots are roughly equal in the alternative, and the Lawley–Hotelling trace is more powerful when the roots are substantially unequal. Wilks's likelihood ratio test seems to come in second best; in a sense it is maximin.

As noted in Section 8.6.4, the Roy largest root has a limiting distribution which is not a χ^2-distribution under the null hypothesis and is not a noncentral χ^2-distribution under a sequence of alternative hypotheses. Hence the comparison of Rothenberg cannot be extended to this case. In fact, the distributions in the nonnull case are difficult to evaluate. However, the Monte Carlo results of Schatzoff (1966b) and Olson (1974) are clear-cut. The maximum root test has greatest power if the alternative is one-dimensional, that is, if $\nu_2 = \cdots = \nu_p = 0$. On the other hand, if the alternative is not one-dimensional, then the maximum root test is inferior.

These test procedures tend to be robust. Under the null hypothesis the limiting distribution of $\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^*$ suitably normalized is normal with mean $\mathbf{0}$ and covariances the same as if X were normal, as long as its distribution satisfies some condition such as bounded fourth-order moments. Then $\hat{\boldsymbol{\Sigma}}_\Omega = (1/N)G$ converges with probability one. The limiting distribution of each criterion suitably normalized is the same as if X were normal. Olson (1974) studied the robustness under departures from covariance homogeneity as well as departures from normality. His conclusion was that the two trace tests and the likelihood ratio test were rather robust, and the maximum root test least robust. See also Pillai and Hsu (1979).

Berndt and Savin (1977) have noted that

$$\operatorname{tr} H(H+G)^{-1} \le \log U^{-1} \le \operatorname{tr} HG^{-1}. \tag{33}$$

(See Problem 8.19.) If the χ^2 significance point is used, then a larger criterion may lead to rejection while a smaller one may not.

8.7. TESTS OF HYPOTHESES ABOUT MATRICES OF REGRESSION COEFFICIENTS AND CONFIDENCE REGIONS

8.7.1. Testing Hypotheses

Suppose we are given a set of vector observations $x_1, \dots, x_N$ with accompanying fixed vectors $z_1, \dots, z_N$, where x_α is an observation from $N(\boldsymbol{\beta} z_\alpha, \boldsymbol{\Sigma})$. We let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1 \ \boldsymbol{\beta}_2)$ and $z'_\alpha = (z_\alpha^{(1)\prime}, z_\alpha^{(2)\prime})$, where $\boldsymbol{\beta}_1$ and $z_\alpha^{(1)\prime}$ have q_1 $(= q - q_2)$ columns. The null hypothesis is

$$H: \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^*, \tag{1}$$

where $\boldsymbol{\beta}_1^*$ is a specified matrix. Suppose the desired significance level is α. A test procedure is to compute

$$U = \frac{|N\hat{\boldsymbol{\Sigma}}_\Omega|}{|N\hat{\boldsymbol{\Sigma}}_\omega|} \tag{2}$$

and compare this number with $u_{p,q_1,n}(\alpha)$, the α significance point of the $U_{p,q_1,n}$-distribution. For $p = 2,\dots,10$ and even m, Table 1 in Appendix B can be used. For $m = 2,\dots,10$ and even p the same table can be used with m replaced by p and p replaced by m. (M as given in the table remains unchanged.) For p and m both odd, interpolation between even values of either p or m will give sufficient accuracy for most purposes. For reasonably large n, the asymptotic theory can be used. An equivalent procedure is to calculate $\Pr\{U_{p,m,n} \le U\}$; if this is less than α, the null hypothesis is rejected.

Alternatively one can use the Lawley–Hotelling trace criterion

$$\begin{aligned} W &= \operatorname{tr}\left(N\hat{\boldsymbol{\Sigma}}_\omega - N\hat{\boldsymbol{\Sigma}}_\Omega\right)\left(N\hat{\boldsymbol{\Sigma}}_\Omega\right)^{-1} \\ &= \operatorname{tr}\left(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*\right)A_{11\cdot2}\left(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*\right)'\left(N\hat{\boldsymbol{\Sigma}}_\Omega\right)^{-1}, \end{aligned} \tag{3}$$

the Pillai trace criterion

$$\begin{aligned} V &= \operatorname{tr}\left(N\hat{\boldsymbol{\Sigma}}_\omega - N\hat{\boldsymbol{\Sigma}}_\Omega\right)\left(N\hat{\boldsymbol{\Sigma}}_\omega\right)^{-1} \\ &= \operatorname{tr}\left(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*\right)A_{11\cdot2}\left(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*\right)'\left(N\hat{\boldsymbol{\Sigma}}_\omega\right)^{-1}, \end{aligned} \tag{4}$$

or the Roy maximum root criterion R, where R is the maximum root of

$$|N\hat{\boldsymbol{\Sigma}}_\omega - N\hat{\boldsymbol{\Sigma}}_\Omega - rN\hat{\boldsymbol{\Sigma}}_\Omega| = \left|\left(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*\right)A_{11\cdot2}\left(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*\right)' - rN\hat{\boldsymbol{\Sigma}}_\Omega\right| = 0. \tag{5}$$

These criteria can be referred to the appropriate tables in Appendix B.

We outline an approach to computing the criterion. If we let $\boldsymbol{y}_\alpha = \boldsymbol{x}_\alpha - \boldsymbol{\beta}_1^* \boldsymbol{z}_\alpha^{(1)}$, then $\boldsymbol{y}_\alpha$ can be considered as an observation from $N(\boldsymbol{\Delta}\boldsymbol{z}_\alpha, \boldsymbol{\Sigma})$, where $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1 \ \ \boldsymbol{\Delta}_2) = (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^* \ \ \boldsymbol{\beta}_2)$. Then the null hypothesis is $H: \boldsymbol{\Delta}_1 = \boldsymbol{0}$, and

$$\sum \boldsymbol{y}_\alpha \boldsymbol{y}_\alpha' = \sum \boldsymbol{x}_\alpha \boldsymbol{x}_\alpha' - \boldsymbol{\beta}_1^* \boldsymbol{C}_1' - \boldsymbol{C}_1 \boldsymbol{\beta}_1^{*\prime} + \boldsymbol{\beta}_1^* A_{11} \boldsymbol{\beta}_1^{*\prime}, \tag{6}$$

$$\sum \boldsymbol{y}_\alpha \boldsymbol{z}_\alpha' = \boldsymbol{C} - \boldsymbol{\beta}_1^*\left(A_{11} \quad A_{12}\right). \tag{7}$$

Thus the problem of testing the hypothesis $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^*$ is equivalent to testing the hypothesis $\boldsymbol{\Delta}_1 = \boldsymbol{0}$, where $\mathscr{E}\boldsymbol{y}_\alpha = \boldsymbol{\Delta}\boldsymbol{z}_\alpha$. Hence let us suppose the problem is testing the hypothesis $\boldsymbol{\beta}_1 = \boldsymbol{0}$. Then $N\hat{\boldsymbol{\Sigma}}_\omega = \sum \boldsymbol{x}_\alpha \boldsymbol{x}_\alpha' - \hat{\boldsymbol{\beta}}_{2\omega} A_{22} \hat{\boldsymbol{\beta}}_{2\omega}'$ and

$N\hat{\boldsymbol{\Sigma}}_\Omega = \sum x_\alpha x'_\alpha - \hat{\boldsymbol{\beta}}_\Omega A \hat{\boldsymbol{\beta}}'_\Omega$. We have discussed in Section 8.2.2 the computation of $\hat{\boldsymbol{\beta}}_\Omega A \hat{\boldsymbol{\beta}}'_\Omega$ and hence $N\hat{\boldsymbol{\Sigma}}_\Omega$. Then $\hat{\boldsymbol{\beta}}_{2\omega} A_{22} \hat{\boldsymbol{\beta}}'_{2\omega}$ can be computed in a similar manner. If the method is laid out as

$$(8) \qquad \begin{pmatrix} A_{22} & A_{21} \\ A_{12} & A_{11} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}}'_{2\Omega} \\ \hat{\boldsymbol{\beta}}'_{1\Omega} \end{pmatrix} = \begin{pmatrix} C'_2 \\ C'_1 \end{pmatrix},$$

the first q_2 rows and columns of A^* and of A^{**} are the same as the result of applying the forward solution to the left-hand side of

$$(9) \qquad A_{22} \hat{\boldsymbol{\beta}}'_{2\omega} = C'_2,$$

and the first q_2 rows of $\bar{C}^*$ and $\bar{C}^{**}$ are the same as the result of applying the forward solution to the right-hand side of (9). Thus $\hat{\boldsymbol{\beta}}_{2\omega} A_{22} \hat{\boldsymbol{\beta}}'_{2\omega} = \bar{C}_2^{*} \bar{C}_2^{**\prime}$, where $\bar{C}^{*\prime} = (\bar{C}_2^{*\prime} \ \bar{C}_1^{*\prime})$ and $\bar{C}^{**\prime} = (\bar{C}_2^{**\prime} \ \bar{C}_1^{**\prime})$.

The method implies a method for computing a determinant. In Section A.5 of the Appendix it is shown that the result of the forward solution is $FA = A^*$. Thus $|F| \cdot |A| = |A^*|$. Since the determinant of a triangular matrix is the product of its diagonal elements, $|F| = 1$ and $|A| = |A^*| = \prod_{i=1}^{q} a_{ii}^*$. This result holds for any positive definite matrix in place of A (with a suitable modification of F) and hence can be used to compute $|N\hat{\boldsymbol{\Sigma}}_\Omega|$ and $|N\hat{\boldsymbol{\Sigma}}_\omega|$.

8.7.2. Confidence Regions Based on U

We have considered tests of hypotheses $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^*$, where $\boldsymbol{\beta}_1^*$ is specified. In the usual way we can deduce from the family of tests a confidence region for $\boldsymbol{\beta}_1$. From the theory given before, we know that the probability is $1 - \alpha$ of drawing a sample so that

$$(10) \qquad \frac{|N\hat{\boldsymbol{\Sigma}}_\Omega|}{\left|N\hat{\boldsymbol{\Sigma}}_\Omega + (\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1) A_{11\cdot 2} (\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1)'\right|} \ge u_{p,q_1,n}(\alpha).$$

Thus if we make the confidence-region statement that $\boldsymbol{\beta}_1$ satisfies

$$(11) \qquad \frac{|N\hat{\boldsymbol{\Sigma}}_\Omega|}{\left|N\hat{\boldsymbol{\Sigma}}_\Omega + (\hat{\boldsymbol{\beta}}_{1\Omega} - \bar{\boldsymbol{\beta}}_1) A_{11\cdot 2} (\hat{\boldsymbol{\beta}}_{1\Omega} - \bar{\boldsymbol{\beta}}'_1)'\right|} \ge u_{p,q_1,n}(\alpha),$$

where (11) is interpreted as an inequality on $\boldsymbol{\beta}_1 = \bar{\boldsymbol{\beta}}_1$, then the probability is $1 - \alpha$ of drawing a sample such that the statement is true.

Theorem 8.7.1. *The region* (11) *in the* $\bar{\boldsymbol{\beta}}_1$*-space is a confidence region for* $\boldsymbol{\beta}_1$ *with confidence coefficient* $1 - \alpha$.

Usually the set of $\bar{\boldsymbol{\beta}}_1$ satisfying (11) is difficult to visualize. However, the inequality can be used to determine whether trial matrices are included in the region.

8.7.3. Simultaneous Confidence Intervals Based on the Lawley–Hotelling Trace

Each test procedure implies a set of confidence regions. The Lawley–Hotelling trace criterion can be used to develop simultaneous confidence intervals for linear combinations of elements of $\boldsymbol{\beta}_1$. A confidence region with confidence coefficient $1-\alpha$ is

$$\text{(12)} \qquad \operatorname{tr}\left(\hat{\boldsymbol{\beta}}_{1\Omega}-\bar{\boldsymbol{\beta}}_1'\right)A_{11\cdot 2}\left(\hat{\boldsymbol{\beta}}_{1\Omega}-\bar{\boldsymbol{\beta}}_1'\right)'\left(N\hat{\boldsymbol{\Sigma}}_\Omega\right)^{-1} \le w_{p,m,n}(\alpha).$$

To derive the confidence bounds we generalize Lemma 5.3.2.

Lemma 8.7.1. *For positive definite matiices A and G,*

$$\text{(13)} \qquad |\operatorname{tr}\boldsymbol{\Phi}'Y| \le \sqrt{\operatorname{tr}A^{-1}\boldsymbol{\Phi}'G\boldsymbol{\Phi}}\sqrt{\operatorname{tr}AY'G^{-1}Y}.$$

Proof. Let $b=\operatorname{tr}\boldsymbol{\Phi}'Y/\operatorname{tr}A^{-1}\boldsymbol{\Phi}'G\boldsymbol{\Phi}$. Then

$$\text{(14)} \qquad \begin{aligned} 0 &\le \operatorname{tr}A(Y-bG\boldsymbol{\Phi}A^{-1})'G^{-1}(Y-bG\boldsymbol{\Phi}A^{-1}) \\ &= \operatorname{tr}AY'G^{-1}Y - b\operatorname{tr}\boldsymbol{\Phi}'Y - b\operatorname{tr}Y'\boldsymbol{\Phi} + b^2\operatorname{tr}\boldsymbol{\Phi}'G\boldsymbol{\Phi}A^{-1} \\ &= \operatorname{tr}AY'G^{-1}Y - \frac{(\operatorname{tr}\boldsymbol{\Phi}'Y)^2}{\operatorname{tr}A^{-1}\boldsymbol{\Phi}'G\boldsymbol{\Phi}}, \end{aligned}$$

which yields (13). ■

Now (12) and (13) imply that

$$\text{(15)} \qquad \begin{aligned} \left|\operatorname{tr}\boldsymbol{\Phi}'\hat{\boldsymbol{\beta}}_{1\Omega} - \operatorname{tr}\boldsymbol{\Phi}'\bar{\boldsymbol{\beta}}_1\right| &= \left|\operatorname{tr}\boldsymbol{\Phi}'\left(\hat{\boldsymbol{\beta}}_{1\Omega}-\bar{\boldsymbol{\beta}}_1\right)\right| \\ &\le \sqrt{\operatorname{tr}A_{11\cdot 2}^{-1}\boldsymbol{\Phi}'N\hat{\boldsymbol{\Sigma}}_\Omega\boldsymbol{\Phi}\cdot\operatorname{tr}A_{11\cdot 2}\left(\hat{\boldsymbol{\beta}}_{1\Omega}-\bar{\boldsymbol{\beta}}_1\right)'\left(N\hat{\boldsymbol{\Sigma}}_\Omega\right)^{-1}\left(\hat{\boldsymbol{\beta}}_{1\Omega}-\bar{\boldsymbol{\beta}}_1\right)} \\ &\le \sqrt{\operatorname{tr}A_{11\cdot 2}^{-1}\boldsymbol{\Phi}'N\hat{\boldsymbol{\Sigma}}_\Omega\boldsymbol{\Phi}}\sqrt{w_{p,m,n}(\alpha)} \end{aligned}$$

holds for all $p\times m$ matrices $\boldsymbol{\Phi}$. We assert that

$$\text{(16)} \qquad \begin{aligned} \operatorname{tr}\boldsymbol{\Phi}'\hat{\boldsymbol{\beta}}_{1\Omega} - \sqrt{N\operatorname{tr}A_{11\cdot 2}^{-1}\boldsymbol{\Phi}'\hat{\boldsymbol{\Sigma}}_\Omega\boldsymbol{\Phi}}\sqrt{w_{p,m,n}(\alpha)} &\le \operatorname{tr}\boldsymbol{\Phi}'\bar{\boldsymbol{\beta}}_1 \\ &\le \operatorname{tr}\boldsymbol{\Phi}'\hat{\boldsymbol{\beta}}_{1\Omega} + \sqrt{N\operatorname{tr}A_{11\cdot 2}^{-1}\boldsymbol{\Phi}'\hat{\boldsymbol{\Sigma}}_\Omega\boldsymbol{\Phi}}\sqrt{w_{p,m,n}(\alpha)} \end{aligned}$$

holds for all $\boldsymbol{\Phi}$ with confidence $1-\alpha$.

The confidence region (12) can be explored by use of (16) for various Φ. If $\phi_{ik} = 1$ for some pair (I, K) and 0 for other elements, then (16) gives an interval for β_{IK}. If $\phi_{ik} = 1$ for a pair (I, K), -1 for (I, L), and 0 otherwise, the interval pertains to $\beta_{IK} - \beta_{IL}$, the difference of coefficients of two independent variables. If $\phi_{ik} = 1$ for a pair (I, K), -1 for (J, K), and 0 otherwise, one obtains an interval for $\beta_{IK} - \beta_{JK}$, the difference of coefficients for two dependent variables.

8.7.4. Simultaneous Confidence Intervals Based on the Roy Maximum Root Criterion

A confidence region with confidence $1 - \alpha$ based on the maximum root criterion is

(17) $$\mathrm{ch}_1\big(\hat{\boldsymbol{\beta}}_{1\Omega} - \overline{\boldsymbol{\beta}}_1\big)A_{11\cdot 2}\big(\hat{\boldsymbol{\beta}}_{1\Omega} - \overline{\boldsymbol{\beta}}_1\big)'\big(N\hat{\boldsymbol{\Sigma}}_\Omega\big)^{-1} \le r_{p,m,n}(\alpha),$$

where $\mathrm{ch}_1(C)$ denotes the largest characteristic root of C. We can derive simultaneous confidence bounds from (17). From Lemma 5.3.2, we find for any vectors $\boldsymbol{a}$ and $\boldsymbol{b}$

(18)
$$\begin{aligned}
\big[\boldsymbol{a}'(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1)\boldsymbol{b}\big]^2 &= \big\{\big[(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1)'\boldsymbol{a}\big]'\boldsymbol{b}\big\}^2 \\
&\le \big[(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1)'\boldsymbol{a}\big]'A_{11\cdot 2}\big[(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1)'\boldsymbol{a}\big]\cdot \boldsymbol{b}'A_{11\cdot 2}^{-1}\boldsymbol{b} \\
&= \frac{\boldsymbol{a}'(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1)A_{11\cdot 2}(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1)'\boldsymbol{a}}{\boldsymbol{a}'G\boldsymbol{a}}\cdot \boldsymbol{a}'G\boldsymbol{a}\cdot \boldsymbol{b}'A_{11\cdot 2}^{-1}\boldsymbol{b} \\
&\le \mathrm{ch}_1\big[(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1)A_{11\cdot 2}(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1)'G^{-1}\big]\cdot \boldsymbol{a}'G\boldsymbol{a}\cdot \boldsymbol{b}'A_{11\cdot 2}^{-1}\boldsymbol{b} \\
&\le r_{p,m,n}(\alpha)\cdot \boldsymbol{a}'G\boldsymbol{a}\cdot \boldsymbol{b}'A_{11\cdot 2}^{-1}\boldsymbol{b}
\end{aligned}$$

with probability $1 - \alpha$; the second inequality follows from Theorem A.2.4 of the Appendix. Then a set of confidence intervals on all linear combinations $\boldsymbol{a}'\boldsymbol{\beta}_1\boldsymbol{b}$ holding with confidence $1 - \alpha$ is

(19) $$\begin{aligned}
\boldsymbol{a}'\hat{\boldsymbol{\beta}}_{1\Omega}\boldsymbol{b} - \sqrt{r_{p,m,n}(\alpha)\cdot \boldsymbol{a}'G\boldsymbol{a}\cdot \boldsymbol{b}'A_{11\cdot 2}^{-1}\boldsymbol{b}} &\le \boldsymbol{a}'\overline{\boldsymbol{\beta}}_1\boldsymbol{b} \\
&\le \boldsymbol{a}'\hat{\boldsymbol{\beta}}_{1\Omega}\boldsymbol{b} + \sqrt{r_{p,m,n}(\alpha)\cdot \boldsymbol{a}'G\boldsymbol{a}\cdot \boldsymbol{b}'A_{11\cdot 2}^{-1}\boldsymbol{b}}\,.
\end{aligned}$$

The linear combinations are $\boldsymbol{a}'\boldsymbol{\beta}_1\boldsymbol{b} = \sum_{i=1}^{p}\sum_{h=1}^{m} a_i\beta_{ih}b_h$. If $a_1 = 1$, $a_i = 0$, $i \neq 1$, and $b_1 = 1$, $b_h = 0$, $h \neq 1$, the linear combination is simply β_{11}. If $a_1 = 1$, $a_i = 0$, $i \neq 1$, and $b_1 = 1$, $b_2 = -1$, $b_h = 0$, $h \neq 1, 2$, the linear combination is $\beta_{11} - \beta_{12}$.

We can compare these intervals with (16) for $\Phi = ab'$, which is of rank 1. The term subtracted from and added to tr $\Phi'\hat{\beta}_{1\Omega} = a'\hat{\beta}_{1\Omega}b$ is the square root of

$$(20) \qquad w_{p,m,n}(\alpha) \cdot \operatorname{tr} A_{11\cdot 2}^{-1} ba'Gab' = w_{p,m,n}(\alpha) \cdot a'Ga \cdot b'A_{11\cdot 2}^{-1}b.$$

This is greater than the term subtracted and added to $a'\hat{\beta}_{1\Omega}b$ in (19) because $w_{p,m,n}(\alpha)$, pertaining to the sum of the roots, is greater than $r_{p,m,n}(\alpha)$, relating to one root. The bounds (16) hold for all $p \times m$ matrices Φ, while (19) holds only for matrices ab' of rank 1.

Mudholkar (1966) gives a very general method of constructing simultaneous confidence intervals based on symmetric gauge functions. Gabriel (1969) relates confidence bounds to simultaneous test procedures. Wijsman (1979) showed that under certain conditions the confidence sets based on the maximum root are smallest. [See also Wijsman (1980).]

8.8. TESTING EQUALITY OF MEANS OF SEVERAL NORMAL DISTRIBUTIONS WITH COMMON COVARIANCE MATRIX

In univariate analysis it is well known that many hypotheses can be put in the form of hypotheses concerning regression coefficients. The same is true for the corresponding multivariate cases. As an example we consider testing the hypothesis that the means of, say, q normal distributions with a common covariance matrix are equal.

Let $y_\alpha^{(i)}$ be an observation from $N(\mu^{(i)}, \Sigma)$, $\alpha = 1, \ldots, N_i$, $i = 1, \ldots, q$. The null hypothesis is

$$(1) \qquad H: \mu^{(1)} = \cdots = \mu^{(q)}.$$

To put the problem in the form considered earlier in this chapter, let

$$(2) \qquad X = (x_1\ x_2 \cdots x_{N_1}\ x_{N_1+1} \cdots x_N) = \left(y_1^{(1)}\ y_2^{(2)} \cdots y_{N_1}^{(1)}\ y_1^{(1)} \cdots y_{N_q}^{(q)}\right)$$

with $N = N_1 + \cdots + N_q$. Let

$$(3) \qquad Z = (z_1\quad z_2\quad \cdots\quad z_{N_1}\quad z_{N_1+1}\quad \cdots\quad z_N) = \begin{pmatrix} 1 & 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 1 & 1 & \cdots & 1 \end{pmatrix};$$

that is, $z_{i\alpha} = 1$ if $N_1 + \cdots + N_{i-1} < \alpha \le N_1 + \cdots + N_i$, and $z_{i\alpha} = 0$ otherwise, for $i = 1, \ldots, q-1$, and $z_{q\alpha} = 1$ (all α). Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1 \ \boldsymbol{\beta}_2)$, where

$$\begin{aligned} \boldsymbol{\beta}_1 &= \left(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(q)}, \ldots, \boldsymbol{\mu}^{(q-1)} - \boldsymbol{\mu}^{(q)}\right), \\ \boldsymbol{\beta}_2 &= \boldsymbol{\mu}^{(q)}. \end{aligned} \tag{4}$$

Then $\boldsymbol{x}_\alpha$ is an observation from $N(\boldsymbol{\beta} z_\alpha, \boldsymbol{\Sigma})$, and the null hypothesis is $\boldsymbol{\beta}_1 = \mathbf{0}$. Thus we can use the above theory for finding the criterion for testing the hypothesis.

We have

$$A = \sum_{\alpha=1}^{N} z_\alpha z_\alpha' = \begin{pmatrix} N_1 & 0 & \cdots & 0 & N_1 \\ 0 & N_2 & \cdots & 0 & N_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & N_{q-1} & N_{q-1} \\ N_1 & N_2 & \cdots & N_{q-1} & N \end{pmatrix}, \tag{5}$$

$$C = \sum_{\alpha=1}^{N} \boldsymbol{x}_\alpha z_\alpha' = \left(\sum_\alpha \boldsymbol{y}_\alpha^{(1)} \quad \sum_\alpha \boldsymbol{y}_\alpha^{(2)} \cdots \sum_\alpha \boldsymbol{y}_\alpha^{(q-1)} \quad \sum_{i,\alpha} \boldsymbol{y}_\alpha^{(i)} \right). \tag{6}$$

Here $A_{22} = N$ and $\boldsymbol{C}_2 = \sum_{i,\alpha} \boldsymbol{y}_\alpha^{(i)}$. Thus $\hat{\boldsymbol{\beta}}_{2\omega} = \sum_{i,\alpha} \boldsymbol{y}_\alpha^{(i)} \cdot (1/N) = \bar{\boldsymbol{y}}$, say, and

$$\begin{aligned} N\hat{\boldsymbol{\Sigma}}_\omega &= \sum_\alpha \boldsymbol{x}_\alpha \boldsymbol{x}_\alpha' - \bar{\boldsymbol{y}} N \bar{\boldsymbol{y}}' \\ &= \sum_{i,\alpha} \boldsymbol{y}_\alpha^{(i)} \boldsymbol{y}_\alpha^{(i)\prime} - N\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}' \\ &= \sum_{i,\alpha} \left(\boldsymbol{y}_\alpha^{(i)} - \bar{\boldsymbol{y}}\right)\left(\boldsymbol{y}_\alpha^{(i)} - \bar{\boldsymbol{y}}\right)'. \end{aligned} \tag{7}$$

For $\hat{\boldsymbol{\Sigma}}_\Omega$, we use the formula $N\hat{\boldsymbol{\Sigma}}_\Omega = \sum \boldsymbol{x}_\alpha \boldsymbol{x}_\alpha' - \hat{\boldsymbol{\beta}}_\Omega A \hat{\boldsymbol{\beta}}_\Omega' = \sum \boldsymbol{x}_\alpha \boldsymbol{x}_\alpha' - \boldsymbol{C} A^{-1} \boldsymbol{C}'$. Let

$$\boldsymbol{D} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ -1 & -1 & \cdots & -1 & 1 \end{pmatrix}; \tag{8}$$

then

$$D^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}. \tag{9}$$

Thus

$$\begin{aligned}
CA^{-1}C' &= CD'D^{-1\prime}A^{-1}D^{-1}DC' \\
&= CD'(DAD')^{-1}DC' \\
&= \left(\sum_\alpha y_\alpha^{(1)} \cdots \sum_\alpha y_\alpha^{(q)}\right) \begin{pmatrix} N_1 & 0 & \cdots & 0 \\ 0 & N_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & N_q \end{pmatrix}^{-1} \begin{pmatrix} \sum_\alpha y_\alpha^{(1)\prime} \\ \vdots \\ \sum_\alpha y_\alpha^{(q)\prime} \end{pmatrix} \\
&= \sum_i \left(\sum_\alpha y_\alpha^{(i)} \frac{1}{N_i} \sum_\gamma y_\gamma^{(i)\prime} \right) \\
&= \sum_i N_i \bar{y}^{(i)} \bar{y}^{(i)\prime},
\end{aligned} \tag{10}$$

where $\bar{y}^{(i)} = (1/N_i)\Sigma_\alpha y_\alpha^{(i)}$. Thus

$$\begin{aligned}
N\hat{\Sigma}_\Omega &= \sum_{i,\alpha} y_\alpha^{(i)} y_\alpha^{(i)\prime} - \sum_i N_i \bar{y}^{(i)} \bar{y}^{(i)\prime} \\
&= \sum_{i,\alpha} \left(y_\alpha^{(i)} - \bar{y}^{(i)}\right)\left(y_\alpha^{(i)} - \bar{y}^{(i)}\right)'.
\end{aligned} \tag{11}$$

It will be seen that $\hat{\Sigma}_\omega$ is the estimator of Σ when $\mu^{(1)} = \cdots = \mu^{(q)}$ and $\hat{\Sigma}_\Omega$ is the weighted average of the estimators of Σ based on the separate samples.

When the null hypothesis is true, $|N\hat{\Sigma}_\Omega| / |N\hat{\Sigma}_\omega|$ is distributed as $U_{p,q-1,n}$, where $n = N - q$. Therefore, the rejection region at the α significance level is

$$\lambda = \frac{|N\hat{\Sigma}_\Omega|}{|N\hat{\Sigma}_\omega|} < u_{p,q-1,n}(\alpha). \tag{12}$$

The left-hand side of (12) is (11) of Section 8.3, and

$$(13)\quad N\hat{\boldsymbol{\Sigma}}_\omega - N\hat{\boldsymbol{\Sigma}}_\Omega = \sum_{i,\alpha} y_\alpha^{(i)} y_\alpha^{(i)\prime} - N\bar{y}\bar{y}' - \left(\sum_{i,\alpha} y_\alpha^{(i)} y_\alpha^{(i)\prime} - \sum_i N_i \bar{y}^{(i)} \bar{y}^{(i)\prime} \right)$$
$$= \sum N_i (\bar{y}^{(i)} - \bar{y})(\bar{y}^{(i)} - \bar{y})' = \boldsymbol{H},$$

as implied by (4) and (5) of Section 8.4. Here $\boldsymbol{H}$ has the distribution $W(\boldsymbol{\Sigma}, q-1)$. It will be seen that when $p = 1$, this test reduces to the usual F-test

$$(14)\qquad \frac{\sum N_i (\bar{y}^{(i)} - \bar{y})^2}{\left(y_\alpha^{(i)} - \bar{y}^{(i)}\right)^2} \cdot \frac{n}{q-1} > F_{q-1,n}(\alpha).$$

We give an example of the analysis. The data are taken from Barnard's study of Egyptian skulls (1935). The 4 $(=q)$ populations are Late Predynastic $(i = 1)$, Sixth to Twelfth $(i = 2)$, Twelfth to Thirteenth $(i = 3)$, and Ptolemaic Dynasties $(i = 4)$. The 4 $(=p)$ measurements (i.e., components of $y_\alpha^{(i)}$) are maximum breadth, basialveolar length, nasal height, and basibregmatic height. The numbers of observations are $N_1 = 91$, $N_2 = 162$, $N_3 = 70$, $N_4 = 75$. The data are sumn arized as

$$(15)\quad \begin{pmatrix} \bar{y}^{(1)} & \bar{y}^{(2)} & \bar{y}^{(3)} & \bar{y}^{(4)} \end{pmatrix}$$
$$= \begin{pmatrix} 133.582418 & 134.265432 & 134.371429 & 135.306667 \\ 98.307692 & 96.462963 & 95.857143 & 95.040000 \\ 50.835165 & 51.148148 & 50.100000 & 52.093333 \\ 133.000000 & 134.882716 & 133.642857 & 131.466667 \end{pmatrix}.$$

$$(16)\quad N\hat{\boldsymbol{\Sigma}}_\Omega$$
$$= \begin{pmatrix} 9661.997470 & 445.573301 & 1130.623900 & 2148.584210 \\ 445.573301 & 9073.115027 & 1239.211990 & 2255.812722 \\ 1130.623900 & 1239.211990 & 3938.320351 & 1271.054662 \\ 2148.584210 & 2255.812722 & 1271.054662 & 8741.508829 \end{pmatrix}.$$

From these data we find

$$(17)\quad N\hat{\boldsymbol{\Sigma}}_\omega$$
$$= \begin{pmatrix} 9785.178098 & 214.197666 & 1217.929248 & 2019.820216 \\ 214.197666 & 9559.460890 & 1131.716372 & 2381.126040 \\ 1217.929248 & 1131.716372 & 4088.731856 & 1133.473898 \\ 2019.820216 & 2381.126040 & 1133.473898 & 9382.242720 \end{pmatrix}.$$

We shall use the likelihood ratio test. The ratio of determinants is

$$U = \frac{|N\hat{\Sigma}_\Omega|}{|N\hat{\Sigma}_\omega|} = \frac{2.426\,905\,4 \times 10^5}{2.954\,447\,5 \times 10^5} = 0.821\,434\,4. \tag{18}$$

Here $N = 398$, $n = 394$, $p = 4$, and $q = 4$. Thus $k = 393$. Since n is very large, we may assume $-k \log U_{4,3,394}$ is distributed as χ^2 with 12 degrees of freedom (when the null hypothesis is true). Here $-k \log U = 77.30$. Since the 1% point of the χ^2_{12}-distribution is 26.2, the hypothesis of $\boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)} = \boldsymbol{\mu}^{(3)} = \boldsymbol{\mu}^{(4)}$ is rejected.†

8.9. MULTIVARIATE ANALYSIS OF VARIANCE

The univariate analysis of variance has a direct generalization for vector variables leading to an analysis of vector sums of squares (i.e., sums such as $\sum \boldsymbol{x}_\alpha \boldsymbol{x}'_\alpha$). In fact, in the preceding section this generalization was considered for an analysis of variance problem involving a single classification.

As another example consider a two-way layout. Suppose that we are interested in the question whether the column effects are zero. We shall review the analysis for a scalar variable and then show the analysis for a vector variable. Let Y_{ij}, $i = 1, \ldots, r$, $j = 1, \ldots, c$, be a set of rc random variables. We assume that

$$\mathscr{E} Y_{ij} = \mu + \lambda_i + \nu_j, \qquad i = 1, \ldots, r, \quad j = 1, \ldots, c, \tag{1}$$

with the restrictions

$$\sum_{i=1}^{r} \lambda_i = \sum_{j=1}^{c} \nu_j = 0, \tag{2}$$

that the variance of Y_{ij} is σ^2, and that the Y_{ij} are independently normally distributed. To test that column effects are zero is to test that

$$\nu_j = 0, \qquad j = 1, \ldots, c. \tag{3}$$

This problem can be treated as a problem of regression by the introduction

†The above computations were given by Bartlett (1947).

of dummy fixed variates. Let

$$
\begin{aligned}
(4)\qquad z_{00,ij} &= 1, && \\
z_{k0,ij} &= 1, && k=i, \\
&= 0, && k \neq i, \\
z_{0k,ij} &= 1, && k=j, \\
&= 0, && k \neq j.
\end{aligned}
$$

Then (1) can be written

$$
(5)\qquad \mathscr{E}Y_{ij} = \mu z_{00,ij} + \sum_{k=1}^{r} \lambda_k z_{k0,ij} + \sum_{k=1}^{c} \nu_k z_{0k,ij}.
$$

The hypothesis is that the coefficients of $z_{0k,ij}$ are zero. Since the matrix of fixed variates here,

$$
(6)\qquad \begin{pmatrix} z_{00,11} & \cdots & z_{00,rc} \\ z_{10,11} & \cdots & z_{10,rc} \\ z_{20,11} & \cdots & z_{20,rc} \\ \vdots & & \vdots \\ z_{0c,11} & \cdots & z_{0c,rc} \end{pmatrix},
$$

is singular (for example, row 00 is the sum of rows $10, 20, \ldots, r0$), one must elaborate the regression theory. When one does, one finds that the test criterion indicated by the regression theory is the usual F-test of analysis of variance.

Let

$$
\begin{aligned}
Y_{..} &= \frac{1}{rc} \sum_{i,j} Y_{ij}, \\
(7)\qquad Y_{i.} &= \frac{1}{c} \sum_{j} Y_{ij}, \\
Y_{.j} &= \frac{1}{r} \sum_{i} Y_{ij},
\end{aligned}
$$

and let

$$
\begin{aligned}
a &= \sum_{i,j} (Y_{ij} - Y_{i.} - Y_{.j} + Y_{..})^2 \\
&= \sum_{i,j} Y_{ij}^2 - c \sum_i Y_{i.}^2 - r \sum_j Y_{.j}^2 + rcY_{..}^2, \\
b &= r \sum_j (Y_{.j} - Y_{..})^2 \\
&= r \sum_j Y_{.j}^2 - rcY_{..}^2 .
\end{aligned} \tag{8}
$$

Then the F-statistic is given by

$$F = \frac{b}{a} \cdot \frac{(c-1)(r-1)}{c-1}. \tag{9}$$

Under the null hypothesis, this has the F-distribution with $c-1$ and $(r-1)\cdot(c-1)$ degrees of freedom. The likelihood ratio criterion for the hypothesis is the $rc/2$ power of

$$\frac{a}{a+b} = \frac{1}{1 + \{(c-1)/[(r-1)(c-1)]\}F}. \tag{10}$$

Now let us turn to the multivariate analysis of variance. We have a set of p-dimensional random vectors Y_{ij}, $i = 1, \ldots, r$, $j = 1, \ldots, c$, with expected values (1), where $\boldsymbol{\mu}$, the $\boldsymbol{\lambda}$'s, and the $\boldsymbol{\nu}$'s are vectors, and with covariance matrix $\boldsymbol{\Sigma}$, and they are independently normally distributed. Then the same algebra may be used to reduce this problem to the regression problem. We define $Y_{..}, Y_{i.}, Y_{.j}$ by (7) and

$$
\begin{aligned}
A &= \sum_{i,j} (Y_{ij} - Y_{i.} - Y_{.j} + Y_{..})(Y_{ij} - Y_{i.} - Y_{.j} + Y_{..})' \\
&= \sum_{i,j} Y_{ij}Y_{ij}' - c \sum_i Y_{i.}Y_{i.}' - r \sum_j Y_{.j}Y_{.j}' + rcY_{..}Y_{..}', \\
B &= r \sum_j (Y_{.j} - Y_{..})(Y_{.j} - Y_{..})' \\
&= r \sum_j Y_{.j}Y_{.j}' - rcY_{..}Y_{..}' .
\end{aligned} \tag{11}
$$

Table 8.1

	Varieties					
Location	M	S	V	T	P	Sums
UF	81	105	120	110	98	514
	81	82	80	87	84	414
W	147	142	151	192	146	778
	100	116	112	148	108	584
M	82	77	78	131	90	458
	103	105	117	140	130	595
C	120	121	124	141	125	631
	99	62	96	126	76	459
GR	99	89	69	89	104	450
	66	50	97	62	80	355
D	87	77	79	102	96	441
	68	67	67	92	94	338
Sums	616	611	621	765	659	3272
	517	482	569	655	572	2795

A statistic analogous to (10) is

$$\frac{|A|}{|A+B|}. \tag{12}$$

Under the null hypothesis, this has the distribution of U for $p, n=(r-1)\cdot(c-1)$ and $q_1 = c-1$ given in Section 8.4. In order for A to be nonsingular (with probability 1), we must require $p \le (r-1)(c-1)$.

As an example we use data first published by Immer, Hayes, and Powers (1934), and later used by Fisher (1947a), by Yates and Cochran (1938), and by Tukey (1949). The first component of the observation vector is the barley yield in a given year; the second component is the same measurement made the following year. Column indices run over the varieties of barley, and row indices over the locations. The data are given in Table 8.1 [e.g., $\begin{smallmatrix}81\\81\end{smallmatrix}$ in the upper left-hand corner indicates a yield of 81 in each year of variety M in location UF]. The numbers along the borders are sums.

We consider the square of (147, 100) to be

$$\begin{pmatrix}147\\100\end{pmatrix}(147 \quad 100) = \begin{pmatrix}21{,}609 & 14{,}700\\ 14{,}700 & 10{,}000\end{pmatrix}.$$

Then

$$\sum_{i,j} Y_{ij}Y'_{ij} = \begin{pmatrix} 380{,}944 & 315{,}381 \\ 315{,}381 & 277{,}625 \end{pmatrix}, \tag{13}$$

$$\sum_{j} (6Y_{.j})(6Y_{.j})' = \begin{pmatrix} 2{,}157{,}924 & 1{,}844{,}346 \\ 1{,}844{,}346 & 1{,}579{,}583 \end{pmatrix}, \tag{14}$$

$$\sum_{i} (5Y_{i.})(5Y_{i.})' = \begin{pmatrix} 1{,}874{,}386 & 1{,}560{,}145 \\ 1{,}560{,}145 & 1{,}353{,}727 \end{pmatrix}, \tag{15}$$

$$(30Y_{..})(30Y_{..})' = \begin{pmatrix} 10{,}750{,}984 & 9{,}145{,}240 \\ 9{,}145{,}240 & 7.812{,}025 \end{pmatrix}. \tag{16}$$

Then the *error sum of squares* is

$$A = \begin{pmatrix} 3279 & 802 \\ 802 & 4017 \end{pmatrix}, \tag{17}$$

the *row sum of squares* is

$$5\sum_{j} (Y_{i.} - Y_{..})(Y_{i.} - Y_{..})' = \begin{pmatrix} 18{,}011 & 7.188 \\ 7{,}188 & 10{,}345 \end{pmatrix}, \tag{18}$$

and the *column sum of squares* is

$$B = \begin{pmatrix} 2788 & 2550 \\ 2550 & 2863 \end{pmatrix}. \tag{19}$$

The test criterion is

$$\frac{|A|}{|A+B|} = \frac{\begin{vmatrix} 3279 & 802 \\ 802 & 4017 \end{vmatrix}}{\begin{vmatrix} 6067 & 3352 \\ 3352 & 6880 \end{vmatrix}} = 0.4107. \tag{20}$$

This result is to be compared with the significant point for $U_{2,4,20}$. Using the result of Section 8.4, we see that

$$\frac{1-\sqrt{0.4107}}{\sqrt{0.4107}} \cdot \frac{19}{4} = 2.66$$

is to be compared with the significance point of $F_{3,38}$. This is significant at the 5% level. Our data show that there are differer ces between varieties.

Now let us see that each F-test in the univariate analysis of variance has analogous tests in the multivariate analysis of variance. In the linear hypothesis model for the univariate analysis of variance, one assumes that the random variables $Y_1, \ldots, Y_N$ have expected values that are linear combinations of unknown parameters

$$\mathscr{E} Y_\alpha = \sum_g \beta_g z_{g\alpha}, \tag{21}$$

where the β's are the parameters and the z's are the known coefficients. The variables $\{Y_\alpha\}$ are assumed to be normally and independently distributed with common variance σ^2. In this model there are a set of linear combinations, say $\sum_{\alpha=1}^N \gamma_{i\alpha} Y_\alpha$, where the γ's are known, such that

$$a = \sum_{i=1}^n \left(\sum_{\alpha=1}^N \gamma_{i\alpha} Y_\alpha \right)^2 = \sum_{\alpha,\beta=1}^N d_{\alpha\beta} Y_\alpha Y_\beta \tag{22}$$

is distributed as $\sigma^2 \chi^2$ with n degrees of freedom. There is another set of linear combinations, say $\sum_\alpha \phi_{g\alpha} Y_\alpha$, where the ϕ's are known, such that

$$b = \sum_{g=1}^m \left(\sum_{\alpha=1}^N \phi_{g\alpha} Y_\alpha \right)^2 = \sum_{\alpha,\beta=1}^N c_{\alpha\beta} Y_\alpha Y_\beta \tag{23}$$

is distributed as $\sigma^2 \chi^2$ with m degrees of freedom when the null hypothesis is true and as σ^2 times a noncentral χ^2 when the null hypothesis is not true; and in either case b is distributed independently of a. Then

$$\frac{b}{a} \cdot \frac{n}{m} = \frac{\sum c_{\alpha\beta} Y_\alpha Y_\beta}{\sum d_{\alpha\beta} Y_\alpha Y_\beta} \cdot \frac{n}{m} \tag{24}$$

has the F-distribution with m and n degrees of freedom, respectively, when the null hypothesis is true. The null hypothesis is that certain β's are zero.

In the multivariate analysis of variance, $Y_1, \ldots, Y_N$ are vector variables with p components. The expected value of Y_α is given by (21) where $\boldsymbol{\beta}_g$ is a vector of p parameters. We assume that the $\{Y_\alpha\}$ are normally and independently distributed with common covariance matrix $\boldsymbol{\Sigma}$. The linear combinations $\sum \gamma_{i\alpha} Y_\alpha$ can be formed for the vectors. Then

$$A = \sum_{i=1}^n \left(\sum_{\alpha=1}^N \gamma_{i\alpha} Y_\alpha \right) \left(\sum_{\alpha=1}^N \gamma_{i\alpha} Y_\alpha \right)' = \sum_{\alpha,\beta=1}^N d_{\alpha\beta} Y_\alpha Y_\beta' \tag{25}$$

has the distribution $W(\mathbf{\Sigma}, n)$. When the null hypothesis is true,

$$\mathbf{B} = \sum_{g=1}^{m}\left(\sum_{\alpha=1}^{N}\phi_{g\alpha}\mathbf{Y}_\alpha\right)\left(\sum_{\alpha=1}^{N}\phi_{g\alpha}\mathbf{Y}_\alpha\right)' = \sum_{\alpha,\beta=1}^{N} c_{\alpha\beta}\mathbf{Y}_\alpha\mathbf{Y}_\beta' \tag{26}$$

has the distribution $W(\mathbf{\Sigma}, m)$, and $\mathbf{B}$ is independent of $\mathbf{A}$. Then

$$\frac{|\mathbf{A}|}{|\mathbf{A}+\mathbf{B}|} = \frac{\left|\sum d_{\alpha\beta}\mathbf{Y}_\alpha\mathbf{Y}_\beta'\right|}{\left|\sum d_{\alpha\beta}\mathbf{Y}_\alpha\mathbf{Y}_\beta' + \sum c_{\alpha\beta}\mathbf{Y}_\alpha\mathbf{Y}_\beta'\right|} \tag{27}$$

has the $U_{p,m,n}$-distribution.

The argument for the distribution of a and b involves showing that $\mathscr{E}\sum_\alpha \gamma_{i\alpha}Y_\alpha = 0$ and $\mathscr{E}\sum_\alpha \phi_{g\alpha}Y_\alpha = 0$ when certain β's are equal to zero as specified by the null hypothesis (as identities in the unspecified β's). Clearly this argument holds for the vector case as well. Secondly, one argues, in the univariate case, that there is an orthogonal matrix $\mathbf{\Psi} = (\psi_{\alpha\beta})$ such that when the transformation $Y_\beta = \sum_\alpha \psi_{\beta\alpha}Z_\alpha$ is made

$$\begin{aligned} a &= \sum_{\alpha,\beta,\gamma,\delta} d_{\alpha\beta}\psi_{\alpha\gamma}\psi_{\beta\delta}Z_\gamma Z_\delta = \sum_{\alpha=1}^{n} Z_\alpha^2, \\ b &= \sum_{\alpha,\beta,\gamma,\delta} c_{\alpha\beta}\psi_{\alpha\gamma}\psi_{\beta\delta}Z_\gamma Z_\delta = \sum_{\alpha=n+1}^{n+m} Z_\alpha^2. \end{aligned} \tag{28}$$

Because the transformation is orthogonal, the $\{Z_\alpha\}$ are independently and normally distributed with common variance σ^2. Since the Z_α, $\alpha = 1, \dots, n$, must be linear combinations of $\sum_\alpha \gamma_{i\alpha}Y_\alpha$ and since Z_α, $\alpha = n+1, \dots, n+m$, must be linear combinations of $\sum_\alpha \phi_{g\alpha}Y_\alpha$, they must have means zero (under the null hypothesis). Thus a/σ^2 and b/σ^2 have the stated independent χ^2-distributions.

In the multivariate case the transformation $\mathbf{Y}_\beta = \sum_\alpha \psi_{\beta\alpha}\mathbf{Z}_\alpha$ is used, where $\mathbf{Y}_\beta$ and $\mathbf{Z}_\alpha$ are vectors. Then

$$\begin{aligned} \mathbf{A} &= \sum_{\alpha,\beta,\gamma,\delta} d_{\alpha\beta}\psi_{\alpha\gamma}\psi_{\beta\delta}\mathbf{Z}_\gamma\mathbf{Z}_\delta' = \sum_{\alpha=1}^{n}\mathbf{Z}_\alpha\mathbf{Z}_\alpha', \\ \mathbf{B} &= \sum_{\alpha,\beta,\gamma,\delta} c_{\alpha\beta}\psi_{\alpha\gamma}\psi_{\beta\delta}\mathbf{Z}_\gamma\mathbf{Z}_\delta' = \sum_{\alpha=n+1}^{n+m}\mathbf{Z}_\alpha\mathbf{Z}_\alpha' \end{aligned} \tag{29}$$

because it follows from (28) that $\sum_{\alpha,\beta} d_{\alpha\beta}\psi_{\alpha\gamma}\psi_{\beta\delta} = 1$, $\gamma = \delta \le n$, and $= 0$ otherwise, and $\sum_{\alpha,\beta} c_{\alpha\beta}\psi_{\alpha\gamma}\psi_{\beta\delta} = 1$, $n+1 \le \gamma = \delta \le n+m$, and $= 0$ otherwise. Since $\mathbf{\Psi}$ is orthogonal, the $\{\mathbf{Z}_\alpha\}$ are independently normally distributed with covariance matrix Σ. The same argument shows $\mathscr{E}\mathbf{Z}_\alpha = \mathbf{0}$, $\alpha = 1, \ldots, n+m$, under the null hypothesis. Thus $\boldsymbol{A}$ and $\boldsymbol{B}$ are independently distributed according to $W(\boldsymbol{\Sigma}, n)$ and $W(\boldsymbol{\Sigma}, m)$, respectively.

8.10. SOME OPTIMAL PROPERTIES OF TESTS

8.10.1. Admissibility of Invariant Tests

In this chapter we have considered several tests of a linear hypothesis which are invariant with respect to transformations that leave the null hypothesis invariant. We raise the question of which invariant tests are good tests. In particular we ask for admissible procedures, that is, procedures that cannot be improved on in the sense of smaller probabilities of Type I and/or Type II error. The competing tests are not necessarily invariant. Clearly, if an invariant test is admissible in the class of all tests, it is admissible in the class of invariant tests.

Testing the general linear hypothesis as treated here is a generalization of testing the hypothesis concerning one mean vector as treated in Chapter 5. The invariant procedures in Chapter 8 are generalizations of the T^2-test. One way of showing a procedure is admissible is to display a prior distribution on the parameters such that the Bayes procedure is a given test procedure. This approach requires some ingenuity in constructing the prior, but the verification of the property given the prior is straightforward. Problems 8.26 and 8.27 show that the Bartlett–Nanda–Pillai trace criterion V and Wilks's likelihood ratio criterion U yield admissible tests. The disadvantage of this approach to admissibility is that one must invent a prior distribution for each procedure; a general theorem does not cover many cases.

The other approach to admissibility is to apply Stein's theorem (Theorem 5.6.5), which yields general results. The invariant tests can be stated in terms of the roots of the determinantal equation

$$|\boldsymbol{H} - \lambda(\boldsymbol{H} + \boldsymbol{G})| = 0, \tag{1}$$

where $\boldsymbol{H} = \hat{\boldsymbol{\beta}}_1 \boldsymbol{A}_{11\cdot 2} \hat{\boldsymbol{\beta}}_1' = \boldsymbol{W}_1 \boldsymbol{W}_1'$ and $\boldsymbol{G} = N\hat{\boldsymbol{\Sigma}}_\Omega = \boldsymbol{W}_3 \boldsymbol{W}_3'$. There is also a matrix $\hat{\boldsymbol{\beta}}_2$ (or $\boldsymbol{W}_2$) associated with the nuisance parameters $\boldsymbol{\beta}_2$. For convenience, we define the canonical form in the following notation. Let $\boldsymbol{W}_1 = \boldsymbol{X}$ $(p \times m)$, $\boldsymbol{W}_2 = \boldsymbol{Y}$ $(p \times r)$, $\boldsymbol{W}_3 = \boldsymbol{Z}$ $(p \times n)$, $\mathscr{E}\boldsymbol{X} = \boldsymbol{\Xi}$, $\mathscr{E}\boldsymbol{Y} = \boldsymbol{H}$, and $\mathscr{E}\boldsymbol{Z} = \mathbf{0}$; the columns are independently normally distributed with covariance matrix $\boldsymbol{\Sigma}$. The null hypothesis is $\boldsymbol{\Xi} = \mathbf{0}$, and the alternative hypothesis is $\boldsymbol{\Xi} \neq \mathbf{0}$.

The usual tests are given in terms of the (nonzero) roots of

$$|\boldsymbol{XX}' - \lambda(\boldsymbol{ZZ}' + \boldsymbol{XX}')| = |\boldsymbol{XX}' - \lambda(\boldsymbol{U} - \boldsymbol{YY}')| = 0. \tag{2}$$

where $U = XX' + YY' + ZZ'$. Expect for roots that are identically zero, the roots of (2) coincide with the nonzero characteristic roots of $X'(U - YY')^{-1}X$. Let $V = (X, Y, U)$ and

$$M(V) = X'(U - YY')^{-1}X. \tag{3}$$

The vector of ordered characteristic roots of $M(V)$ is denoted by

$$(\lambda_1, \ldots, \lambda_m)' = \boldsymbol{\lambda}(M(V)), \tag{4}$$

where $\lambda_1 \geq \cdots \geq \lambda_m \geq 0$. Since the inclusion of zero roots (when $m > p$) causes no trouble in the sequel, we assume that the tests depend on $\boldsymbol{\lambda}(M(V))$.

The admissibility of these tests can be stated in terms of the geometric characteristics of the acceptance regions. Let

$$\begin{aligned} R^m_< &= \{\boldsymbol{\lambda} \in R^m \mid \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0\}, \\ R^m_+ &= \{\boldsymbol{\lambda} \in R^m \mid \lambda_1 \geq 0, \ldots, \lambda_m \geq 0\}. \end{aligned} \tag{5}$$

It seems reasonable that if a set of sample roots leads to acceptance of the null hypothesis, then a set of smaller roots would as well (Figure 8.2).

Definition 8.10.1. *A region $A \subset R^m_<$ is* monotone *if $\boldsymbol{\lambda} \in A$, $\boldsymbol{\nu} \in R^m_<$, and $\nu_i \leq \lambda_i$, $i = 1, \ldots, m$, imply $\boldsymbol{\nu} \in A$.*

Definition 8.10.2. *For $A \subset R^m_<$ the* extended region A^* *is*

$$A^* = \bigcup_\pi \{(x_{\pi(1)}, \ldots, x_{\pi(m)})' \mid x \in A\}, \tag{6}$$

where π ranges over all permutations of $(1, \ldots, m)$.

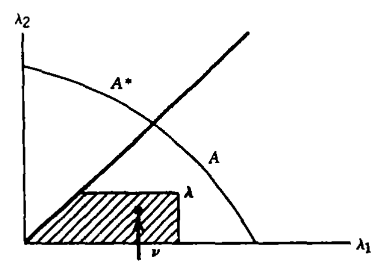

Figure 8.2. A monotone acceptance region.

The main result, first proved by Schwartz (1967), is the following theorem:

Theorem 8.10.1. *If the region $A \subset R^m_<$ is monotone and if the extended region A^* is closed and convex, then A is the acceptance region of an admissible test.*

Another characterization of admissible tests is given in terms of *majorization.*

Definition 8.10.3. *A vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)'$* weakly majorizes *a vector $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_m)'$ if*

$$(7) \quad \lambda_{[1]} \geq \nu_{[1]}, \; \lambda_{[1]} + \lambda_{[2]} \geq \nu_{[1]} + \nu_{[2]}, \ldots, \lambda_{[1]} + \cdots + \lambda_{[m]} \geq \nu_{[1]} + \cdots + \nu_{[m]},$$

where $\lambda_{[i]}$ and $\nu_{[i]}$, $i = 1, \ldots, m$, are the coordinates rearranged in nonascending order.

We use the notation $\boldsymbol{\lambda} \succ_w \boldsymbol{\nu}$ or $\boldsymbol{\nu} \prec_w \boldsymbol{\lambda}$ if $\boldsymbol{\lambda}$ weakly majorizes $\boldsymbol{\nu}$. If $\boldsymbol{\lambda}, \boldsymbol{\nu} \in \mathrm{R}^m_<$, then $\boldsymbol{\lambda} \succ_w \boldsymbol{\nu}$ is simply

$$(8) \quad \lambda_1 \geq \nu_1, \quad \lambda_1 + \lambda_2 \geq \nu_1 + \nu_2, \ldots, \quad \lambda_1 + \cdots + \lambda_m \geq \nu_1 + \cdots + \nu_m.$$

If the last inequality in (7) is replaced by an equality, we say simply that $\boldsymbol{\lambda}$ *majorizes* $\boldsymbol{\nu}$ and denote this by $\boldsymbol{\lambda} \succ \boldsymbol{\nu}$ or $\boldsymbol{\nu} \prec \boldsymbol{\lambda}$. The theory of majorization and the related inequalities are developed in detail in Marshall and Olkin (1979).

Definition 8.10.4. *A region $A \subset R^m_<$ is* monotone in majorization *if $\boldsymbol{\lambda} \in A$, $\boldsymbol{\nu} \in R^m_<$, $\boldsymbol{\nu} \prec_w \boldsymbol{\lambda}$ imply $\boldsymbol{\nu} \in A$.* (See Figure 8.3.)

Theorem 8.10.2. *If a region $A \subset R^m_<$ is closed, convex, and monotone in majorization, then A is the acceptance region of an admissible test.*

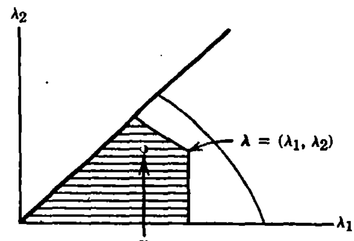

Figure 8.3. A region monotone in majorization.

Theorems 8.10.1 and 8.10.2 are equivalent; it will be convenient to prove Theorem 8.10.2 first. Then an argument about the extreme points of a certain convex set (Lemma 8.10.11) establishes the equivalence of the two theorems.

Theorem 5.6.5 (Stein's theorem) will be used because we can write the distribution of (X, Y, Z) in exponential form. Let $U = XX' + YY' + ZZ' = (u_{ij})$ and $\Sigma^{-1} = (\sigma^{ij})$. For a general matrix $C = (c_1, \dots, c_k)$, let $\operatorname{vec}(C) = (c_1', \dots, c_k')'$. The density of (X, Y, Z) can be written as

$$
(9)\quad \begin{aligned} f(X,Y,Z) &= K(\Xi, \mathrm{H}, \Sigma)\exp\{\operatorname{tr}\Xi'\Sigma^{-1}X + \operatorname{tr}\mathrm{H}'\Sigma^{-1}Y - \tfrac{1}{2}\operatorname{tr}\Sigma^{-1}U\} \\ &= K(\Xi, \mathrm{H}, \Sigma)\exp\{\omega_{(1)}'y_{(1)} + \omega_{(2)}'y_{(2)} + \omega_{(3)}'y_{(3)}\}, \end{aligned}
$$

where $K(X, \mathrm{H}, \Sigma)$ is a constant,

$$
(10)\quad \begin{aligned} &\omega_{(1)} = \operatorname{vec}(\Sigma^{-1}\Xi), \qquad \omega_{(2)} = \operatorname{vec}(\Sigma^{-1}\mathrm{H}), \\ &\omega_{(3)} = -\tfrac{1}{2}(\sigma^{11}, 2\sigma^{12}, \dots, 2\sigma^{1p}, \sigma^{22}, \dots, \sigma^{pp})', \\ &y_{(1)} = \operatorname{vec}(X), \qquad y_{(2)} = \operatorname{vec}(Y), \\ &y_{(3)} = (u_{11}, u_{12}, \dots, u_{1p}, u_{22}, \dots, u_{pp})'. \end{aligned}
$$

If we denote the mapping $(X, Y, Z) \to y = (y_{(1)}', y_{(2)}', y_{(3)}')'$ by g, $y = g(X, Y, Z)$, then the measure of a set A in the space of y is $m(A) = \mu(g^{-1}(A))$, where μ is the ordinary Lebesgue measure on $R^{p(m+r+n)}$. We note that (X, Y, U) is a sufficient statistic and so is $y = (y_{(1)}', y_{(2)}', y_{(3)}')'$. Because a test that is admissible with respect to the class of tests based on a sufficient statistic is admissible in the whole class of tests, we consider only tests based on a sufficient statistic. Then the acceptance regions of these tests are subsets in the space of y. The density of y given by the right-hand side of (9) is of the form of the exponential family, and therefore we can apply Stein's theorem. Furthermore, since the transformation $(X, Y, U) \to y$ is linear, we prove the convexity of an acceptance region of (X, Y, U). The acceptance region of an invariant test is given in terms of $\lambda(M(V)) = (\lambda_1, \dots, \lambda_m)'$. Therefore, in order to prove the admissibility of these tests we have to check that the inverse image of A, namely, $\tilde{A} = \{V \mid \lambda(M(V)) \in A\}$, satisfies the conditions of Stein's theorem, namely, is convex.

Suppose $V_i = (X_i, X_i, U_i) \in \tilde{A}$, $i = 1, 2$, that is, $\lambda[M(V_i)] \in A$. By the convexity of A, $p\lambda[M(V_1)] + q\lambda[M(V_2)] \in A$ for $0 \le p = 1 - q \le 1$. To show $pV_1 + qV_2 \in \tilde{A}$, that is, $\lambda[M(pV_1 + qV_2)] \in A$, we use the property of monotonicity of majorization of A and the following theorem.

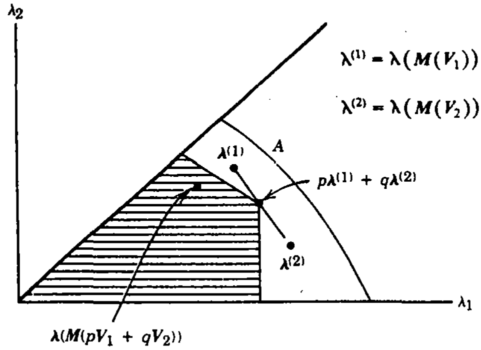

Figure 8.4. Theorem 8.10.3.

Theorem 8.10.3.

$$\lambda[M(pV_1 + gV_2)] \succ_w p\lambda[M(V_1)] + q\lambda[M(V_2)]. \tag{11}$$

The proof of Theorem 8.10.3 (Figure 8.4) follows from the pair of majorizations

$$\begin{aligned} \lambda[M(pV_1 + qV_2)] &\succ_w \lambda[pM(V_1) + qM(V_2)] \\ &\succ_w p\lambda[M(V_1)] + q\lambda[M(V_2)]. \end{aligned} \tag{12}$$

The second majorization in (12) is a special case of the following lemma.

Lemma 8.10.1. *For A and B symmetric,*

$$\lambda(A + B) \succ_w \lambda(A) + \lambda(B). \tag{13}$$

Proof. By Corollary A.4.2 of the Appendix,

$$\begin{aligned} \sum_{i=1}^{k} \lambda_i(A + B) &= \max_{R'R = I_k} \operatorname{tr} R'(A + B)R \\ &\le \max_{R'R = I_k} \operatorname{tr} R'AR + \max_{R'R = I_k} \operatorname{tr} R'BR \\ &= \sum_{i=1}^{k} \lambda_i(A) + \sum_{i=1}^{k} \lambda_i(B) \\ &= \sum_{i=1}^{k} \{\lambda_i(A) + \lambda_i(B)\}, \qquad k = 1, \ldots, p. \end{aligned} \tag{14}$$

■

Let $A > B$ mean $A - B$ is positive definite and $A \geq B$ mean $A - B$ is positive semidefinite.

The first majorization in (12) follows from several lemmas.

Lemma 8.10.2

$$pU_1 + qU_2 - (pY_1 + qY_2)(pY_1 + qY_2)' \geq p(U_1 - Y_1Y_1') + q(U_2 - Y_2Y_2'). \tag{15}$$

Proof. The left-hand side minus the right-hand side is

$$\begin{aligned} &pY_1Y_1' + qY_2Y_2' - p^2Y_1Y_1' - q^2Y_2Y_2' - pq(Y_1Y_2' + Y_2Y_1') \\ &= p(1-p)Y_1Y_1' + q(1-q)Y_2Y_2' - pq(Y_1Y_2' + Y_2Y_1') \\ &= pq(Y_1 - Y_2)(Y_1 - Y_2)' \geq 0. \quad \blacksquare \end{aligned} \tag{16}$$

Lemma 8.10.3. *If* $A \geq B > 0$, *then* $A^{-1} \leq B^{-1}$.

Proof. See Problem 8.31. ■

Lemma 8.10.4. *If* $A > 0$, *then* $f(x, A) = x'A^{-1}x$ *is convex in* (x, A).

Proof. See Problem 5.17. ■

Lemma 8.10.5. *If* $A_1 > 0$, $A_2 > 0$, *then*

$$(pB_1 + qB_2)'(pA_1 + qA_2)^{-1}(pB_1 + qB_2) \leq pB_1'A_1^{-1}B_1 + qB_2'A_2^{-1}B_2. \tag{17}$$

Proof. From Lemma 8.10.4 we have for all y

$$\begin{aligned} &py'B_1'A_1^{-1}B_1y + qy'B_2'A_2^{-1}B_2y \\ &\qquad - y'(pB_1 + qB_2)'(pA_1 + qA_2)^{-1}(pB_1 + qB_2)y \\ &= p(B_1y)'A_1^{-1}(B_1y) + q(B_2y)'A_2^{-1}(B_2y) \\ &\qquad - (pB_1y + qB_2y)'(pA_1 + qA_2)^{-1}(pB_1y + qB_2y) \\ &\geq 0. \quad \blacksquare \end{aligned} \tag{18}$$

Thus the matrix of the quadratic form in y is positive semidefinite. ■

The relation as in (17) is sometimes called *matrix convexity*. [See Marshall and Olkin (1979).]

Lemma 8.10.6.

$$M(pV_1 + qV_2) \leq pM(V_1) + qM(V_2), \tag{19}$$

where $V_1 = (X_1, Y_1, U_1)$, $V_2 = (X_2, Y_2, U_2)$, $U_1 - Y_1 Y_1' > 0$, $U_2 - Y_2 Y_2' > 0$, $0 \leq p = 1 - q \leq 1$.

Proof. Lemmas 8.10.2 and 8.10.3 show that

$$\begin{aligned}[pU_1 + qU_2 - (pY_1 + qY_2)(pY_1 + qY_2)']^{-1} \\ \leq [p(U_1 - Y_1Y_1') + q(U_2 - Y_2Y_2')]^{-1}.\end{aligned} \tag{20}$$

This implies

$$\begin{aligned}&M(pV_1 + qV_2) \\ &\quad \leq (pX_1 + qX_2)'[p(U_1 - Y_1Y_1') + q(U_2 - Y_2Y_2')]^{-1}(pX_1 + qX_2).\end{aligned} \tag{21}$$

Then Lemma 8.10.5 implies that the right-hand side of (21) is less than or equal to

$$pX_1'(U_1 - Y_1Y_1')^{-1}X_1 + qX_2'(U_2 - Y_2Y_2')^{-1}X_2 = pM(V_1) + qM(V_2). \tag{22}$$

■

Lemma 8.10.7. *If* $A \leq B$, *then* $\lambda(A) \prec_w \lambda(B)$.

Proof. From Corollary A.4.2 of the Appendix,

$$\sum_{i=1}^{k} \lambda_i(A) = \max_{R'R = I_k} \operatorname{tr} R'AR \leq \max_{R'R = I_k} \operatorname{tr} R'BR = \sum_{i=1}^{k} \lambda_i(B), \qquad k = 1, \ldots, p. \tag{23}$$

■

From Lemma 8.10.7 we obtain the first majorization in (12) and hence Theorem 8.10.3, which in turn implies the convexity of $\tilde{A}$. Thus the acceptance region satisfies condition (i) of Stein's theorem.

Lemma 8.10.8. *For the acceptance region A of Theorem* 8.10.1 *or Theorem* 8.10.2, *condition* (ii) *of Stein's theorem is satisfied.*

Proof. Let $\boldsymbol{\omega}$ correspond to $(\boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Theta})$; then

$$\begin{aligned}\boldsymbol{\omega}'y &= \boldsymbol{\omega}_{(1)}'y_{(1)} + \boldsymbol{\omega}_{(2)}'y_{(2)} + \boldsymbol{\omega}_{(3)}'y_{(3)} \\ &= \operatorname{tr} \boldsymbol{\Phi}'X + \operatorname{tr} \boldsymbol{\Psi}'Y - \tfrac{1}{2}\operatorname{tr} \boldsymbol{\Theta}U,\end{aligned} \tag{24}$$

where $\mathbf{\Theta}$ is symmetric. Suppose that $\{y \mid \omega' y > c\}$ is disjoint from $\tilde{A} = \{V \mid \lambda(M(V)) \in A\}$. We want to show that in this case $\mathbf{\Theta}$ is positive semidefinite. If this were not true, then

$$\mathbf{\Theta} = D \begin{bmatrix} I & 0 & 0 \\ 0 & -I & 0 \\ 0 & 0 & 0 \end{bmatrix} D', \tag{25}$$

where D is nonsingular and $-I$ is not vacuous. Let $X = (1/\gamma)X_0$, $Y = (1/\gamma)Y_0$,

$$U = (D')^{-1} \begin{pmatrix} I & 0 & 0 \\ 0 & \gamma I & 0 \\ 0 & 0 & I \end{pmatrix} D^{-1}, \tag{26}$$

and $V = (X, Y, U)$, where X_0, Y_0 are fixed matrices and γ is a positive number. Then

$$\omega' y = \frac{1}{\gamma} \operatorname{tr} \mathbf{\Phi}' X_0 + \frac{1}{\gamma} \operatorname{tr} \mathbf{\Psi}' Y_0 + \frac{1}{2} \operatorname{tr} \begin{pmatrix} -I & 0 & 0 \\ 0 & \gamma I & 0 \\ 0 & 0 & 0 \end{pmatrix} > c \tag{27}$$

for sufficiently large γ. On the other hand,

$$\begin{aligned} \lambda(M(V)) &= \lambda\{X'(U - YY')^{-1} X\} \\ &= \frac{1}{\gamma^2} \lambda \left\{ X_0' \left[(D')^{-1} \begin{pmatrix} I & 0 & 0 \\ 0 & \gamma I & 0 \\ 0 & 0 & I \end{pmatrix} D^{-1} - \frac{1}{\gamma^2} Y_0 Y_0' \right]^{-1} X_0 \right\} \\ &\to 0 \end{aligned} \tag{28}$$

as $\gamma \to \infty$. Therefore, $V \in \tilde{A}$ for sufficiently large γ. This is a contradiction. Hence $\mathbf{\Theta}$ is positive semidefinite.

Now let ω_1 correspond to $(\mathbf{\Phi}_1, \mathbf{0}, I)$, where $\mathbf{\Phi}_1 \neq \mathbf{0}$. Then $I + \lambda\mathbf{\Theta}$ is positive definite and $\mathbf{\Phi}_1 + \lambda\mathbf{\Phi} \neq \mathbf{0}$ for sufficiently large λ. Hence $\omega_1 + \lambda\omega \in \Omega - \Omega_0$ for sufficiently large λ. ■

The preceding proof was suggested by Charles Stein.

By Theorem 5.6.5, Theorem 8.10.3 and Lemma 8.10.8 now imply Theorem 8.10.2.

To obtain Theorem 8.10.1 from Theorem 8.10.2, we use the following lemmas.

Lemma 8.10.9. *$A \subset R^m_{\leq}$ is convex and monotone in majorization if and only if A is monotone and A^* is convex.*

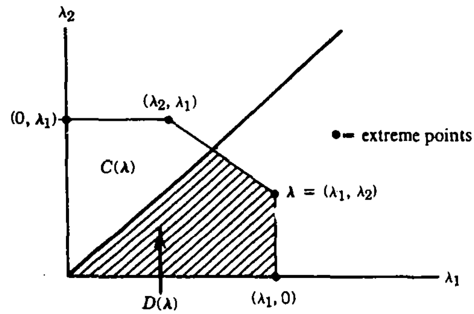

Figure 8.5

Proof. Necessity. If A is monotone in majorization, then it is obviously monotone. A^* is convex (see Problem 8.35).

Sufficiency. For $\lambda \in R^m_<$ let

$$(29)\qquad \begin{aligned} C(\lambda) &= \{x | x \in R^m_+, x \succ_w \lambda\},\\ D(\lambda) &= \{x | x \in R^m_<, x \succ_w \lambda\}. \end{aligned}$$

It will be proved in Lemma 8.10.10, Lemma 8.10.11, and its corollary that monotonicity of A and convexity of A^* implies $C(\lambda) \subset A^*$. Then $D(\lambda) = C(\lambda) \cap R^m_< \subset A^* \cap R^m_< = A$. Now suppose $\nu \in R^m_<$ and $\nu \prec_w \lambda$. Then $\nu \in D(\lambda) \subset A$. This shows that A is monotone in majorization. Furthermore, if A^* is convex, then $A = R^m_< \cap A^*$ is convex. (See Figure 8.5.) ■

Lemma 8.10.10. *Let C be compact and convex, and let D be convex. If the extreme points of C are contained in D, then $C \subset D$.*

Proof. Obvious. ■

Lemma 8.10.11. *Every extreme point of $C(\lambda)$ is of the form*

$$(30)\qquad \left(\delta_{\pi(1)}\lambda_{\pi(1)}, \ldots, \delta_{\pi(m)}\lambda_{\pi(m)}\right),$$

where π is a permutation of $(1, \ldots, m)$ and $\delta_1 = \cdots = \delta_k = 1$, $\delta_{k+1} = \cdots = \delta_m = 0$ for some k.

Proof. $C(\lambda)$ is convex. (See Problem 8.34.) Now note that $C(\lambda)$ is permutation-symmetric, that is, if $(x_1, \ldots, x_m)' \in C(\lambda)$, then $(x_{\pi(1)}, \ldots, x_{\pi(m)})' \in C(\lambda)$ for any permutation π. Therefore, for any permutation π, $\pi(C(\lambda)) =$

$\{(x_{\pi(1)}, \ldots, x_{\pi(m)})' | x \in C(\lambda)\}$ coincides with $C(\lambda)$. This implies that if $(x_1, \ldots, x_m)'$ is an extreme point of $C(\lambda)$, then $(x_{\pi(1)}, \ldots, x_{\pi(m)})'$ is also an extreme point. In particular, $(x_{[1]}, \ldots, x_{[m]}) \in R^m_<$ is an extreme point. Conversely, if $(x_1, \ldots, x_m) \in R^m_<$ is an extreme point of $C(\lambda)$, then $(x_{\pi(1)}, \ldots, x_{\pi(m)})'$ is an extreme point.

We see that once we enumerate the extreme points of $C(\lambda)$ in $R^m_<$, the rest of the extreme points can be obtained by permutation.

Suppose $x \in R^m_<$. An extreme point, being the intersection of m hyperplanes, has to satisfy m or more of the following $2m$ equations:

$$
(31) \qquad \begin{array}{ll}
E_1 : x_1 = 0, & F_1 : x_1 = \lambda_1, \\
E_2 : x_2 = 0, & F_2 : x_1 + x_2 = \lambda_1 + \lambda_2, \\
\vdots & \vdots \\
E_m : x_m = 0, & F_m : x_1 + \cdots + x_m = \lambda_1 + \cdots + \lambda_m.
\end{array}
$$

Suppose that k is the first index such that E_k holds. Then $x \in R^m_<$ implies $0 = x_k \geq x_{k+1} \geq \cdots \geq x_m \geq 0$. Therefore, $E_k, \ldots, E_m$ hold. The remaining $k - 1 = m - (m - k + 1)$ or more equations are among the F's. We order them as $F_{i_1}, \ldots, F_{i_l}$, where $i_1 < \cdots < i_l$, $l \geq k - 1$. Now $i_1 < \cdots < i_l$ implies $i_l \geq l$ with equality if and only if $i_1 = 1, \ldots, i_l = l$. In this case $F_1, \ldots, F_{k-1}$ hold ($l \geq k - 1$). Now suppose $i_l > l$. Since $x_k = \cdots = x_m = 0$,

$$
(32) \qquad F_{i_l} : x_1 + \cdots + x_{k-1} = \lambda_1 + \cdots + \lambda_{k-1} + \cdots + \lambda_{i_l}.
$$

But $x_1 + \cdots + x_{k-1} \leq \lambda_1 + \cdots + \lambda_{k-1}$, and we have $\lambda_k + \cdots + \lambda_{i_l} = 0$. Therefore, $0 = \lambda_k + \cdots + \lambda_{i_l} \geq \lambda_k \geq \cdots \geq \lambda_m \geq 0$. In this case $F_{k-1}, \ldots, F_m$ reduce to the same equation $x_1 + \cdots + x_{k-1} = \lambda_1 + \cdots + \lambda_{k-1}$. It follows that x satisfies $k - 2$ more equations, which have to be $F_1, \ldots, F_{k-2}$. We have shown that in either case $E_k, \ldots, E_m, F_1, \ldots, F_{k-1}$ hold and this gives the point $\beta = (\lambda_1, \ldots, \lambda_{k-1}, 0, \ldots, 0)$, which is in $R^m_< \cap C(\lambda)$. Therefore, β is an extreme point. ■

Corollary 8.10.1. $C(\lambda) \subset A^*$.

Proof. If A is monotone, then A^* is monotone in the sense that if $\lambda = (\lambda_1, \ldots, \lambda_m)' \in A^*$, $\nu = (\nu_1, \ldots, \nu_m)'$, $\nu_i \leq \lambda_i$, $i = 1, \ldots, m$, then $\nu \in A^*$. (See Problem 8.35.) Now the extreme points of $C(\lambda)$ given by (30) are in A^* because of permutation symmetry and monotonicity of A^*. Hence, by Lemma 8.10.10, $C(\lambda) \subset A^*$. ■

Proof of Theorem 8.10.1. Immediate from Theorem 8.10.2 and Lemma 8.10.9. ∎

Application of the theory of Schur-convex functions yields several corollaries to Theorem 8.10.2

Corollary 8.10.2. *Let g be continuous, nondecreasing, and convex in* $[0,1)$. *Let*

$$f(\boldsymbol{\lambda}) = f(\lambda_1, \dots, \lambda_m) = \sum_{i=1}^{m} g(\lambda_i). \tag{33}$$

Then a test with the acceptance region $A = \{\boldsymbol{\lambda} | f(\boldsymbol{\lambda}) \le c\}$ *is admissible.*

Proof. Being a sum of convex functions f is convex, and hence A is convex. A is closed because f is continuous. We want to show that if $f(\boldsymbol{x}) \le c$ and $\boldsymbol{y} \prec_w \boldsymbol{x}$ $(\boldsymbol{x}, \boldsymbol{y} \in R^m_{<})$, then $f(\boldsymbol{y}) \le c$. Let $\tilde{x}_k = \sum_{i=1}^{k} x_i$, $\tilde{y}_k = \sum_{i=1}^{k} y_i$. Then $\boldsymbol{y} \prec_w \boldsymbol{x}$ if and only if $\tilde{x}_k \ge \tilde{y}_k$, $k = 1, \dots, m$. Let $f(\boldsymbol{x}) = h(\tilde{x}_1, \dots, \tilde{x}_m) = g(\tilde{x}_1) + \sum_{i=2}^{m} g(\tilde{x}_i - \tilde{x}_{i-1})$. It suffices to show that $h(\tilde{x}_1, \dots, \tilde{x}_m)$ is increasing in each $\tilde{x}_i$. For $i \le m-1$ the convexity of g implies that

$$\begin{aligned} h(\tilde{x}_1, \dots, \tilde{x}_i + \varepsilon, \dots, \tilde{x}_m) - h(\tilde{x}_1, \dots, \tilde{x}_i, \dots, \tilde{x}_m) \\ = g(x_1 + \varepsilon) - g(x_i) - \{g(x_{i+1}) - g(x_{i+1} - \varepsilon)\} \ge 0. \end{aligned} \tag{34}$$

For $i = m$ the monotonicity of g implies

$$h(\tilde{x}_1, \dots, \tilde{x}_m + \varepsilon) - h(\tilde{x}_1, \dots, \tilde{x}_m) = g(x_m + \varepsilon) - g(x_m) \ge 0. \tag{35}$$ ∎

Setting $g(\lambda) = -\log(1-\lambda)$, $g(\lambda) = \lambda/(1-\lambda)$, $g(\lambda) = \lambda$, respectively, shows that Wilks' likelihood ratio test, the Lawley–Hotelling trace test, and the Bartlett–Nanda–Pillai test are admissible. Admissibility of Roy's maximum root test $A: \lambda_1 \le c$ follows directly from Theorem 8.10.1 or Theorem 8.10.2. On the contrary, the minimum root test, $\lambda_t \le c$, where $t = \min(m, p)$, does not satisfy the convexity condition. The following theorem shows that this test is actually inadmissible.

Theorem 8.10.4. *A necessary condition for an invariant test to be admissible is that the extended region in the space of* $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_t}$ *is convex and monotone.*

We shall only sketch the proof of this theorem [following Schwartz (1967)]. Let $\sqrt{\lambda_i} = d_i$, $i = 1, \dots, t$, and let the density of $d_1, \dots, d_t$ be $f(\boldsymbol{d} | \boldsymbol{\nu})$, where $\boldsymbol{\nu} = (\nu_1, \dots, \nu_t)'$ is defined in Section 8.6.5 and $f(\boldsymbol{d} | \boldsymbol{\nu})$ is given in Chapter 13.

The ratio $f(\boldsymbol{d}|\boldsymbol{\nu})/f(\boldsymbol{d}|\boldsymbol{0})$ can be extended symmetrically to the unit cube $(0 \le d_i \le 1,\ i = 1,\dots,t)$. The extended ratio is then a convex function and is strictly increasing in each d_i. A proper Bayes procedure has an acceptance region

$$\int \frac{f(\boldsymbol{d}|\boldsymbol{\nu})}{(\boldsymbol{d}|\boldsymbol{0})}\, d\Pi(\boldsymbol{\nu}) \le c, \tag{36}$$

where $\Pi(\boldsymbol{\nu})$ is a finite measure on the space of $\boldsymbol{\nu}$'s. Then the symmetric extension of the set of $\boldsymbol{d}$ satisfying (36) is convex and monotone [as shown by Birnbaum (1955)]. The closure (in the weak$*$topology) of the set of Bayes procedures forms an essentially complete class [Wald (1950)]. In this case the limit of the convex monotone acceptance regions is convex and monotone. The exposition of admissibility here was developed by Anderson and Takemura (1982).

8.10.2. Unbiasedness of Tests and Monotonicity of Power Functions

A test T is called *unbiased* if the power achieves its minimum at the null hypothesis. When there is a natural parametrization and a notion of distance in the parameter space, the power function is *monotone* if the power increases as the distance between the alternative hypothesis and the null hypothesis increases. Note that monotonicity implies unbiasedness. In this section we shall show that the power functions of many of the invariant tests of the general linear hypothesis are monotone in the invariants of the parameters, namely, the roots; these can be considered as measures of distance.

To introduce the approach, we consider the acceptance interval $(-a, a)$ for testing the null hypothesis $\mu = 0$ against the alternative $\mu \ne 0$ on the basis of an observation from $N(\mu, \sigma^2)$. In Figure 8.6 the probabilities of acceptance are represented by the shaded regions for three values of μ. It is clear that the probability of acceptance decreases monotonically (or equivalently the power increases monotonically) as μ moves away from zero. In fact, this property depends only on the density function being unimodal and symmetric.

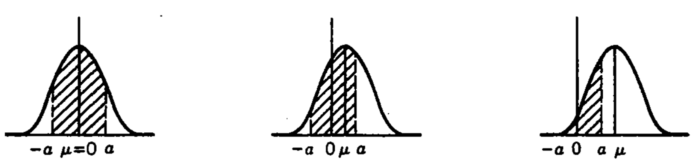

Figure 8.6. Three probabilities of acceptance.

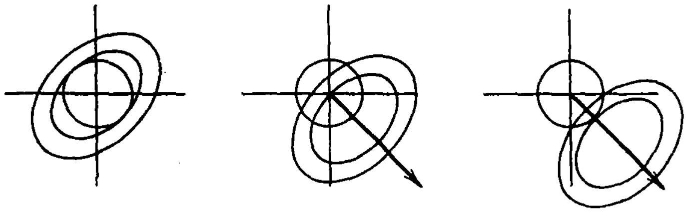

Figure 8.7. Acceptance regions.

In higher dimensions we generalize the interval by a symmetric convex set, and we ask that the density function be symmetric and unimodal in the sense that every contour of constant density surrounds a convex set. In Figure 8.7 we illustrate that in this case the probability of acceptance decreases monotonically. The following theorem is due to Anderson (1955b).

Theorem 8.10.5. *Let E be a convex set in n-space, symmetric about the origin. Let $f(\boldsymbol{x}) \geq 0$ be a function such that* (i) $f(\boldsymbol{x}) - f(-\boldsymbol{x})$, (ii) $\{\boldsymbol{x} | f(\boldsymbol{x}) \geq u\} = K_u$ *is convex for every u* $(0 < u < \infty)$, *and* (iii) $\int_E f(\boldsymbol{x})\, d\boldsymbol{x} < \infty$. *Then*

$$\int_E f(\boldsymbol{x} + k\boldsymbol{y})\, d\boldsymbol{x} \geq \int_E f(\boldsymbol{x} + \boldsymbol{y})\, d\boldsymbol{x} \tag{37}$$

for $0 \leq k \leq 1$.

The proof of Theorem 8.10.5 is based on the following lemma.

Lemma 8.10.12. *Let E, F be convex and symmetric about the origin. Then*

$$V\{(E + k\boldsymbol{y}) \cap F\} \geq V\{(E + \boldsymbol{y}) \cap F\}, \tag{38}$$

where $0 \leq k \leq 1$ and V denotes the n-dimensional volume.

Proof. Consider the set $\alpha(E + \boldsymbol{y}) + (1 - \alpha)(E - \boldsymbol{y}) = \alpha E + (1 - \alpha)E + (2\alpha - 1)\boldsymbol{y}$ which consists of points $\alpha(\boldsymbol{x} + \boldsymbol{y}) + (1 - \alpha)(\boldsymbol{z} - \boldsymbol{y})$ with $\boldsymbol{x}, \boldsymbol{z} \in E$. Let $\alpha_0 = (k + 1)/2$, so that $2\alpha_0 - 1 = k$. Then by convexity of E we have

$$\alpha_0(E + \boldsymbol{y}) + (1 - \alpha_0)(E - \boldsymbol{y}) \subset E + k\boldsymbol{y}. \tag{39}$$

Hence by convexity of F

$$\alpha_0[(E + \boldsymbol{y}) \cap F] + (1 - \alpha_0)[(E - \boldsymbol{y}) \cap F] \subset (E + k\boldsymbol{y}) \cap F$$

and

$$V\{\alpha_0[(E+y)\cap F]+(1-\alpha_0)[(E-y)\cap F]\}\le V\{(E+ky)\cap F\}. \tag{40}$$

Now by the Brunn–Minkowski inequality [e.g., Bonnesen and Fenchel (1948), Section 48], we have

$$\begin{aligned}V^{1/n}&\{\alpha_0[(E+y)\cap F]+(1-\alpha_0)[(E-y)\cap F]\}\\ &\ge \alpha_0 V^{1/n}\{(E+y)\cap F\}+(1-\alpha_0)V^{1/n}\{(E-y)\cap F\}\\ &=\alpha_0 V^{1/n}\{(E+y)\cap F\}+(1-\alpha_0)V^{1/n}\{(-E+y)\cap(-F)\}\\ &=V^{1/n}\{(E+y)\cap F\}.\end{aligned} \tag{41}$$

The last equality follows from the symmetry of E and F. ■

Proof of Theorem 8.10.5. Let

$$H(u)=V\{(E+ky)\cap K_u\}, \tag{42}$$

$$H^*(u)=V\{(E+y)\cap K_u\}. \tag{43}$$

Then

$$\begin{aligned}\int_E f(x+y)\,dx&=\int_{E+y} f(x)\,dx\\ &=\int_{E+y}\int_0^\infty I_{\{0\le u\le f(x)\}}(u)\,du\,dx\\ &=\int_0^\infty\int_{E+y} I_{\{0\le u\le f(x)\}}(u)\,dx\,du\\ &=\int_0^\infty H^*(u)\,du.\end{aligned} \tag{44}$$

Similarly,

$$\int_E f(x+ky)\,dx=\int_0^\infty H(u)\,du. \tag{45}$$

By Lemma 8.10.12, $H(u)\ge H^*(u)$. Hence Theorem 8.10.5 follows from (44) and (45). ■

We start with the canonical form given in Section 8.10.1. We further simplify the problem as follows. Let $t = \min(m, p)$, and let $\nu_1, \dots, \nu_t$ $(\nu_1 \geq \nu_2 \geq \cdots \geq \nu_t)$ be the nonzero characteristic roots of $\boldsymbol{\Xi}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Xi}$, where $\boldsymbol{\Xi} = \mathscr{E}\boldsymbol{X}$.

Lemma 8.10.13. *There exist matrices* $\boldsymbol{B}$ $(p \times p)$ *and* $\boldsymbol{F}$ $(m \times m)$ *such that*

$$
\begin{aligned}
&\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}' = \boldsymbol{I}_p, \qquad \boldsymbol{F}\boldsymbol{F}' = \boldsymbol{I}_m, \\
&\boldsymbol{B}\boldsymbol{\Xi}\boldsymbol{F}' = \left(\boldsymbol{D}_\nu^{\frac{1}{2}}, \boldsymbol{0}\right), \qquad p \leq m, \\
&\qquad\quad = \begin{pmatrix} \boldsymbol{D}_\nu^{\frac{1}{2}} \\ \boldsymbol{0} \end{pmatrix}, \qquad p > m,
\end{aligned} \tag{46}
$$

where $\boldsymbol{D}_\nu = \mathrm{diag}(\nu_1, \dots, \nu_t)$.

Proof. We prove this for the case $p \leq m$ and $\nu_p > 0$. Other cases can be proved similarly. By Theorem A.2.2 of the Appendix there is a matrix $\boldsymbol{B}$ such that

$$\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}' = \boldsymbol{I}, \qquad \boldsymbol{B}\boldsymbol{\Xi}\boldsymbol{\Xi}'\boldsymbol{B}' = \boldsymbol{D}_\nu. \tag{47}$$

Let

$$\boldsymbol{F}_1 = \boldsymbol{D}_\nu^{-\frac{1}{2}}\boldsymbol{B}\boldsymbol{\Xi} \quad (p \times m). \tag{48}$$

Then

$$\boldsymbol{F}_1\boldsymbol{F}_1' = \boldsymbol{I}_p. \tag{49}$$

Let $\boldsymbol{F}' = (\boldsymbol{F}_1', \boldsymbol{F}_2')$ be a full $m \times m$ orthogonal matrix. Then

$$\boldsymbol{B}\boldsymbol{\Xi}\boldsymbol{F}_2' = \boldsymbol{D}_\nu^{\frac{1}{2}}\boldsymbol{F}_1\boldsymbol{F}_2' = \boldsymbol{0} \tag{50}$$

and

$$\boldsymbol{B}\boldsymbol{\Xi}\boldsymbol{F}' = \boldsymbol{B}\boldsymbol{\Xi}(\boldsymbol{F}_1', \boldsymbol{F}_2') = \boldsymbol{B}\boldsymbol{\Xi}\left(\boldsymbol{\Xi}'\boldsymbol{B}'\boldsymbol{D}_\nu^{-\frac{1}{2}}, \boldsymbol{F}_2'\right) = \left(\boldsymbol{D}_\nu^{\frac{1}{2}}, \boldsymbol{0}\right). \tag{51}$$ ■

Now let

$$\boldsymbol{U} = \boldsymbol{B}\boldsymbol{X}\boldsymbol{F}', \qquad \boldsymbol{V} = \boldsymbol{B}\boldsymbol{Z}. \tag{52}$$

Then the columns of $\boldsymbol{U}, \boldsymbol{V}$ are independently normally distributed with covariance matrix $\boldsymbol{I}$ and means when $p \leq m$

$$
\begin{aligned}
\mathscr{E}\boldsymbol{U} &= \left(\boldsymbol{D}_\nu^{\frac{1}{2}}, \boldsymbol{0}\right), \\
\mathscr{E}\boldsymbol{V} &= \boldsymbol{0}.
\end{aligned} \tag{53}
$$

Invariant tests are given in terms of characteristic roots $l_1, \dots, l_t$ $(l_1 \geq \cdots \geq l_t)$ of $U'(VV')^{-1}U$. Note that for the admissibility we used the characteristic roots of λ_i of $U'(UU' + VV')^{-1}U$ rather than $l_i = \lambda_i/(1 - \lambda_i)$. Here it is more natural to use l_i, which corresponds to the parameter value ν_i. The following theorem is given by Das Gupta, Anderson, and Mudholkar (1964).

Theorem 8.10.6. *If the acceptance region of an invariant test is convex in the space of each column vector of U for each set of fixed values of V and of the other column vectors of U, then the power of the test increases monotonically in each ν_i.*

Proof. Since UU' is unchanged when any column vector of U is multiplied by -1, the acceptance region is symmetric about the origin in each of the column vectors of U. Now the density of $U = (u_{ij})$, $V = (v_{ij})$ is

$$
(54) \quad f(U,V) = (2\pi)^{-\frac{1}{2}(n+m)p} \exp\left[-\frac{1}{2}\left\{ \operatorname{tr} VV' + \sum_{i=1}^{t} \left(u_{ii} - \sqrt{\nu_i}\right)^2 + \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \neq i}}^{m} u_{ij}^2 \right\} \right].
$$

Applying Theorem 8.10.5 to (54), we see that the power increases monotonically in each $\sqrt{\nu_i}$. ∎

Since the section of a convex set is convex, we have the following corollary.

Corollary 8.10.3. *If the acceptance region A of an invariant test is convex in U for each fixed V, then the power of the test increases monotonically in each ν_i.*

From this we see that Roy's maximum root test $A: l_1 \leq K$ and the Lawley–Hotelling trace test $A: \operatorname{tr} U'(VV')^{-1}U \leq K$ have power functions that are monotonically increasing in each ν_i.

To see that the acceptance region of the likelihood ratio test

$$
(55) \qquad A: \prod_{i=1}^{t} (1 + l_i) \leq K
$$

satisfies the condition of Theorem 8.10.6 let

$$
(56) \qquad \begin{aligned} (VV')^{-1} &= T'T, \qquad T: p \times p \\ U^* &= (u_1^*, \dots, u_m^*) = TU. \end{aligned}
$$

Then

$$\prod_{i=1}^{t}(1+l_i) = |U'(VV')^{-1}U + I| = |U^{*\prime}U^* + I| \tag{57}$$

$$= |U^*U^{*\prime} + I| = |u_1^*u_1^{*\prime} + B|$$

$$= (u_1^{*\prime}B^{-1}u_1^* + 1)|B|$$

$$= (u_1'T'B^{-1}Tu_1 + 1)|B|,$$

where $B = u_2^*u_2^{*\prime} + \cdots + u_m^*u_m^{*\prime} + I$. Since $T'B^{-1}T$ is positive definite, (55) is convex in u_1. Therefore, the likelihood ratio test has a power function which is monotone increasing in each ν_i.

The Bartlett–Nanda–Pillai trace test

$$A: \operatorname{tr} U'(UU' + VV')^{-1}U = \sum_{i=1}^{t} \frac{l_i}{1+l_i} \le K \tag{58}$$

has an acceptance region that is an ellipsoid if $K < 1$ and is convex in each column u_i of U *provided* $K \le 1$. (See Problem 8.36.) For $K > 1$ (58) may not be convex in each column of U. The reader can work out an example for $p = 2$.

Eaton and Perlman (1974) have shown that if an invariant test is convex in U and $W = VV'$, then the power at $(\nu_1^0, \dots, \nu_t^0)$ is greater than at $(\nu_1, \dots, \nu_t)$ if $(\sqrt{\nu_1}, \dots, \sqrt{\nu_t}) \prec_w (\sqrt{\nu_1^0}, \dots, \sqrt{\nu_t^0})$. We shall not prove this result. Roy's maximum root test and the Lawley–Hotelling trace test satisfy the condition, but the likelihood ratio and the Bartlett–Nanda–Pillai trace test do not.

Takemura has shown that if the acceptance region is convex in U and W, the set of $\sqrt{\nu_1}, \dots, \sqrt{\nu_t}$ for which the power is not greater than a constant is monotone and convex.

It is enlightening to consider the contours of the power function, $\Pi(\sqrt{\nu_1}, \dots, \sqrt{\nu_t})$. Theorem 8.10.6 does not exclude case (a) of Figure 8.8.

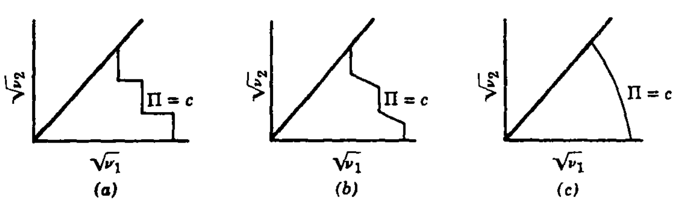

Figure 8.8. Contours of power functions.

and similarly the Eaton–Perlman result does not exclude (b). The last result guarantees that the contour looks like (c) for Roy's maximum root test and the Lawley–Hotelling trace test. These results relate to the fact that these two tests are more likely to detect alternative hypotheses where few $\boldsymbol{\nu}_i$'s are far from zero. In contrast with this, the likelihood ratio test and the Bartlett–Nanda–Pillai trace test are sensitive to the overall departure from the null hypothesis. It might be noted that the convexity in $\sqrt{\nu}$-space cannot be translated into the convexity in ν-space.

By using the noncentral density of l_i's which depends on the parameter values $\nu_1, \dots, \nu_t$, Perlman and Olkin (1980) showed that any invariant test with monotone acceptance region (in the space of roots) is unbiased. Note that this result covers all the standard tests considered earlier.

8.11. ELLIPTICALLY CONTOURED DISTRIBUTIONS

8.11.1. Observations Elliptically Contoured

The regression model of Section 8.2 can be written

$$x_\alpha = \boldsymbol{\beta} z_\alpha + e_\alpha, \qquad \alpha = 1, \dots, N, \tag{1}$$

where e_α is an unobserved disturbance with $\mathscr{E} e_\alpha = \mathbf{0}$ and $\mathscr{E} e_\alpha e_\alpha' = \boldsymbol{\Sigma}$. We assume that e_α has a density $|\boldsymbol{\Lambda}|^{-\frac{1}{2}} g(e'\boldsymbol{\Lambda}^{-1}e)$; then $\boldsymbol{\Sigma} = (\mathscr{E}R^2/p)\boldsymbol{\Lambda}$, where $R^2 = e_\alpha' \boldsymbol{\Lambda}^{-1} e_\alpha$. In general the exact distribution of $\boldsymbol{B} = \sum_{\alpha=1}^N \boldsymbol{x}_\alpha z_\alpha' \boldsymbol{A}^{-1}$ and $N\hat{\boldsymbol{\Sigma}} = \sum_{\alpha=1}^N (\boldsymbol{x}_\alpha - \boldsymbol{B}z_\alpha)(\boldsymbol{x}_\alpha - \boldsymbol{B}z_\alpha)'$ is difficult to obtain and cannot be expressed concisely. However, the expected value of $\boldsymbol{B}$ is $\boldsymbol{\beta}$, and the covariance matrix of $\operatorname{vec} \boldsymbol{B}$ is $\boldsymbol{\Sigma} \otimes \boldsymbol{A}^{-1}$ with $\boldsymbol{A} = \sum_{\alpha=1}^N z_\alpha z_\alpha'$. We can develop a large-sample distribution for $\boldsymbol{B}$ and $N\hat{\boldsymbol{\Sigma}}$.

Theorem 8.11.1. *Suppose $(1/N)A \to A_0$, $z_\alpha' z_\alpha <$ constant, $\alpha = 1, 2, \dots$, and either the e_α's are independent identically distributed or the e_α's are independent with $\mathscr{E}|e_\alpha' e_\alpha|^{2+\varepsilon} <$ constant for some $\varepsilon > 0$. Then $\boldsymbol{B} \xrightarrow{p} \boldsymbol{\beta}$ and $\sqrt{N}\operatorname{vec}(\boldsymbol{B} - \boldsymbol{\beta})$ has a limiting normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma} \otimes A_0^{-1}$.*

Theorem 8.11.1 appears in Anderson (1971) as Theorem 5.5.13. There are many alternatives to its assumptions in the literature. Under its assumptions $\hat{\boldsymbol{\Sigma}}_\Omega \xrightarrow{p} \boldsymbol{\Sigma}$. This result permits a large-sample theory for the criteria for testing null hypotheses about $\boldsymbol{\beta}$.

Consider testing the null hypothesis

$$H: \boldsymbol{\beta} = \boldsymbol{\beta}^*, \tag{2}$$

where $\boldsymbol{\beta}^*$ is completely specified. In Section 8.3 a more general hypothesis was considered for $\boldsymbol{\beta}$ partitioned as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$. However, as shown in that section by the transformation (4), the hypothesis $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^*$ can be reduced to a hypothesis of the form (1) above.

Let

$$G = \sum_{\alpha=1}^{N} (\boldsymbol{x}_\alpha - \boldsymbol{B}\boldsymbol{z}_\alpha)(\boldsymbol{x}_\alpha - \boldsymbol{B}\boldsymbol{z}_\alpha)' = N\hat{\boldsymbol{\Sigma}}_\Omega, \tag{3}$$

$$\boldsymbol{H} = (\boldsymbol{B} - \boldsymbol{\beta})\boldsymbol{A}(\boldsymbol{B} - \boldsymbol{\beta})'. \tag{4}$$

Lemma 8.11.1. *Under the conditions of Theorem* 8.11.1 *the limiting distribution of* $\boldsymbol{H}$ *is* $W(\boldsymbol{\Sigma}, q)$.

Proof. Write $\boldsymbol{H}$ as

$$\boldsymbol{H} = \sqrt{N}(\boldsymbol{B} - \boldsymbol{\beta})\frac{1}{N}\boldsymbol{A}\sqrt{N}(\boldsymbol{B} - \boldsymbol{\beta})'. \tag{5}$$

Then the lemma follows from Theorem 8.11.1 and (4) of Section 8.4. ■

We can express the likelihood ratio criterion in the form

$$\begin{aligned} -2\log\lambda &= -N\log U = N\log|\boldsymbol{I} + \boldsymbol{G}^{-1}\boldsymbol{H}| \\ &= N\log\left|\boldsymbol{I} + \frac{1}{N}\left(\frac{1}{N}\boldsymbol{G}\right)^{-1}\boldsymbol{H}\right|. \end{aligned} \tag{6}$$

Theorem 8.11.2. *Under the conditions of Theorem* 8.11.1, *when the null hypothesis is true,*

$$-2\log\lambda \xrightarrow{d} \chi^2_{pq}. \tag{7}$$

Proof. We use the fact that $N\log|\boldsymbol{I} + N^{-1}\boldsymbol{C}| = \operatorname{tr}\boldsymbol{C} + O_p(N^{-1})$ when $N \to \infty$, since $|\boldsymbol{I} + x\boldsymbol{C}| = 1 + x\operatorname{tr}\boldsymbol{C} + O(x^2)$ (Theorem A.4.8).

We have

$$\begin{aligned} \operatorname{tr}\left(\frac{1}{N}\boldsymbol{G}\right)^{-1}\boldsymbol{H} &= N\sum_{i,j=1}^{p}\sum_{g,h=1}^{q} g^{ij}(b_{ig} - \beta_{ig})a_{gh}(b_{jh} - \beta_{jh}) \\ &= [\operatorname{vec}(\boldsymbol{B}' - \boldsymbol{\beta}')]'\left(\frac{1}{N}\boldsymbol{G}^{-1}\otimes\boldsymbol{A}\right)\operatorname{vec}(\boldsymbol{B}' - \boldsymbol{\beta}') \xrightarrow{d} \chi^2_{pq} \end{aligned} \tag{8}$$

because $(1/N)\boldsymbol{G} \xrightarrow{p} \boldsymbol{\Sigma}$, $(1/N)\boldsymbol{A} \to \boldsymbol{A}_0$, and the limiting distribution of $\sqrt{N}\operatorname{vec}(\boldsymbol{B}' - \boldsymbol{\beta}')$ is $N(\boldsymbol{\Sigma}\otimes\boldsymbol{A}_0^{-1})$. ■

Theorem 8.11.2 agrees with the first term of the asymptotic expansion of $-2\log\lambda$ given by Theorem 8.5.2 for sampling from a normal distribution. The test and confidence procedures discussed in Sections 8.3 and 8.4 can be applied using this χ^2-distribution.

The criterion $U = \lambda^{2/N}$ can be written as $U = \prod_{i=1}^{p} V_i$, where V_i is defined in (8) of Section 8.4. The term V_i has the form of U; that is, it is the ratio of the sum of squares of residuals of $x_{i\alpha}$ regressed on $x_{1\alpha}, \dots, x_{i-1,\alpha}, z_\alpha$ to the sum regressed on $x_{1\alpha}, \dots, x_{i-1,\alpha}$. It follows that under the null hypothesis $V_1, \dots, V_p$ are asymptotically independent and $-N\log V_i \xrightarrow{d} \chi_q^2$. Thus $-N\log U = -N\sum_{i=1}^{p}\log V_i \xrightarrow{d} \chi_{pq}^2$. This argument justifies the step-down procedure asymptotically.

Section 8.6 gave several other criteria for the general linear hypothesis: the Lawley–Hotelling trace $\operatorname{tr} \boldsymbol{H}\boldsymbol{G}^{-1}$, the Bartlett–Nanda–Pillai trace $\operatorname{tr} \boldsymbol{H}(\boldsymbol{G}+\boldsymbol{H})^{-1}$, and the Roy maximum root of $\boldsymbol{H}\boldsymbol{G}^{-1}$ or $\boldsymbol{H}(\boldsymbol{G}+\boldsymbol{H})^{-1}$. The limiting distributions of $N\operatorname{tr}\boldsymbol{H}\boldsymbol{G}^{-1}$ and $N\operatorname{tr}\boldsymbol{H}(\boldsymbol{G}+\boldsymbol{H})^{-1}$ are again χ_{pq}^2. The limiting distribution of the maximum characteristic root of $N\boldsymbol{H}\boldsymbol{G}^{-1}$ or $N\boldsymbol{H}(\boldsymbol{G}+\boldsymbol{H})^{-1}$ is the distribution of the maximum characteristic root of $\boldsymbol{H}$ having the distributions $W(\boldsymbol{I}, q)$ (Lemma 8.11.1). Significance points for these test criteria are available in Appendix B.

8.11.2. Elliptically Contoured Matrix Distributions

In Section 8.3.2 the $p \times N$ matrix of observations on the dependent variable was defined as $\boldsymbol{X} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_N)$, and the $q \times N$ matrix of observations on the independent variables as $\boldsymbol{Z} = (\boldsymbol{z}_1, \dots, \boldsymbol{z}_N)$; the two matrices are related by $\mathscr{E}\boldsymbol{X} = \boldsymbol{\beta}\boldsymbol{Z}$. Note that in this chapter the matrices of observations have N columns instead of N rows.

Let $\boldsymbol{E} = (\boldsymbol{e}_1, \dots, \boldsymbol{e}_N)$ be a $p \times N$ random matrix with density $|\boldsymbol{\Lambda}|^{-N/2} g[\boldsymbol{F}^{-1}\boldsymbol{E}\boldsymbol{E}'(\boldsymbol{F}')^{-1}]$, where $\boldsymbol{\Lambda} = \boldsymbol{F}\boldsymbol{F}'$. Define $\boldsymbol{X}$ by

$$\boldsymbol{X} = \boldsymbol{\beta}\boldsymbol{Z} + \boldsymbol{E}. \tag{9}$$

In these terms the least squares estimator of $\boldsymbol{\beta}$ is

$$\boldsymbol{B} = \boldsymbol{X}\boldsymbol{Z}'(\boldsymbol{Z}\boldsymbol{Z}')^{-1} = \boldsymbol{C}\boldsymbol{A}^{-1}, \tag{10}$$

where $\boldsymbol{C} = \boldsymbol{X}\boldsymbol{Z}' = \sum_{\alpha=1}^{N} \boldsymbol{x}_\alpha \boldsymbol{z}'_\alpha$ and $\boldsymbol{A} = \boldsymbol{Z}\boldsymbol{Z}' = \sum_{\alpha=1}^{N} \boldsymbol{z}_\alpha \boldsymbol{z}'_\alpha$. Note that the density of $\boldsymbol{E}$ is invariant with respect to multiplication on the right by $N \times N$ orthogonal matrices; that is, $\boldsymbol{E}'$ is left spherical. Then $\boldsymbol{E}'$ has the stochastic representation

$$\boldsymbol{E}' \overset{d}{=} \boldsymbol{U}\boldsymbol{T}\boldsymbol{F}', \tag{11}$$

where U has the uniform distribution on $U'U = I_p$, T is the lower triangular matrix with nonnegative diagonal elements satisfying $EE' = TT'$, and F is a lower triangular matrix with nonnegative diagonal elements satisfying $FF' = \Sigma$. We can write

$$(12)\qquad B - \boldsymbol{\beta} = EZ'A^{-1} \stackrel{d}{=} FT'U'Z'A^{-1},$$

$$(13)\qquad H = (B - \boldsymbol{\beta})A(B - \boldsymbol{\beta})' = EZ'A^{-1}ZE' \stackrel{d}{=} FT'U'(Z'A^{-1}Z)UTF'.$$

$$(14)\qquad \begin{aligned} G &= (X - \boldsymbol{\beta}Z)(X - \boldsymbol{\beta}Z)' - H = EE' - H \\ &= E(I_N - Z'A^{-1}Z)E' = FT'U'(I_N - Z'A^{-1}Z)UTF'. \end{aligned}$$

It was shown in Section 8.6 that the likelihood ratio criterion for $H: \boldsymbol{\beta} = 0$, the Lawley–Hotelling trace criterion, the Bartlett–Nanda–Pillai trace criterion, and the Roy maximum root test are invariant with respect to linear transformations $x \to Kx$. Then Corollary 4.5.5 implies the following theorem.

Theorem 8.11.3. *Under the null hypothesis $\boldsymbol{\beta} = 0$, the distribution of each invariant criterion when the distribution of E' is left spherical is the same as the distribution under normality.*

Thus the tests and confidence regions described in Section 8.7 are valid for left-spherical distributions E'.

The matrices $Z'A^{-1}Z$ and $I_N - Z'A^{-1}Z$ are idempotent of ranks q and $N-q$. There is an orthogonal matrix O_N such that

$$(15)\qquad OZ'A^{-1}ZO' = \begin{bmatrix} I_q & 0 \\ 0 & 0 \end{bmatrix}, \qquad O(I_N - Z'A^{-1}Z)O' = \begin{bmatrix} 0 & 0 \\ 0 & I_{N-q} \end{bmatrix}.$$

The transformation $V = O'U$ is uniformly distributed on $V'V = I_p$, and

$$(16)\qquad H = KV'\begin{bmatrix} I_q & 0 \\ 0 & 0 \end{bmatrix}VK', \qquad G = KV'\begin{bmatrix} 0 & 0 \\ 0 & I_{N-q} \end{bmatrix}VK',$$

where $K = FT'$.

The trace criterion $\operatorname{tr} HG^{-1}$, for example, is

$$(17)\qquad \operatorname{tr} HG^{-1} = V'\begin{bmatrix} I_1 & 0 \\ 0 & 0 \end{bmatrix}V\left(V'\begin{bmatrix} 0 & 0 \\ 0 & I_{N-q} \end{bmatrix}V\right)^{-1}.$$

The distribution of any invariant criterion depends only on U (or V), not on T.

Since $G+H=FT'TF'$, it is independent of U. A selection of a linear transformation of X can be made on the basis of $G+H$. Let D be a $p\times r$ matrix of rank r that may depend on $G+H$. Define $x_\alpha^* = D'x_\alpha$. Then $\mathscr{E}x_\alpha^* = (D'\boldsymbol{\beta})z_\alpha$, and the hypothesis $\boldsymbol{\beta}=\mathbf{0}$ implies $D'\boldsymbol{\beta}=\mathbf{0}$. Let $X^* = (x_1^*,\dots,x_N^*) = D'X$, $\boldsymbol{\beta}_D = D'\boldsymbol{\beta}$, $E_D = D'E$, $H_D = D'HD$, $G_D = D'GD$. Then $E_D' = E'D \overset{d}{=} UTF'D'$. The invariant test criteria for $\boldsymbol{\beta}_D = \mathbf{0}$ are those for $\boldsymbol{\beta}=\mathbf{0}$ and have the same distributions under the null hypothesis as for the normal distribution with p replaced by r.

PROBLEMS

8.1. (Sec. 8.2.2) Consider the following sample (for $N=8$):

Weight of grain	40	17	9	15	6	12	5	9
Weight of straw	53	19	10	29	13	27	19	30
Amount of fertilizer	24	11	5	12	7	14	11	18

Let $z_{2\alpha}=1$, and let $z_{1\alpha}$ be the amount of fertilizer on the αth plot. Estimate $\boldsymbol{\beta}$ for this sample. Test the hypothesis $\boldsymbol{\beta}_1 = \mathbf{0}$ at the 0.01 significance level.

8.2. (Sec. 8.2) Show that Theorem 3.2.1 is a special case of Theorem 8.2.1. [*Hint:* Let $q=1$, $z_\alpha = 1$, $\boldsymbol{\beta} = \mu$.]

8.3. (Sec. 8.2) Prove Theorem 8.2.3.

8.4. (Sec. 8.2) Show that $\hat{\boldsymbol{\beta}}$ minimizes the generalized variance

$$\left|\sum_{\alpha=1}^{N}(x_\alpha - \boldsymbol{\beta}z_\alpha)(x_\alpha - \boldsymbol{\beta}z_\alpha)'\right|.$$

8.5. (Sec. 8.3) In the following data [Woltz, Reid, and Colwell (1948), used by R. L. Anderson and Bancroft (1952)] the variables are x_1, rate of cigarette burn; x_2, the percentage of nicotine; z_1, the percentage of nitrogen; z_2, of chlorine; z_3, of potassium; z_4, of phosphorus; z_5, of calcium; and z_6, of magnesium; and $z_7=1$; and $N=25$:

$$\sum_{\alpha=1}^{N} x_\alpha = \begin{pmatrix} 42.20 \\ 54.03 \end{pmatrix}, \qquad \sum_{\alpha=1}^{N} z_\alpha = \begin{pmatrix} 53.92 \\ 62.02 \\ 56.00 \\ 12.25 \\ 89.79 \\ 24.10 \\ 25 \end{pmatrix},$$

$$\sum_{\alpha=1}^{N}(x_\alpha - \bar{x})(x_\alpha - \bar{x})' = \begin{pmatrix} 0.6690 & 0.4527 \\ 0.4527 & 6.5921 \end{pmatrix},$$

$$\sum_{\alpha=1}^{N} (z_\alpha - \bar{z})(z_\alpha - \bar{z})'$$

$$= \begin{pmatrix} 1.8311 & -0.3589 & -0.0125 & -0.0244 & 1.6379 & 0.5057 & 0 \\ -0.3589 & 8.8102 & -0.3469 & 0.0352 & 0.7920 & 0.2173 & 0 \\ -0.0125 & -0.3469 & 1.5818 & -0.0415 & -1.4278 & -0.4753 & 0 \\ -0.0244 & 0.0352 & -0.0415 & 0.0258 & 0.0043 & 0.0154 & 0 \\ 1.6379 & 0.7920 & -1.4278 & 0.0043 & 3.7248 & 0.9120 & 0 \\ 0.5057 & 0.2173 & -0.4753 & 0.0154 & 0.9120 & 0.3828 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\sum_{\alpha=1}^{N} (z_\alpha - \bar{z})(x_\alpha - \bar{x})' = \begin{pmatrix} 0.2501 & 2.6691 \\ -1.5136 & -2.0617 \\ 0.5007 & -0.9503 \\ -0.0421 & -0.0187 \\ -0.1914 & 3.4020 \\ -0.1586 & 1.1663 \\ 0 & 0 \end{pmatrix}.$$

(a) Estimate the regression of x_1 and x_2 on z_1, z_5, z_6, and z_7.

(b) Estimate the regression on all seven variables.

(c) Test the hypothesis that the regression on z_2, z_3, and z_4 is 0.

8.6. (Sec. 8.3) Let $q = 2$, $z_{1\alpha} = w_\alpha$ (scalar), $z_{2\alpha} = 1$. Show that the U-statistic for testing the hypothesis $\boldsymbol{\beta}_1 = \mathbf{0}$ is a monotonic function of a T^2-statistic, and give the T^2-statistic in a simple form. (See Problem 5.1.)

8.7. (Sec. 8.3) Let $z_{q\alpha} = 1$, let $q_2 = 1$, and let

$$A^* = \left[\sum_\alpha (z_{i\alpha} - \bar{z}_i)(z_{j\alpha} - \bar{z}_j)\right], \qquad i, j = 1, \dots, q_1 = q - 1.$$

Prove that

$$(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1)(A_{11} - A_{12}A_{22}^{-1}A_{21})(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1)' = (\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1)A^*(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1)'.$$

8.8. (Sec. 8.3) Let $q_1 = q_2$. How do you test the hypothesis $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$?

8.9. (Sec. 8.3) Prove

$$\hat{\boldsymbol{\beta}}_{1\Omega} = \sum_\alpha x_\alpha (z_\alpha^{(1)} - A_{12}A_{22}^{-1}z_\alpha^{(2)})' \left[\sum_\alpha (z_\alpha^{(1)} - A_{12}A_{22}^{-1}z_\alpha^{(2)})(z_\alpha^{(1)} - A_{12}A_{22}^{-1}z_\alpha^{(2)})'\right]^{-1}$$

$$= (C_1 - C_2A_{22}^{-1}A_{21})(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}.$$

8.10. (Sec. 8.4) By comparing Theorem 8.2.2 and Problem 8.9, prove Lemma 8.4.1.

8.11. (Sec. 8.4) Prove Lemma 8.4.1 by showing that the density of $\hat{\mathbf{B}}_{1\Omega}$ and $\hat{\mathbf{B}}_{2\omega}$ is

$$K_1 \exp\left[-\tfrac{1}{2}\operatorname{tr} \Sigma^{-1}(\hat{\mathbf{B}}_{1\Omega} - \mathbf{B}_1^*)A_{11\cdot 2}(\hat{\mathbf{B}}_{1\Omega} - \mathbf{B}_1^*)'\right]$$
$$\cdot K_2 \exp\left[-\tfrac{1}{2}\operatorname{tr} \Sigma^{-1}(\hat{\mathbf{B}}_{2\omega} - \mathbf{B}_2)A_{22}(\hat{\mathbf{B}}_{2\omega} - \mathbf{B}_2)'\right].$$

8.12. (Sec. 8.4) Show that the cdf of $U_{3,3,n}$ is

$$I_u(\tfrac{1}{2}n - 1, \tfrac{3}{2}) + \frac{\Gamma(n+2)\Gamma[\frac{1}{2}(n+1)]}{\Gamma(n-1)\Gamma(\frac{1}{2}n - 1)\sqrt{\pi}}$$
$$\cdot\left\{\frac{2u^{\frac{1}{2}n-1}\sqrt{1-u}}{n(n-1)} + \frac{u^{\frac{1}{2}(n-1)}}{n-1}\left[\arcsin(2u-1) - \tfrac{1}{2}\pi\right]\right.$$
$$\left. + \frac{2u^{\frac{1}{2}n}}{n}\log\left(\frac{1+\sqrt{1-u}}{\sqrt{u}}\right) + \frac{2u^{\frac{1}{2}n-1}(1-u)^{\frac{3}{2}}}{3(n+1)}\right\}.$$

[*Hint:* Use Theorem 8.4.4. The region $\{0 \le z_1 \le 1, 0 \le z_2 \le 1, z_1^2 z_2 \le u\}$ is the union of $\{0 \le z_1 \le 1, 0 \le z_2 \le u\}$ and $\{0 \le z_1 \le u/z_2,\ u \le z_2 \le 1\}$.]

8.13. (Sec. 8.4) Find $\Pr\{U_{4,3,n} \ge u\}$.

8.14. (Sec. 8.4) Find $\Pr\{U_{4,4,n} \ge u\}$.

8.15. (Sec. 8.4) For $p \le m$ find $\mathscr{E}U^h$ from the density of G and H. [*Hint:* Use the fact that the density of $K + \sum_{i=1}^t V_i V_i'$ is $W(\Sigma, s+t)$ if the density of K is $W(\Sigma, s)$ and $V_1, \dots, V_t$ are independently distributed as $N(\mathbf{0}, \Sigma)$.]

8.16. (Sec. 8.4)

(a) Show that when p is even, the characteristic function of $Y = \log U_{p,m,n}$, say $\phi(t) = \mathscr{E} e^{itY}$, is the reciprocal of a polynomial.

(b) Sketch a method of inverting the characteristic function of Y by the method of residues.

(c) Show that the resulting density of U is a polynomial in $\sqrt{u}$ and $\log u$ with possibly a factor of $u^{-\frac{1}{2}}$.

8.17. (Sec. 8.5) Use the asymptotic expansion of the distribution to compute $\Pr\{-k \log U_{3,3,n} \le M^*\}$ for

(a) $n = 8$, $M^* = 14.7$,
(b) $n = 8$, $M^* = 21.7$,
(c) $n = 16$, $M^* = 14.7$,
(d) $n = 16$, $M^* = 21.7$.

(Either compute to the third decimal place or use the expansion to the k^{-4} term.)

8.18. (Sec. 8.5) In case $p = 3$, $q_1 = 4$, and $n = N - q = 20$, find the 50% significance point for $k \log U$ (a) using $-2\log\lambda$ as χ^2 and (b) using $-k\log U$ as χ^2. Using more terms of this expansion, evaluate the exact significance levels for your answers to (a) and (b).

8.19. (Sec. 8.6.5) Prove for $l_i \geq 0$, $i = 1, \dots, p$,

$$\sum_{i=1}^{p} \frac{l_i}{1+l_i} \leq \log \prod_{i=1}^{p} (1 + l_i) \leq \sum_{i=1}^{p} l_i.$$

Comment: The inequalities imply an ordering of the values of the Bartlett–Nanda–Pillai trace, the negative logarithm of the likelihood ratio criterion, and the Lawley–Hotelling trace.

8.20. (Sec. 8.6) *The multivariate beta density.* Let $\boldsymbol{H}$ and $\boldsymbol{G}$ be independently distributed according to $W(\boldsymbol{\Sigma}, m)$ and $W(\boldsymbol{\Sigma}, n)$, respectively. Let $\boldsymbol{C}$ be a matrix such that $\boldsymbol{CC}' = \boldsymbol{H} + \boldsymbol{G}$, and let

$$\boldsymbol{L} = \boldsymbol{C}^{-1}\boldsymbol{H}\boldsymbol{C}'^{-1}.$$

Show that the density of $\boldsymbol{L}$ is

$$\frac{\Gamma_p\left[\frac{1}{2}(m+n)\right]}{\Gamma_p\left(\frac{1}{2}m\right)\Gamma_p\left(\frac{1}{2}n\right)} |\boldsymbol{L}|^{\frac{1}{2}(m-p-1)} |\boldsymbol{I} - \boldsymbol{L}|^{\frac{1}{2}(n-p-1)}$$

for $\boldsymbol{L}$ and $\boldsymbol{I} - \boldsymbol{L}$ positive definite, and 0 otherwise.

8.21. (Sec. 8.9) Let $\boldsymbol{Y}_{ij}$ (a p-component vector) be distributed according to $N(\boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma})$, where $\mathscr{E}\boldsymbol{Y}_{ij} = \boldsymbol{\mu}_{ij} = \boldsymbol{\mu} + \boldsymbol{\lambda}_i + \boldsymbol{\nu}_j + \boldsymbol{\gamma}_{ij}$, $\Sigma_i \boldsymbol{\lambda}_i = \boldsymbol{0} = \Sigma_j \boldsymbol{\nu}_j = \Sigma_i \boldsymbol{\gamma}_{ij} = \Sigma_j \boldsymbol{\gamma}_{ij}$; the $\boldsymbol{\gamma}_{ij}$ are the interactions. If m observations are made on each $\boldsymbol{Y}_{ij}$ (say $\boldsymbol{y}_{ij1}, \dots, \boldsymbol{y}_{ijm}$), how do you test the hypothesis $\boldsymbol{\lambda}_i = \boldsymbol{0}$, $i = 1, \dots, r$? How do you test the hypothesis $\boldsymbol{\gamma}_{ij} = \boldsymbol{0}$, $i = 1, \dots, r$, $j = 1, \dots, c$?

8.22. (Sec. 8.9) *The Latin square.* Let $\boldsymbol{Y}_{ij}$, $i, j = 1, \dots, r$, be distributed according to $N(\boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma})$, where $\mathscr{E}\boldsymbol{Y}_{ij} = \boldsymbol{\mu}_{ij} = \boldsymbol{\gamma} + \boldsymbol{\lambda}_i + \boldsymbol{\nu}_j + \boldsymbol{\mu}_k$ and $k = j - i + 1 \pmod r$ with $\Sigma \boldsymbol{\lambda}_i = \Sigma \boldsymbol{\nu}_j = \Sigma \boldsymbol{\mu}_k = \boldsymbol{0}$.

(a) Give the univariate analysis of variance table for main effects and error (including sums of squares, numbers of degrees of freedom, and mean squares).

(b) Give the table for the vector case.

(c) Indicate in the vector case how to test the hypothesis $\boldsymbol{\lambda}_i = \boldsymbol{0}$, $i = 1, \dots, r$.

8.23. (Sec. 8.9) Let x_1 be the yield of a process and x_2 a quality measure. Let $z_1 = 1$, $z_2 = \pm 10°$ (temperature relative to average) $z_3 = \pm 0.75$ (relative measure of flow of one agent), and $z_4 = \pm 1.50$ (relative measure of flow of another agent). [See Anderson (1955a) for details.] Three observations were made on x_1

and x_2 for each possible triplet of values of z_2, z_3, and z_4. The estimate of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} 58.529 & -0.3829 & -5.050 & 2.308 \\ 98.675 & 0.1558 & 4.144 & -0.700 \end{pmatrix};$$

$s_1 = 3.090$, $s_2 = 1.619$, and $r = -0.6632$ can be used to compute S or $\hat{\boldsymbol{\Sigma}}$.

(a) Formulate an analysis of variance model for this situation.
(b) Find a confidence region for the effects of temperature (i.e., β_{12}, β_{22}).
(c) Test the hypothesis that the two agents have no effect on the yield and quantity.

8.24. (Sec. 8.6) Interpret the transformations referred to in Theorem 8.6.1 in the original terms; that is, $H: \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^*$ and $z_\alpha^{(1)}$.

8.25. (Sec. 8.6) Find the cdf of $\operatorname{tr} HG^{-1}$ for $p = 2$. [*Hint:* Use the distribution of the roots given in Chapter 13.]

8.26. (Sec. 8.10.1) *Bartlett–Nanda–Pillai V-test as a Bayes procedure.* Let $\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_{m+n}$ be independently normally distributed with covariance matrix $\boldsymbol{\Sigma}$ and means $\mathscr{E}\boldsymbol{w}_i = \boldsymbol{\gamma}_i$, $i = 1, \dots, m$, $\mathscr{E}\boldsymbol{w}_i = \boldsymbol{0}$, $i = m+1, \dots, m+n$. Let Π_0 be defined by $[\boldsymbol{\Gamma}_1, \boldsymbol{\Sigma}] = [\boldsymbol{0}, (\boldsymbol{I} + \boldsymbol{CC}')^{-1}]$, where the $p \times m$ matrix $\boldsymbol{C}$ has a density proportional to $|\boldsymbol{I} + \boldsymbol{CC}'|^{-\frac{1}{2}(n+m)}$, and $\boldsymbol{\Gamma}_1 = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_m)$; let Π_1 be defined by $[\boldsymbol{\Gamma}_1, \boldsymbol{\Sigma}] = [(\boldsymbol{I} + \boldsymbol{CC}')^{-1}\boldsymbol{C}, (\boldsymbol{I} + \boldsymbol{CC}')^{-1}]$ where $\boldsymbol{C}$ has a density proportional to $|\boldsymbol{I} + \boldsymbol{CC}'|^{-\frac{1}{2}(n+m)} e^{\frac{1}{2}\operatorname{tr} \boldsymbol{C}'(\boldsymbol{I}+\boldsymbol{CC}')^{-1}\boldsymbol{C}}$.

(a) Show that the measures are finite for $n \ge p$ by showing $\operatorname{tr} \boldsymbol{C}'(\boldsymbol{I} + \boldsymbol{CC}')^{-1}\boldsymbol{C} < m$ and verifying that the integral of $|\boldsymbol{I} + \boldsymbol{CC}'|^{-\frac{1}{2}(n+m)}$ is finite. [*Hint:* Let $\boldsymbol{C} = (\boldsymbol{c}_1, \dots, \boldsymbol{c}_m)$, $\boldsymbol{D}_j = \boldsymbol{I} + \sum_{i=1}^{j} \boldsymbol{c}_i\boldsymbol{c}_i' = \boldsymbol{E}_j\boldsymbol{E}_j'$, $\boldsymbol{c}_j = \boldsymbol{E}_{j-1}\boldsymbol{d}_j$, $j = 1, \dots, m$ $(\boldsymbol{E}_0 = \boldsymbol{I})$. Show $|\boldsymbol{D}_j| = |\boldsymbol{D}_{j-1}|(1 + \boldsymbol{d}_j'\boldsymbol{d}_j)$ and hence $|\boldsymbol{D}_m| = \prod_{j=1}^{m}(1 + \boldsymbol{c}_j'\boldsymbol{d}_j)$. Then refer to Problem 5.15.]
(b) Show that the inequality (26) of Section 5.6 is equivalent to

$$\operatorname{tr}\left(\sum_{i=1}^{m+n} \boldsymbol{w}_i\boldsymbol{w}_i'\right)^{-1} \sum_{i=1}^{m} \boldsymbol{w}_i\boldsymbol{w}_i' \ge k.$$

Hence the Bartlett–Nanda–Pillai V-test is Bayes and thus admissible.

8.27. (Sec. 8.10.1) *Likelihood ratio test as a Bayes procedure.* Let $\boldsymbol{w}_1, \dots, \boldsymbol{w}_{m+n}$ be independently normally distributed with covariance matrix $\boldsymbol{\Sigma}$ and means $\mathscr{E}\boldsymbol{w}_i = \boldsymbol{\gamma}_i$, $i = 1, \dots, m$, $\mathscr{E}\boldsymbol{w}_i = \boldsymbol{0}$, $i = m+1, \dots, m+n$, with $n \ge m + p$. Let Π_0 be defined by $[\boldsymbol{\Gamma}_1, \boldsymbol{\Sigma}] = [\boldsymbol{0}, (\boldsymbol{I} + \boldsymbol{CC}')^{-1}]$, where the $p \times m$ matrix $\boldsymbol{C}$ has a density proportional to $|\boldsymbol{I} + \boldsymbol{CC}'|^{-\frac{1}{2}(n+m)}$ and $\boldsymbol{\Gamma}_1 = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_m)$; let Π_1 be defined by

$$[\boldsymbol{\Gamma}_1, \boldsymbol{\Sigma}] = \left[(\boldsymbol{I} + \boldsymbol{CC}')^{-1}\boldsymbol{CD}, (\boldsymbol{I} + \boldsymbol{CC}')^{-1}\right],$$

where the m columns of D are conditionally independently normally distributed with means $\mathbf{0}$ and covariance matrix $[I - C'(I+CC')^{-1}C]^{-1}$, and C has (marginal) density proportional to

$$|I+CC'|^{-\frac{1}{2}(n+m)}|I - C'(I+CC')^{-1}C|^{-\frac{1}{2}m}.$$

(a) Show the measures are finite. [*Hint:* See Problem 8.26.]

(b) Show that the inequality (26) of Section 5.6 is equivalent to

$$\frac{|\sum_{i=1}^{m+n} w_i w_i'|}{|\sum_{i=m+1}^{m+n} w_i w_i'|} \geq k.$$

Hence the likelihood ratio test is Bayes and thus admissible.

8.28. (Sec. 8.10.1) *Admissibility of the likelihood ratio test*. Show that the acceptance region $|ZZ'|/|ZZ'+XX'| \geq c$ satisfies the conditions of Theorem 8.10.1. [*Hint:* The acceptance region can be written $\prod_{i=1}^{t} m_i > c$, where $m_i = 1 - \lambda_i$, $i = 1,\dots,t$.]

8.29. (Sec. 8.10.1) *Admissibility of the Lawley–Hotelling test*. Show that the acceptance region $\operatorname{tr} XX'(ZZ')^{-1} \leq c$ satisfies the conditions of Theorem 8.10.1.

8.30. (Sec. 8.10.1) *Admissibility of the Bartlett–Nanda–Pillai trace test*. Show that the acceptance region $\operatorname{tr} X'(ZZ'+XX')^{-1}X \leq c$ satisfies the conditions of Theorem 8.10.1.

8.31. (Sec. 8.10.1) Show that if A and B are positive definite and $A - B$ is positive semidefinite, then $B^{-1} - A^{-1}$ is positive semidefinite.

8.32. (Sec. 8.10.1) Show that the boundary of $\tilde{A}$ has m-measure 0. [*Hint:* Show that (closure of $\tilde{A}$) $\subset \tilde{A} \cup C$, where $C = \{V | U - YY' \text{ is singular}\}$.]

8.33. (Sec. 8.10.1) Show that if $A \subset R_{<}^{m}$ is convex and monotone in majorization, then A^* is convex. [*Hint:* Show

$$(px+qy)_{\downarrow} \succ_{w} px_{\downarrow} + qy_{\downarrow},$$

where

$$z_{\downarrow} = (z_{[1]},\dots,z_{[m]})' \in R_{<}^{m}.]$$

8.34. (Sec. 8.10.1) Show that $C(\boldsymbol{\lambda})$ is convex. [*Hint:* Follow the solution of Problem 8.33 to show $(px+qy) \prec_{w} \boldsymbol{\lambda}$ if $x \prec_{w} \boldsymbol{\lambda}$ and $y \prec_{w} \boldsymbol{\lambda}$.]

8.35. (Sec. 8.10.1) Show that if A is monotone, then A^* is monotone. [*Hint:* Use the fact that

$$x_{[k]} = \max_{i_1,\dots,i_k}\left\{\min(x_{i_1},\dots,x_{i_k})\right\}.]$$

8.36. (Sec. 8.10.2) *Monotonicity of the power function of the Bartlett–Nanda–Pillai trace test.* Show that

$$\operatorname{tr}(uu' + B)(uu' + B + W)^{-1} \le K$$

is convex in u for fixed positive semidefinite B and positive definite $B + W$ if $0 \le K \le 1$. [*Hint:* Verify

$$\begin{aligned}(uu' + B + W)^{-1}\\ &= (B+W)^{-1} - \frac{1}{1 + u'(B+W)^{-1}u}(B+W)^{-1}uu'(B+W)^{-1}.\end{aligned}$$

The resulting quadratic form in u involves the matrix $(\operatorname{tr} A)I - A$ for $A = (B+W)^{-\frac{1}{2}}B(B+W)^{-\frac{1}{2}}$; show that this matrix is positive semidefinite by diagonalizing A.]

8.37. (Sec. 8.8) Let $x_\alpha^{(\nu)}$, $\alpha = 1, \dots, N_\nu$, be observations from $N(\mu^{(\nu)}, \Sigma)$, $\nu = 1, \dots, q$. What criterion may be used to test the hypothesis that

$$\mu^{(\nu)} = \sum_{h=1}^{m} \gamma_h c_{h\nu} + \mu,$$

where $c_{h\nu}$ are given numbers and γ_ν, μ are unknown vectors? [*Note:* This hypothesis (that the means lie on an m-dimensional hyperplane with ratios of distances known) can be put in the form of the general linear hypothesis.]

8.38. (Sec. 8.2) Let x_α be an observation from $N(\beta z_\alpha, \Sigma)$, $\alpha = 1, \dots, N$. Suppose there is a known fixed vector γ such that $\beta\gamma = 0$. How do you estimate β?

8.39. (Sec. 8.8) What is the largest group of transformations on $y_\alpha^{(i)}$, $\alpha = 1, \dots, N_i$, $i = 1, \dots, q$, that leaves (1) invariant? Prove the test (12) is invariant under this group.