COMPARING MULTIVARIATE REGRESSION IN TWO POPULATIONS
WHEN MEASUREMENT ERRORS HAVE KNOWN VARIANCES

BY

T. W. F. STROUD

TECHNICAL REPORT NO. 63
OCTOBER 25, 1971

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT GP-27210

Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
COMPARING MULTIVARIATE REGRESSION IN TWO POPULATIONS
WHEN MEASUREMENT ERRORS HAVE KNOWN VARIANCES

by

T. W. F. Stroud

TECHNICAL REPORT NO. 63
OCTOBER 25, 1971

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT GP-27210
Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
COMPARING MULTIVARIATE REGRESSION IN TWO POPULATIONS
WHEN MEASUREMENT ERRORS HAVE KNOWN VARIANCES

by

T. W. F. Stroud
Queen's University at Kingston

1. Introduction

1.1 Model, application and related work.

Let \( \tilde{X} \) and \( \tilde{Y} \) be two observable random column vectors of the same dimension, defined in terms of four mutually independent normally distributed random column vectors \( X^*, X^{**}, Y^* \) and \( Y^{**} \) by

\[
\tilde{X} = X^* + X^{**} \quad \text{and} \quad \tilde{Y} = Y^* + Y^{**}.
\]

\( X^* \) and \( Y^* \) are to be considered as vectors of interest, and \( X^{**} \) and \( Y^{**} \) are to be interpreted as errors of measurement. The vectors of interest and errors of measurement cannot be observed individually.

Let all vectors be partitioned into a first part and a second part. Let all the first parts have dimension \( p \), and all the second parts have dimension \( q \); so that the full vectors have dimension \( p+q \). Denote the first and second parts with single and double dots, thus:

\[
X^* = \begin{pmatrix} \ddots \vline & \ddots \vline & \vdots \\ \vdots & \ddots & \ddots \\ \vdots & \ddots & \ddots \end{pmatrix} X^*, \quad X^{**} = \begin{pmatrix} \ddots \vline & \ddots \vline & \vdots \\ \vdots & \ddots & \ddots \\ \vdots & \ddots & \ddots \end{pmatrix} X^{**}, \quad Y^* = \begin{pmatrix} \ddots \vline & \ddots \vline & \vdots \\ \vdots & \ddots & \ddots \\ \vdots & \ddots & \ddots \end{pmatrix} Y^*, \quad Y^{**} = \begin{pmatrix} \ddots \vline & \ddots \vline & \vdots \\ \vdots & \ddots & \ddots \\ \vdots & \ddots & \ddots \end{pmatrix} Y^{**}.
\]
We assume that $\dot{X}^{**}$ is statistically independent of $\dot{X}^{**}$, that $\dot{Y}^{**}$ is statistically independent of $\dot{Y}^{**}$, and that $X^{**}$ and $Y^{**}$ have zero mean vectors and the same known nonsingular covariance matrix $\Delta = \text{diag}(\Delta_1, \Delta_2)$, where $\Delta_1$ has dimension $p \times p$ and $\Delta_2$ has dimension $q \times q$.

As may be seen from the example following, we wish to test separately, against all alternatives, the hypotheses

(1.1) \hspace{1cm} H_1 : E[\dot{X}^*|\dot{X}^*=u] = E[\dot{Y}^*|\dot{Y}^*=u] \text{ for all } u \ (q \times 1)

and

(1.2) \hspace{1cm} H_2 : \text{Cov}[\dot{X}^*|\dot{X}^*=u] = \text{Cov}[\dot{Y}^*|\dot{Y}^*=u] \text{ for all } u.

Because these hypotheses are invariant when $X^*$, $X^{**}$, $Y^*$ and $Y^{**}$ are each premultiplied by $\Delta^{-\frac{1}{2}}$, we may, without loss of generality, assume that $\Delta=I$.

These two problems originated from a desire to compare the learning taking place in two schools, where batteries of standarized achievement tests were administered to students at intervals two years apart ("pretests" and "posttests"). The conditional expectation of the posttest vector, given the pretest vector, may be taken as a measure of learning, and the conditional covariance matrix may be taken as a measure of lack of uniformity of learning. We may be interested in whether the average learning is the same in the two schools, even if the dispersions are different.
A more complete discussion of this application for the case \( p=q=1 \) is presented in Stroud (1972), where asymptotic tests for that case are proposed. The \( p=q=1 \) model is appropriate if one considers a fixed linear functional of the vector score (composite score), rather than the vector score itself. The present paper extends these results to the multivariate case, so that the vector score may be treated.

In the mental testing application, the vectors of interest \( \mathbf{X}^* \) and \( \mathbf{Y}^* \) are thought of as "true scores" (Lord and Novick, 1968). It is proposed that the "known" measurement error covariance matrices \( \Delta_1 \) and \( \Delta_2 \) be based on the "standard error of measurement" figures provided by the achievement test publishers. In mental test theory it is common to assume the subtest measurement errors to be uncorrelated (i.e., \( \Delta_1 \) and \( \Delta_2 \) diagonal); otherwise one could use arbitrary correlation figures which seemed reasonable.

The regression hypotheses \( H_1 \) and \( H_2 \) are treated in this paper by the consideration of the joint distribution of the \( \hat{\mathbf{X}} \) and \( \bar{\mathbf{X}} \) vectors. Lord (1960) has treated \( H_1 \) in the case where \( p=q=1 \), equal slopes are assumed, and duplicate measurements of the predictor variable are available instead of known measurement error variances, by studying instead the conditional distributions of \( \hat{\mathbf{X}}^* \) given \( \bar{\mathbf{X}}^* \) and \( \hat{\mathbf{Y}}^* \) given \( \bar{\mathbf{Y}}^* \) (i.e., with \( \bar{\mathbf{X}}^* \) and \( \bar{\mathbf{Y}}^* \) treated as unobservable non-random quantities). This conditional approach to regression problems is by far the more common; see Moran (1971) and Cochran (1968) and their bibliographies for other results in error-of-measurement regression problems where this approach is taken.
1.2 Formulation of the hypotheses in terms of the parameters of the observed vectors and the sufficient statistics.

Let us formulate the hypotheses $H_1$ and $H_2$ in terms of the parameters of the multivariate normal distributions from which the observations are taken. To begin, let the vectors of interest $X^*$ and $Y^*$ have mean vectors $\mu$ and $\nu$, and positive definite covariance matrices $\Sigma^*$ and $\Psi^*$, respectively. Both $X^{**}$ and $Y^{**}$ have zero mean vectors and covariance matrices equal to the identity. Let

$$\mu = \begin{pmatrix} \hat{\mu} \\ \hat{\mu} \end{pmatrix}, \quad \Sigma^* = \begin{pmatrix} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{21}^* & \Sigma_{22}^* \end{pmatrix},$$

where $\hat{\mu}$, $\hat{\mu}$, $\Sigma_{11}^*$ and $\Sigma_{22}^*$ have dimensions $p \times 1$, $q \times 1$, $p \times p$ and $q \times q$ respectively. Let $\nu$ and $\Psi^*$ be partitioned similarly. Then, by the formulas for the conditional mean vector and covariance matrix of a subvector of a normal random vector (see, e.g., Anderson (1958), p. 29), the hypotheses (1.1) and (1.2) are equivalent to $H_1$ and $H_2$:

$$H_1 : \left\{ \begin{array}{l} \Sigma_{12}^* \Sigma_{22}^{-1} = \Psi_{12}^* \Psi_{22}^{-1} \\ \hat{\mu} - \Sigma_{12}^* \Sigma_{22}^{-1} \hat{\mu} = \hat{\nu} - \Psi_{12}^* \Psi_{22}^{-1} \hat{\nu} \end{array} \right.$$

(1.3)

$$H_2 : \Sigma_{11}^* - \Sigma_{12}^* \Sigma_{22}^{-1} = \Psi_{11}^* - \Psi_{12}^* \Psi_{22}^{-1} \Psi_{21}^*.$$

(1.4)
To write these hypotheses in terms of the parameters of the distributions of the observed vectors $\tilde{X}$ and $\tilde{Y}$, note that $\tilde{X}$ and $\tilde{Y}$ are normally and independently distributed with mean vectors $\mu$, $\nu$ and covariance matrices $\Sigma = \Sigma^* + I$ and $\Psi = \Psi^* + I$ respectively. Partition $\Sigma$, $\Psi$ in the same way as $\Sigma^*$, $\Psi^*$; then (1.3) and (1.4) become, respectively,

\begin{align*}
H_1: & \left\{ \begin{array}{c}
\Sigma_{12} (\Sigma_{22} - I)^{-1} = \Psi_{12} (\Psi_{22} - I)^{-1} \\
\mu - \Sigma_{12} (\Sigma_{22} - I)^{-1} \mu = \nu - \Psi_{12} (\Psi_{22} - I)^{-1} \nu
\end{array} \right.
\end{align*}

\begin{equation}
H_2: \Sigma_{11} - \Sigma_{12} (\Sigma_{22} - I)^{-1} \Sigma_{21} = \Psi_{11} - \Psi_{12} (\Psi_{22} - I)^{-1} \Psi_{21}.
\end{equation}

Suppose we observe a sample of $M$ and of $N$ observations, which are independently distributed as the vector $\tilde{X}$ and as $\tilde{Y}$, respectively. The sample mean vectors $X$ and $Y$ and sample covariance matrices $V/(M-1)$ and $W/(N-1)$ are jointly sufficient statistics for inferences concerning $\mu, \nu, \Sigma$ and $\Psi$. $X$ and $Y$ are independent normally distributed random $(p+q) \times 1$ column vectors, with mean vectors $\mu, \nu$ and covariance matrices $M^{-1} \Sigma$ and $N^{-1} \Psi$, respectively, where $\Sigma > I$ and $\Psi > I$. $V$ and $W$ are two random $(p+q) \times (p+q)$ matrices, statistically independent of each other and of $X$ and $Y$, such that $V$ has the Wishart distribution with $M-1$ degrees of freedom and $EV = (M-1) \Sigma$, denoted by $\mathcal{W}(\Sigma, M-1)$, and such that $W$ has the Wishart distribution $\mathcal{W}(\Psi, N-1)$. In what follows we shall always make reference to $X, Y, V$ and $W$, instead
of the original observations. The symbols \( \tilde{X}, \tilde{Y}, \tilde{Y} \) and \( \tilde{Y} \) refer to the partitioning of \( X \) and \( Y \) in the same manner as \( \mu \).

Possible testing procedures are considered in Section 2. Because other methods do not appear practical, the proposed procedure is an asymptotic test based on unrestricted maximum-likelihood estimators described by Wald (1943). In Section 3 this procedure is applied to \( H_2 \) first, because it is simpler; Section 4 contains the application to \( H_1 \). In Section 5 the formulas are applied to some educational test score data, alongside the univariate \((p=q=1)\) version of the procedures with respect to the same data.

2. Discussion of Possible Test Procedures

2.1 Likelihood-ratio and other procedures.

When we try to apply the likelihood-ratio (LR) method to either \( H_1 \) or \( H_2 \), the result appears to be extremely complicated. Since \( H_1 \) is to be tested independently of \( H_2 \), \( H_1 \) may be regarded as an extension of the Behrens-Fisher problem and possesses all the difficulties of that problem (see, e.g., Kendall and Stuart, 1967, p. 226). Testing \( H_1 \) given \( H_2 \) is no easier since tests must be based on the observed vectors (not the vectors of interest), and the distributions of these vectors are not related in any direct way to conditional covariance matrices given the vectors of interest, whose equality is expressed by \( H_2 \).

There is a treatment of the Behrens-Fisher regression problem (without errors of measurement) by Barankin (1949) and Potthoff (1965) which produces an exact test using the method of Scheffé (1943).
Here the values of the predictor variables are assumed to be fixed and known; we cannot make this assumption here since the predictor variables are the unobservable vectors of interest.

For testing $H_0$, the LR method again produces a set of equations for the restricted maximum likelihood estimators which cannot be solved easily. A complicated sequence of numerical solutions of matrix equations is required. Details may be found in Stroud (1968).

2.2 The proposed procedure.

The method proposed in this paper for testing each of $H_1$ and $H_2$ is an asymptotic test procedure based on unrestricted maximum likelihood estimators (MLE). If the hypothesis is expressed as a vector equation with the right-hand side equal to zero, the test statistic is the quadratic form obtained from the MLE of the left-hand side and the MLE of its asymptotic inverse covariance matrix. The test statistic is asymptotically distributed as central (noncentral) chi-square if the hypothesis is true (false). This result, originally proved by Wald (1943) under rather strict regularity conditions, has been shown by the author (Stroud, 1971) to hold under fairly general circumstances, which apply in particular to the problems of this paper. The version of this result used in the following sections to provide test statistics for $H_1$ and $H_2$ is stated in the following three paragraphs.

Let the random quantities $X, Y, V$ and $W$, the parameters $\mu, \nu, \Sigma$ and $\Psi$ and the integers $M$ and $N$ be as defined near the end of Section 1.2, but assume $M$ to be dependent on $N$ in such a manner that $N/M$
approaches a finite nonzero limit \( \rho \) as \( N \to \infty \). Define \( \hat{\Sigma} = V/M \), \( \hat{\Psi} = W/N \). Note the dependence of \( X, Y, V, W, \hat{\Sigma} \) and \( \hat{\Psi} \) on \( N \), suppressed in the notation. Let \( g \) be an \( r \)-dimensional vector-valued function of \( (\mu, \nu, \Sigma, \Psi) \), possessing continuous second partial derivatives at each point of the parameter space \( \Theta \), assumed to be an open subset of the space of all possible values of \( \Theta = (\mu, \nu, \Sigma, \Psi) \). Assume \( r \leq \dim \Theta = (p+q)(p+q+3) \), and that the matrix of first partial derivatives of \( g \) with respect to \( \Theta \) has full rank \( r \).

To test the hypothesis \( g(\theta) = 0 \) versus the alternative \( g(\theta) \neq 0 \), the proposed statistic is

\[
J_N = N(U_N - NU_N)^{-1} U_N;
\]

where \( U_N = g(X, Y, \hat{\Sigma}, \hat{\Psi}) \), and \( \hat{\Omega}_N \) is the value, when \( \Theta = (X, Y, \hat{\Sigma}, \hat{\Psi}) \), of \( \Omega \), the asymptotic covariance matrix (when \( \Theta \) is fixed) of \( N^{1/2}(U_N - g(\theta)) \). \( \Omega \) is evaluated as follows. For any \( i \) and \( j \), each in the set \( \{1, \ldots, r\} \), denote \( \alpha = g_{i1}(\theta), \beta = g_{ij}(\theta) \); i.e., \( \alpha \) and \( \beta \) are the \( i \)-th and \( j \)-th components, respectively, of \( g(\theta) \). Correspondingly, denote the \( (i, j) \)-th component of the matrix \( \Omega \) by \( \omega_{ij} = \omega(\alpha, \beta) \); then

\[
\omega(\alpha, \beta) = \rho \alpha' \Sigma \beta + \alpha' \Psi \beta + 2 \rho \text{tr} \alpha' \Sigma \beta \Sigma + 2 \text{tr} \alpha' \Psi \beta \Psi ,
\]

where \( \alpha_{\mu}, \alpha_{\nu}, \beta_{\mu} \) and \( \beta_{\nu} \) are \((p+q) \times 1\) vectors of partial derivatives, e.g., the components of \( \alpha_{\mu} \) are, respectively, the partial derivatives of \( \alpha \) with respect to the components of \( \mu \). \( \alpha_{\Sigma}, \alpha_{\Psi}, \beta_{\Sigma} \) and \( \beta_{\Psi} \) are...
\((p+q) \times (p+q)\) symmetric matrices representing partial derivatives with respect to the matrices \(\Sigma\) and \(\Psi\). The off-diagonal components of \(\alpha_{\Sigma}\), etc., include a factor of \(\frac{1}{2}\) for the purpose of obtaining the simplicity of (2.2). (Cf. Stroud (1971, formula 4.6) and Aitken (1953)).

Suppose \(\theta_0\) is a parameter value satisfying \(g(\theta_0) = 0\), and suppose the true parameter assumes values in a sequence \(\{\theta_N\}\) such that \(\frac{1}{N^2}(\theta_N - \theta_0)\) tends to some \(\delta\) as \(N \to \infty\). Then the distribution of \(J_N\) converges, as \(N \to \infty\), to the noncentral chi-square distribution with \(r\) degrees of freedom and noncentrality parameter equal to the limit (cf. Wald, 1943, p. 482) of

\[
N[g(\theta_N)]' \Sigma^{-1}_N g(\theta_N)
\]

where \(\Sigma_N\) is the value of \(\Omega\) when the parameter is \(\theta_N^*\). If \(\theta_N = \theta_0\), or more generally if \(\frac{1}{N^2}(\theta_N - \theta_0) \to 0\), then the limit distribution is central chi-square.

The exact distributions of \(J_N\) for \(H_2\) when \(p=1\) and for \(H_1\) when \(p=q=1\) have been studied. Representations have been obtained which appear too complicated for use in applications (Stroud, 1968). There appears to be no simplification in the results when the hypothesis is true.

For problems such as \(H_2\) where the hypothesis to be tested does not involve \(\mu\) or \(\nu\), we write for brevity \(g(\theta) = g(\Sigma, \Psi)\), and (2.2) becomes

\[
(2.3) \quad \omega(\alpha, \beta) = 2 \rho \text{tr} \Sigma^{-1} \Sigma + 2 \text{tr} \Sigma^{-1} \Psi \Psi
\]
since partial derivatives with respect to \( \mu \) and \( \nu \) are all zero.

In the following section, an explicit version of the components (2.3) of \( \Omega \) for the hypothesis \( H_2 \) is worked out; and in Section 4 the same is done with (2.2) for the hypothesis \( H_1 \). The value of the statistic (2.1) may then be calculated, given the data, by the substitution of the MLE's and the execution of routine matrix arithmetic.

3. Testing, Equality of Conditional Covariance Matrices

3.1 Summary.

Rewrite the hypothesis \( H_2 \) with all parameters on the left side, thus:

\[
H_2 : \Sigma_{11} - \Sigma_{12} (\Sigma_{22} - I)^{-1} \Sigma_{21} - \psi_{11} + \psi_{12} (\psi_{22} - I)^{-1} \psi_{21} = 0 ,
\]

where the left side is a \( p \times p \) symmetric matrix. The proposed statistic is

\[
J_N = \nu^T U_N^N
\]

as described in the previous section. The components of \( U_N \) are the components \( \pi_{i,j} (1 \leq i \leq j \leq p) \) of the matrix

\[
\Pi = \Sigma_{11} - \Sigma_{12} (\Sigma_{22} - I)^{-1} \Sigma_{21} - \psi_{11} + \psi_{12} (\psi_{22} - I)^{-1} \psi_{21} ,
\]
where \((\Sigma, \Psi)\) is replaced by \((\hat{\Sigma}, \hat{\Psi})\). The components of \(\hat{\Omega}_N\) are given by (2.3), where the parameter values \((\Sigma, \Psi)\) are replaced by their MLE \((\hat{\Sigma}, \hat{\Psi})\), thus:

\[
(3.4) \quad \hat{\omega}(\alpha, \beta) = 2\rho \text{tr} \hat{\Sigma} \hat{\alpha} \hat{\Sigma}^\intercal + 2 \text{tr} \hat{\Psi} \hat{\beta} \hat{\Psi}^\intercal
\]

where \(\rho = \lim_{N \to \infty} N/M\). The quantities \(\hat{\alpha}_\Sigma, \hat{\alpha}_\Psi, \hat{\beta}_\Sigma\) and \(\hat{\beta}_\Psi\) are the values of the partial derivatives \(\alpha_\Sigma, \alpha_\Psi, \beta_\Sigma\) and \(\beta_\Psi\) (as defined in Section 2.2) respectively, when \(\Sigma = \hat{\Sigma}\) and \(\Psi = \hat{\Psi}\), where \(\alpha\) and \(\beta\) each represent one of the \(r = p(p+1)/2\) components of \(\Pi\).

It is implied in the definition of \(U_N\) that \(\hat{\Sigma}_{22-I}\) and \(\hat{\Psi}_{22-I}\) are nonsingular. If the data turn out to be such that either \(\hat{\Sigma}-I\) or \(\hat{\Psi}-I\) is near singularity, one may regard this as an indication that, due to insufficient sample size, the measurement error has swamped out the information in the data estimating \(\Sigma-I\) or \(\Psi-I\). Evidence is presented in Stroud (1968, pp. 80-81) that with standardized achievement test data this is very unlikely to occur for sample sizes greater than 40 when \(p = q = 1\). For the multivariate case, perhaps larger sample sizes than this would be necessary to ensure than the anomaly would not occur.

That the matrix of partial derivatives of the transformation from \((\Sigma, \Psi)\) to \(\Pi\) has rank \(r\) follows from the fact that the \(r \times r\) matrix of partial derivatives based on \(\Sigma_{11}\) alone is the identity.

In order to compute the test statistic (3.2), it remains only to evaluate \(\hat{\Omega}_N\) when \(g(\Sigma, \Psi)\) is determined by (3.3). Let the components of \(\Omega\) be denoted by \(\omega_{ij,k\ell}(1 \leq i \leq j \leq p, 1 \leq k \leq \ell \leq p)\), defined
for each $i,j,k$ and $\ell$ as the quantity denoted in (2.2) by $\omega(\alpha,\beta)$ when $\alpha = \pi_{ij}^L$ and $\beta = \pi_{k\ell}^L$. For each $i,j,k$ and $\ell$, let the corresponding component of $\hat{\Omega}_N$, denoted in (3.4) by $\hat{\omega}(\alpha,\beta)$, be $\hat{\omega}_{ij,k\ell}^L$. Then

\begin{equation}
\omega_{ij,k\ell} = 2\rho \text{tr} \sum_{\Sigma} \Sigma_{ij}^L \Sigma + 2 \text{tr} \Sigma_{ij}^L \Sigma_{k\ell}^L \Sigma_{ij}^L \Sigma_{k\ell}^L \Sigma,
\end{equation}

where $\hat{\omega}_{ij,k\ell}^L$ is the value of this expression when $\Sigma = \hat{\Sigma}$ and $\Psi = \hat{\Psi}$. The purpose of the following section is to evaluate (3.5) explicitly.

3.2 Evaluation of the components of $\Omega$

The partial derivatives $\pi_{ij}^L$ and $\pi_{ij}^L$ are evaluated with the aid of matrix differentials and the associated formulas for products $d(AB) = (dA)B + A(dB)$ and for inverses $d(A^{-1}) = -A^{-1}(dA)A^{-1}$ (see, e.g., Deemer and Olkin (1951), results 5A13, 5A15, and 5B3). If $Y = AXB$ (where all capitals denote matrices), the formula

$$\frac{\partial y_{ij}}{\partial x} = A'B_{ij}B'$$

where $E_{ij}$ consists of a "1" in position $(i,j)$ and zeros elsewhere (Dwyer and MacPhail, 1948) is used to evaluate the matrix derivative. In case $X$ is symmetric, the formula becomes

$$\frac{\partial y_{ij}}{\partial x} = \frac{1}{2}(A'E_{ij}B' + BE_{ij}^T A)$$
where \( \frac{\partial \nu_{ij}}{\partial X} \) is as described in Section 2.2 above.

The differential of \( \Pi \) is now obtained. From (3.3) it follows that

\[
(3.6) \quad d\Pi = d\Sigma_{11}^{-1}d\Sigma_{12}^{-1}(\Sigma_{22}^{-1})^{-1}\Sigma_{21}^{-1}d\Sigma_{12}^{-1}+d\Sigma_{12}^{-1}(\Sigma_{22}^{-1})^{-1}d\Sigma_{22}^{-1}\Sigma_{21}^{-1}
- d\nu_{11}^{-1}d\nu_{12}^{-1}(\Psi_{22}^{-1})^{-1}\Psi_{21}^{-1}d\nu_{12}^{-1}+d\nu_{12}^{-1}(\Psi_{22}^{-1})^{-1}d\nu_{22}^{-1}(\Psi_{22}^{-1})^{-1}\Psi_{21}^{-1}.
\]

The following partial derivatives are obtained:

\[
\begin{align*}
\frac{\partial \nu_{ij}}{\partial \Sigma_{11}} &= \frac{\partial (d\nu_{ij})}{\partial (d\Sigma_{11})} = \frac{1}{2}(E_{ij} + E_{ji}) = \frac{1}{2}(E_{ij} + E_{ji}), \\
\frac{\partial \nu_{ij}}{\partial \nu_{11}} &= \frac{\partial (d\nu_{ij})}{\partial (d\nu_{11})} = -\frac{1}{2}(E_{ij} + E_{ji}), \\
\frac{\partial \nu_{ij}}{\partial \Sigma_{22}} &= \frac{1}{2}(\Sigma_{22}^{-1})^{-1}\Sigma_{21}^{-1}E_{ij}^{-1}\Sigma_{12}^{-1}(\Sigma_{22}^{-1})^{-1} + (\Sigma_{22}^{-1})^{-1}\Sigma_{21}^{-1}E_{ij}^{-1}\Sigma_{12}^{-1}(\Sigma_{22}^{-1})^{-1}
= \frac{1}{2}(\Sigma_{22}^{-1})^{-1}\Sigma_{21}^{-1}(E_{ij} + E_{ji})\Sigma_{12}^{-1}(\Sigma_{22}^{-1})^{-1}, \\
\frac{\partial \nu_{ij}}{\partial \nu_{22}} &= -\frac{1}{2}(\Psi_{22}^{-1})^{-1}\Psi_{21}^{-1}(E_{ij} + E_{ji})\Psi_{12}^{-1}(\Psi_{22}^{-1})^{-1}, \\
\frac{\partial \nu_{ij}}{\partial \Sigma_{12}} &= -(E_{ij} + E_{ji})\Sigma_{12}^{-1}(\Sigma_{22}^{-1})^{-1}, \\
\frac{\partial \nu_{ij}}{\partial \nu_{12}} &= (E_{ij} + E_{ji})\Psi_{12}^{-1}(\Psi_{22}^{-1})^{-1}.
\end{align*}
\]
Remembering that the matrix $\partial x_i^j/\partial \Sigma$ is evaluated with a factor of $\frac{1}{2}$ applied to all off-diagonal components, we may write it, using (3.7), as follows, where the $p \times p$ symmetric matrix $F_{ij}$ is defined by $F_{ij} = \frac{1}{2}(E_{ij} + E_{ji})$:

$$
\begin{pmatrix}
F_{ij} & -F_{ij} \Sigma_{12} (\Sigma_{22} - I)^{-1} \\
-\Sigma_{22} (\Sigma_{22} - I)^{-1} F_{ij} & \Sigma_{22} (\Sigma_{22} - I)^{-1} \Sigma_{21} F_{ij} \Sigma_{12} (\Sigma_{22} - I)^{-1}
\end{pmatrix}
$$

(3.8)

Since $\Sigma$ may be written as

$$
\Sigma = \begin{pmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & (\Sigma_{22} - I) + I
\end{pmatrix}
$$

it follows from (3.8), by straightforward evaluation, that

$$
\operatorname{tr} \pi_{\Sigma}^{i j} \Sigma_{\pi}^{k} \Sigma = \operatorname{tr} F_{ij} A^{(1)} F_{kj} A^{(1)} ,
$$

(3.9)

where

$$
A^{(1)} = \Sigma_{11} - \Sigma_{12} (\Sigma_{22} - I)^{-1} \Sigma_{21} + \Sigma_{12} (\Sigma_{22} - I)^{-2} \Sigma_{21} .
$$

(3.10)

Using $F_{ij} = \frac{1}{2}(E_{ij} + E_{ji})$ and the fact that $A^{(1)}$ is symmetric, we may obtain, using (3.9),

$$
\operatorname{tr} \pi_{\Sigma}^{i j} \Sigma_{\pi}^{k} \Sigma = \frac{1}{2} (a_{i}^{(1)} a_{j}^{(1)} + a_{j}^{(1)} a_{i}^{(1)}) .
$$

(3.11)

14
Define $A^{(2)}$ in terms of $\Psi$ as in (3.10). If we substitute into (3.5) the formula (3.11) and the analogous formula involving $\Psi$ and $A^{(2)}$, the following result is yielded:

\begin{equation}
\omega_{i,j,\ell} = \rho(a_{ik}^{(1)} a_{jl}^{(1)} + a_{ij}^{(1)} a_{jk}^{(1)}) + (a_{ik}^{(2)} a_{jl}^{(2)} + a_{ij}^{(2)} a_{jk}^{(2)})
\end{equation}

\[(1 \leq i \leq j \leq p, \ 1 \leq k \leq \ell \leq p),\]

where $\rho$, defined as the limit of $N/M$ as $N \to \infty$, will in practice be taken equal to $N/M$.

In the case $p=1$, $\Omega$ equals the scalar $\omega_{11,11}$. $U_N = \bar{\Psi}$ is also a scalar, so it is straightforward, on the basis of (3.2), (3.10) and (3.12) to write down the following formula for the statistic $J_N$:

\[
J_N = \frac{\frac{1}{2} [\hat{\Psi}_{11} - \hat{\Psi}_{12} (\hat{\Psi}_{22} - I)^{-1} \hat{\Psi}_{21} + \hat{\Psi}_{12} (\hat{\Psi}_{22} - I)^{-1} \hat{\Psi}_{21} ]^2}{M^{-1} [\hat{\Psi}_{11} - \hat{\Psi}_{12} (\hat{\Psi}_{22} - I)^{-1} \hat{\Psi}_{21} + \hat{\Psi}_{12} (\hat{\Psi}_{22} - I)^{-1} \hat{\Psi}_{21}]^2 \cdot \Psi^{-1} [\hat{\Psi}_{11} - \hat{\Psi}_{12} (\hat{\Psi}_{22} - I)^{-1} \hat{\Psi}_{21} + \hat{\Psi}_{12} (\hat{\Psi}_{22} - I)^{-1} \hat{\Psi}_{21}]^2}.
\]

4. Testing for Equality of Conditional Mean Vectors

4.1 Summary.

The development of this section parallels that of the preceding section, with more details where necessary because the statement of the hypothesis requires two equations rather than one. The problem has been stated in Section 1.2, equations (1.6). The proposed test
The statistic is

\[(4.1)\quad J_N = N \hat{U}_N ^{-1} U_N , \]

where \( U_N = g(X,Y,\hat{\Sigma},\hat{\psi}) \), the quantity \( g(\mu,\nu,\Sigma,\psi) \) is the vector of dimension \( r = pq+p \) whose components are the components of the \( p \times q \) matrix

\[(4.2)\quad \Phi = \Sigma_{12} (\Sigma_{22} - I)^{-1} \Sigma_{12} (\Sigma_{22} - I)^{-1} \psi_{12} (\Sigma_{22} - I)^{-1} \]

and the \( p \times 1 \) vector

\[(4.3)\quad \Lambda = \mu - \Sigma_{12} (\Sigma_{22} - I)^{-1} \hat{\mu} + \Sigma_{12} (\Sigma_{22} - I)^{-1} \psi_1 \]

and the components of \( \hat{\Lambda}_N \) are given by the following formula, obtained from (2.2) by replacing all parameters by the corresponding estimators:

\[(4.4)\quad \hat{\omega}(\alpha,\beta) = \rho \hat{\alpha}_\mu \hat{\beta}_\mu + \hat{\alpha}_\psi \hat{\beta}_\psi + 2 \rho \text{tr} \hat{\alpha}_\Sigma \hat{\beta}_\Sigma + 2 \text{tr} \hat{\alpha}_\psi \hat{\beta}_\psi \]

A verification that the matrix of partial derivatives of the transformation from \( (\mu,\nu,\Sigma,\psi) \) to \( (\Phi,\Lambda) \) has rank \( r \) may be found in an appendix at the end of the paper.

The \( pq+p \) components of \( U_N \) are the components \( \phi_{ij} \) \( (1 \leq i \leq p, 1 \leq j \leq q) \) of the matrix \( \Phi \) and the components \( \lambda_i \) \( (1 \leq i \leq p) \) of the vector \( \Lambda \) when \( (\mu,\nu,\Sigma,\psi) \) is replaced by its estimator \( (X,Y,\hat{\Sigma},\hat{\psi}) \).
(Here, as in §3.1, it is assumed that $\hat{\Sigma}_{22}^{-1}$ and $\hat{\Psi}_{22}^{-1}$ are non-
singular.) In order to compute the test statistic (4.1), it remains
only to evaluate $\hat{\Omega}$ when $g(\mu, \nu, \Sigma, \Psi)$ is determined by (4.2) and
(4.3). As in the problem of the preceding section, the components of
$\Omega$ are evaluated, so that the components of $\hat{\Omega}$ may be obtained there-
from by substituting the estimators in place of the parameters.

Each component of $\Omega$, as given by (2.2), is one of three types,
according to whether $\alpha$ and $\beta$ are:

(i) both components of $\Phi$

(ii) both components of $\Lambda$

(iii) one a component of $\Phi$ and the other of $\Lambda$.

Let the components of $\Omega$ of these types be denoted, respectively, by
$\omega_{ij, kl} (i, k = 1, \ldots, p; j, l = 1, \ldots, q)$, $\omega_{i, k} (i, k = 1, \ldots, p)$, and

$\omega_{i, kl} (i, k = 1, \ldots, p; l = 1, \ldots, q)$. The general formulas (2.2) may be
rewritten as follows.

\begin{equation}
(4.5) \quad \omega_{ij, kl} = \rho (\varphi_{ij}^{''})^{'} \Sigma_{\mu}^{\varphi_{ij}^{''} \varphi_{ij}^{''}} \varphi_{ij}^{''} + 2 \varphi_{ij}^{''} \varphi_{ij}^{''} \Sigma_{\mu}^{\varphi_{ij}^{''} \varphi_{ij}^{''}} + 2 \varphi_{ij}^{''} \varphi_{ij}^{''} \Sigma_{\mu}^{\varphi_{ij}^{''} \varphi_{ij}^{''}} + 2 \varphi_{ij}^{''} \varphi_{ij}^{''} \Sigma_{\mu}^{\varphi_{ij}^{''} \varphi_{ij}^{''}} ,
\end{equation}

\begin{equation}
(4.6) \quad \omega_{i, k} = \rho (\lambda_{\mu}^{''})^{'} \Sigma_{\mu}^{\lambda_{\mu}^{''} \lambda_{\mu}^{''}} + 2 \lambda_{\mu}^{''} \lambda_{\mu}^{''} \Sigma_{\mu}^{\lambda_{\mu}^{''} \lambda_{\mu}^{''}} + 2 \lambda_{\mu}^{''} \lambda_{\mu}^{''} \Sigma_{\mu}^{\lambda_{\mu}^{''} \lambda_{\mu}^{''}} ,
\end{equation}

\begin{equation}
(4.7) \quad \omega_{i, kl} = \rho (\lambda_{\mu}^{''})^{'} \Sigma_{\mu}^{\lambda_{\mu}^{''} \lambda_{\mu}^{''}} + 2 \lambda_{\mu}^{''} \lambda_{\mu}^{''} \Sigma_{\mu}^{\lambda_{\mu}^{''} \lambda_{\mu}^{''}} + 2 \lambda_{\mu}^{''} \lambda_{\mu}^{''} \Sigma_{\mu}^{\lambda_{\mu}^{''} \lambda_{\mu}^{''}} ,
\end{equation}

where $i, k = 1, \ldots, p; j, l = 1, \ldots, q$. Denote the corresponding components
of $\hat{\Omega}$ by the symbols $\hat{\omega}_{ij, kl}$, $\hat{\omega}_{i, k}$ and $\hat{\omega}_{i, kl}$. The values of these
quantities are given by the right-hand sides of (4.5), (4.6) and (4.7).
respectively, when $\mu = X$, $\nu = Y$, $\Sigma = \hat{\Sigma}$ and $\Psi = \hat{\Psi}$. These right-hand sides are evaluated explicitly in the following three sections.

### 4.2 Evaluation of components of $\Phi$ involving $\Phi$ alone.

The evaluation of the partial derivatives $\Phi_{\Sigma}^{ij}$ and $\Phi_{\Psi}^{ij}$ is carried out as in §3.2, by first obtaining a formula for the differentials, in which the variables are now related in a linear way:

\[(4.8)\quad \frac{\partial \Phi}{\partial \Sigma_{12}} = \frac{\partial (\Phi_{ij}^{ij})}{\partial (\Sigma_{22}^{ij} - I)} = E_{ij} (\Sigma_{22} - I)^{-1} (i=1, \ldots, p; j=1, \ldots, q),\]

\[+ \psi_{12}^{ij} (\psi_{22}^{ij} - I)^{-1} d\psi_{22}^{ij} (\psi_{22}^{ij} - I)^{-1} .\]

From this it is straightforward to obtain

\[\frac{\partial \Phi^{ij}}{\partial \Sigma_{22}} = \frac{\partial (\Phi^{ij})}{\partial (\Sigma_{22}^{ij} - I)} = -\frac{1}{2} [ (\Sigma_{22} - I)^{-1} (\Sigma_{22} - I)^{-1} + (\Sigma_{22} - I)^{-1} (\Sigma_{22} - I)^{-1} (\Sigma_{22} - I)^{-1} ],\]

with similar formulas with changed sign for $\partial \Phi^{ij}/\partial \Psi_{12}$ and $\partial \Phi^{ij}/\partial \Psi_{22}$.

Note that the off-diagonal components of $\partial \Phi^{ij}/\partial \Sigma_{22}$ and $\partial \Phi^{ij}/\partial \Psi_{22}$, but not $\partial \Phi^{ij}/\partial \Sigma_{12}$ and $\partial \Phi^{ij}/\partial \Psi_{22}$, include a factor of $\frac{1}{2}$. Finally

\[(4.9)\quad \Phi_{\Sigma}^{ij} = \begin{pmatrix} 0 & \frac{1}{2} E_{ij} (\Sigma_{22} - I)^{-1} \\ \frac{1}{2} (\Sigma_{22} - I)^{-1} E_{ij} & -\frac{1}{2} (\Sigma_{22} - I)^{-1} (\Sigma_{22} - I)^{-1} (\Sigma_{22} - I)^{-1} \end{pmatrix} .\]
By straightforward calculation and noting \((\Sigma_{22} - I)^{-1}\Sigma_{22} (\Sigma_{22} - I)^{-1} = \Sigma_{22} (\Sigma_{22} - I)^{-1} + (\Sigma_{22} - I)^{-2}\), we obtain from (4.9):

\[
\text{tr} \varphi^{ij}_{\Sigma}\Sigma^{kl}_{\Sigma} \Sigma = \frac{1}{2} \text{tr} E_{ij} \left[ (\Sigma_{22} - I)^{-1} + (\Sigma_{22} - I)^{-2} \right] E^l_k \left[ (\Sigma_{11} - \Sigma_{12}) (\Sigma_{22} - I)^{-1} \Sigma_{21} \right. \\
+ \left. (\Sigma_{22} - I)^{-2} \Sigma_{21} \right] \\
+ \frac{1}{2} \text{tr} E_{ij} (\Sigma_{22} - I)^{-2} \Sigma_{21} E^l_k (\Sigma_{22} - I)^{-2} \Sigma_{21} \\
= \frac{1}{2} \text{tr} E_{ij} C^{(1)}_{k} A^{(1)}(1) + \frac{1}{2} \text{tr} E_{ij} B^{(1)}_{k} B^{(1)}(1),
\]

where, for \(\alpha=1,2\), the matrices \(A^{(\alpha)}\) are defined by (3.10), and the following notation is introduced:

\[(4.10) \quad B^{(1)} = (\Sigma_{22} - I)^{-2} \Sigma_{21}, \quad C^{(1)} = (\Sigma_{22} - I)^{-1} + (\Sigma_{22} - I)^{-2},\]

with \(B^{(2)}\) and \(C^{(2)}\) defined analogously for \(\Psi\). By straightforward evaluation, we obtain

\[(4.11) \quad \text{tr} \varphi^{ij}_{\Sigma}\Sigma^{kl}_{\Sigma} \Sigma = \frac{1}{2} (c^{(1)}_{kj} a^{(1)}_{ki} + b^{(1)}_{kj} b^{(1)}_{ki}),\]

and a similar formula involving \(\Psi\).

To evaluate (4.5), notice from the form of (4.2) that \(\phi\) does not involve \(\mu\) or \(\nu\). By (4.11), (4.5) becomes
\[ (4.12) \quad \omega_{ij,kl} = \rho (c^1_{ij} a^1_{ki} + b^1_{jk} b^1_{li}) + \rho (c^2_{ij} a^2_{ki} + b^2_{jk} b^2_{li}) \]

\[ (i, k = 1, \ldots, p; j, l = 1, \ldots, q) . \]

4.3 Evaluation of components of $\Omega$ involving $\Lambda$ alone.

In this section the right hand side of (4.6) is evaluated. The method is similar to that used in the preceding section. From (4.3) the following result is obtained for the differential $d\Lambda$:

\[ (4.13) \quad d\Lambda = d\bar{\mu} \Sigma_{22}^{-1} - d\bar{\mu} - d\Sigma_{12} (\Sigma_{22}^{-1})^{-1} d\bar{\mu} + \Sigma_{12} (\Sigma_{22}^{-1})^{-1} d\bar{\mu}
\]
\[ - d \bar{\nu} + \Psi_{12} (\Psi^{-1}_{22})^{-1} d \bar{\nu} + d \Psi_{12} (\Psi^{-1}_{22})^{-1} d \bar{\nu} - \Psi_{12} (\Psi^{-1}_{22})^{-1} d \bar{\nu} . \]

Using the fact that when $E_{ij}$ contains only one column it may be written as $e_i$, defined as a column vector with "1" in the $i$-th position and zeros elsewhere, one may obtain the following derivatives from (4.13):

\[ \frac{\partial \lambda^i}{\partial \mu} = \frac{\partial (d\lambda^i)}{\partial (d\mu)} = e_i , \quad \frac{\partial \lambda^i}{\partial \bar{\mu}} = -(\Sigma_{22}^{-1})^{-1} \Sigma_{21} e_i , \]

\[ \frac{\partial \lambda^i}{\partial \Sigma_{12}} = -e_i \bar{\mu}' (\Sigma_{22}^{-1})^{-1} , \quad \frac{\partial \lambda^i}{\partial \Sigma_{22}} = \frac{1}{2} [ (\Sigma_{22}^{-1})^{-1} (\Sigma_{21} e_i \bar{\mu}' + \bar{\mu} e_i \Sigma_{12}) (\Sigma_{22}^{-1})^{-1} ]; \]

from which

\[ (4.14) \quad \lambda^i_{\Sigma} = \begin{pmatrix} 0 & -\frac{1}{2} e_i \bar{\mu}' (\Sigma_{22}^{-1})^{-1} \\ -\frac{1}{2} (\Sigma_{22}^{-1})^{-1} \bar{\mu} e_i & \frac{1}{2} [ (\Sigma_{22}^{-1})^{-1} (\Sigma_{21} e_i \bar{\mu}' + \bar{\mu} e_i \Sigma_{12}) (\Sigma_{22}^{-1})^{-1} ] \end{pmatrix} . \]
The corresponding expression for $\alpha^i / \alpha^a$ arises directly from (4.13):

\begin{equation}
\lambda^i_\mu = \begin{pmatrix} e_i \\ - (\Sigma_{22} - I)^{-1} \Sigma_{21} e_i \end{pmatrix}.
\end{equation}

(4.15)

It may be noted that formulas (4.9) and (4.14) are identical, except that in (4.14) the matrix \((-e_i B')\) replaces \(E_{ij}\) of (4.9). Thus, in the same manner that (4.11) was derived from (4.9), (4.14) reduces to

\begin{equation}
\text{tr} \frac{i}{2} \Sigma \lambda^k \lambda^l = \frac{1}{2} \text{tr} e_i B'(1) e_k B(1) + \frac{1}{2} \text{tr} e_i B'(1) e_k B(1)
\end{equation}

\begin{equation}
= \frac{1}{2} (B' C (1) B')_{ik} a_{ik} + \frac{1}{2} (B' B(1))_{ik} (B' B(1))_{ik}.
\end{equation}

(4.16)

By further calculation it may be seen that relation (4.15) implies

\begin{equation}
(\lambda^i_\mu)^{,\Sigma^k} = e_i^A A_{ik} = a_{ik}^{(1)}.
\end{equation}

(4.17)

Similarly

\begin{equation}
\text{tr} \frac{i}{2} \lambda^k \lambda^l \lambda^{(2)} = \frac{1}{2} (B' C (2) B')_{ik} a_{ik}^{(2)} + \frac{1}{2} (B' B(2))_{ik} (B' B(2))_{ik}
\end{equation}

(4.18)

and

\begin{equation}
(\lambda^i_\nu)^{,\nu} = a_{ik}^{(2)}.
\end{equation}

(4.19)
Substitution of (4.16), (4.17), (4.18) and (4.19) into (4.6) yields the following formula:

\[
\omega_{i,k} = \rho \left[ (\bar{\nu}' C^{(1)} \bar{\nu}^{1}) a_{ik}^{(1)} + (\bar{\nu}'' B^{(1)})_{i} (\bar{\nu}'' B^{(1)})_{k} \right] \\
+ \left[ (\bar{\nu}' C^{(2)} \bar{\nu}^{1}) a_{ik}^{(2)} + (\bar{\nu}'' B^{(2)})_{i} (\bar{\nu}'' B^{(2)})_{k} \right] \\
\]  

\((i,k=1,...,p)\).

4.4 Evaluation of components of \( \Omega \) involving both \( \Phi \) and \( \Lambda \).

The right hand side of (4.7) is now evaluated. Using (4.14) and (4.9), it is straightforward to obtain, in the same manner as in the preceding two subsections,

\[
\frac{1}{2} \text{tr}_{(\Sigma)} \sum_{\Sigma} \phi^{k\ell}_{\Sigma} = -\frac{1}{2} \text{tr}_{(\Sigma)} e_{i}(\bar{\nu}' C^{(1)}) E_{k\ell} A^{(1)} - \frac{1}{2} \text{tr}_{(\Sigma)} e_{i}(\bar{\nu}'' B^{(1)}) E_{k\ell} B^{(1)} \\
= -\frac{1}{2} a_{ki}^{(1)} (\bar{\nu}' C^{(1)})_{\ell} - \frac{1}{2} b_{ki}^{(1)} (\bar{\nu}'' B^{(1)})_{\ell} ,
\]

with a similar formula for \( \Psi \). Since \( \Lambda_{\mu} = \Psi_{\nu} = 0 \), one obtains, using (4.7),

\[
\omega_{i,\ell} = -\rho \left[ a_{ik}^{(1)} (\bar{\nu}' C^{(1)})_{\ell} + b_{ki}^{(1)} (\bar{\nu}'' B^{(1)})_{\ell} \right] - \\
[ a_{ki}^{(2)} (\bar{\nu}' C^{(2)})_{\ell} + b_{ki}^{(2)} (\bar{\nu}'' B^{(2)})_{\ell} ] \\
\]  

\((i,\ell=1,...,p; \ell=1,...,q)\).
The components of \( \hat{\Omega}_N \) used in the test statistic (4.1) are obtained from (4.12), (4.20) and (4.22) by replacing the parameters by the corresponding estimators, and, as in the previous chapter, replacing \( \rho \) by \( N/M \).

4.5 Formula for \( \hat{\Omega}_N \) when the predicted variable is univariate.

When \( p=1 \), some simplification is possible in the representation of \( \Omega \) and consequently of \( \hat{\Omega}_N \), through the introduction below of matrices \( \hat{G}(\alpha) (\alpha=1,2) \), defined in terms of \( \hat{A}(\alpha), \hat{B}(\alpha) \) and \( \hat{C}(\alpha) \), which in turn are obtained from \( A(\alpha), B(\alpha) \) and \( C(\alpha) \) ((3.10) and (4.10)) by replacing parameters by their estimators.

Assuming \( p=1 \), write \( \hat{\Omega}_N = (\hat{\Omega}_{ij}) \), \( i,j=1,2 \), where \( \hat{\Omega}_{11} \) is the \( q \times q \) symmetric matrix whose \( (j,\ell) \)-th component is \( \hat{\omega}_{1j,\ell} \) \( (j,\ell=1,\ldots,q) \), \( \hat{\Omega}_{22} = \hat{\omega}_{11} \) is a scalar, and \( \hat{\Omega}_{21} = \hat{\Omega}_{12} \) is a \( 1 \times q \) row vector whose \( \ell \)-th component is \( \hat{\omega}_{1,1\ell} \). Formulas for \( \hat{\Omega}_{11}, \hat{\Omega}_{22} \) and \( \hat{\Omega}_{21} \) are now derived.

It will be convenient to eliminate superscripts, wherever possible, and write

\[
4.23 \quad \hat{a}_\alpha = A(\alpha) = a_{11}, \quad \hat{b}_\alpha = B(\alpha), \quad \hat{c}_\alpha = C(\alpha), \quad \alpha = 1,2.
\]

The \( \hat{b}_\alpha \) are \( q \times 1 \) column vectors, and the \( \hat{c}_\alpha \) are \( q \times q \) symmetric matrices. The following formula is obtained from (4.12) by replacing parameters by their estimators:
\[
(4.24) \quad \hat{\omega}_{1j,1l} = \frac{N}{M} \left[ c_{1j}^{(1)} a_{1j}^{(1)} + b_{1j}^{(1)} b_{1j}^{(1)} \right] + \left[ c_{1j}^{(2)} a_{1j}^{(2)} + b_{1j}^{(2)} b_{1j}^{(2)} \right] 
\]
\[(j,l=1,\ldots,q)\,.
\]

Note that, for \(\alpha=1\) and 2, \(b_{j1}^{(\alpha)} b_{l1}^{(\alpha)}\) is the \((j,l)\)-th component of \(\hat{B}_{\alpha} \hat{B}'_{\alpha}\). Accordingly, define

\[
(4.25) \quad \hat{g}_\alpha = \hat{g}^{(\alpha)} = \hat{a}_\alpha \hat{c}_\alpha + \hat{b}_\alpha \hat{b}_\alpha' \quad (\alpha=1,2)\,.
\]

where the scalars \(\hat{a}_\alpha\) are as in (4.23). It follows from (4.24) that

\[
(4.26) \quad \hat{\omega}_{11} = \frac{N}{M} \hat{g}_1 + \hat{g}_2 = M^{-1}(\hat{G}_1 + \hat{M}_2)\,.
\]

The formula for \(\hat{\omega}_{22}\) is obtained from (4.20).

\[
(4.27) \quad \hat{\omega}_{22} = \hat{\omega}_{1,1} = \frac{N}{M} \left[ a_1 (\hat{C}_1^T \hat{X} + 1) + \hat{B}_1 \hat{B}'_1 \hat{X} \right] + \left[ a_2 (\hat{Y}'_1 \hat{C}_2^T \hat{Y} + 1) + \hat{Y}'_2 \hat{B}_2 \hat{B}'_2 \hat{Y} \right]
\]
\[
\quad = \frac{N}{M} (\hat{a}_1 + \hat{X}' \hat{G}_1) + (\hat{a}_2 + \hat{Y}'_1 \hat{G}_2)\,.
\]

Finally, the \(l\)-th component of the row vector \(\hat{\omega}_{21}\) is given by (4.22) to be

\[
\hat{\omega}_{1,1l} = -\frac{N}{M} \left[ a_1 (\hat{X}' \hat{C}^{(1)}_1)_l + b_1^{(1)} (X'B^{(1)}_1)_l \right] - \left[ a_2 (\hat{Y}' \hat{C}^{(2)}_2)_l + b_2^{(2)} (Y'B^{(2)}_2)_l \right] ,
\]

where the subscript "\(l\)" denotes the \(l\)-th component of a vector.
Consequently,

\[ \hat{\omega}_{1,1\ell} = -\frac{N}{M} \hat{X}' [\hat{\mathcal{A}}'_{\ell} + (\hat{\mathcal{B}}'_{\ell})'] - \hat{Y}' [\hat{\mathcal{A}}_{2\ell} + (\hat{\mathcal{B}}_{2\ell})'] , \]

where the subscript "\( \cdot \ell \)" denotes the \( \ell \)-th column of a matrix, whence

\[ \hat{\omega}_{1,1\ell} = -\frac{N}{M} \hat{X}' \hat{G}^{(1)} - \hat{Y}' \hat{G}^{(2)} . \]

Thus the formula for the row-vector \( \hat{\Omega}_{21} \) is

\[ \hat{\Omega}_{21} = -\frac{1}{M} (N \hat{X}' \hat{G}_1 + M \hat{Y}' \hat{G}_2) . \]  

Combining (4.26), (4.27) and (4.28), the formula for \( \hat{\Omega}_N \) in the case \( p=1 \) is

\[
\hat{\Omega}_N = \frac{1}{M} \begin{pmatrix}
NG_1 + MG_2 & -NG_1 \hat{X} - MG_2 \hat{Y} \\
-NX' \hat{G}_1 - M\hat{Y}' \hat{G}_2 & N(a_1 \hat{X}' \hat{G}_1 \hat{X}) + M(a_2 \hat{Y}' \hat{G}_2 \hat{Y})
\end{pmatrix},
\]

where the \( \hat{G}_\alpha \) are defined by (4.25) in terms of the \( \hat{a}_\alpha', \hat{b}_\alpha \) and \( \hat{c}_\alpha' \), which in turn are defined by (4.23), (3.10) and (4.10).

For a further simplification in case \( q=1 \) also holds, see Stroud (1972).
5. **A Comparison of the Application of the Procedures with that of the Univariate Versions**

An application of the univariate \((p=q=1)\) version of the formulas to educational testing data in Portland, Oregon has been reported in Stroud (1972). In this section we report the results of applying the multivariate procedures described in the preceding two sections of this paper to the same data, and we restate the univariate results for comparison purposes.

In the application, \(\hat{X}\) and \(\check{Y}\) represent standardized TAP (Tests of Academic Progress) scores administered to grade 11 students throughout the city of Portland, and \(\bar{X}\) and \(\bar{Y}\) represent grade 9 ITED (Iowa Tests of Educational Development) scores of the same students. We have \(p=3\) and \(q=5\), where the components of \(\hat{X}\) and of \(\check{Y}\) are three TAP subtest scores (Composition, Reading and Mathematics, in that order) and the components of \(\bar{X}\) and \(\bar{Y}\) are the ITED subtest scores in Social Concepts, Correctness of Expression, Quantitative Thinking, Reading (Social Studies) and Reading (Natural Sciences).

First the procedures are applied using as \(\hat{X}\) and \(\bar{X}\) the mean scores of all the girls in a specified school (School 1) and for \(\check{Y}\) and \(\bar{Y}\) the scores of the girls in the rest of the city (Schools 2-12) taken together. Secondly, the same procedure is repeated for the boys. In the third and final application, the girls in the twelve schools \((\hat{X},\check{X})\) are compared with the boys \((\check{Y},\bar{Y})\). Table 1 shows the value of the test statistics \(\mathbf{J}_N\) for comparing (i) conditional covariance matrices and (ii) conditional mean vectors for
both the multivariate data just described and the univariate case utilizing composite scores (averages over the subtests). Beside the value of $J_N$ is given the corresponding value of $P$, the inverse of the approximating chi-square cumulative distribution function ($= 1$ - significance level attained).

The standard errors of measurement for the subtests were taken from the administrators manuals, Scannell (1965) and Science Research Associates (1963). The subtest measurement errors were assumed to be uncorrelated with each other.

It is seen from the significance levels in Table 1 that the results of the univariate and multivariate applications do not completely correspond with each other. Notice that in the school-versus-school comparisons (boys and girls separately) the multivariate test for covariance matrices reveals greater significance than the corresponding univariate test, but the multivariate test for mean vectors shows less significance. In the boy-versus-girl comparison, however, the pattern is reversed.

In trying to account for these phenomena, we may note from Table 2, where the six 8-dimensional mean vectors are tabulated, that subtest scores for School 1 are consistently lower than those of the 2-12 group. However, if we compare boys and girls, we find that boys do better in some subtests (notably the quantitative) whereas girls do better in others (e.g., composition). Thus the composite scores used in the $p=q=1$ analysis are appropriate for comparing schools, but not for comparing sexes because the sex-related differences will tend to be reduced in the averaging of subtest scores. In the school-versus-school comparison
of mean vectors, most of the meaningful variation has been recorded in the composite score analysis; so that for example a chi-square of 11.92 on 2 degrees of freedom is registered as being more significant than a chi-square of 29.94 on 18 degrees of freedom, even though the difference of the chi-squares (18.02) slightly exceeds the difference in the degrees of freedom.

In the boy-versus-girl comparison, on the other hand, the multivariate analysis of mean vectors gives a more impressive result than the univariate for the reason (mentioned above) that the simple average is far from being the best linear combination demonstrating differences between boys and girls.

The interpretation of the results regarding conditional variances and covariance matrices is more difficult. The main factor is probably that residual variances based on predicting a single variable by a single variable cannot be expected to resemble too closely a $3 \times 3$ residual covariance matrix based on five predictors. One would guess that, with the school-versus-school comparisons, there are discrepancies in the residual covariance matrices which are washed out when we look at just the residual variance of the composite score. Regarding the boy-versus-girl comparison, an examination of the data has revealed that the difference between residual variances in the univariate analysis exceeds (in relative terms as well as absolute) the difference in residual variances of any of the three subtest scores in the multivariate results. This may very well be related to the interaction between sex and subtest content, but the pattern appears too complicated to give a detailed account of here.
In conclusion, it would appear that the univariate and multivariate analyses taken together are more informative than either one would be alone. Although the techniques used to study the above data have been derived from the theory of inference, the author has the distinct impression that an honest attempt to get as good a feel for the data as possible is more fruitful than the making of statistical decisions such as accepting or rejecting hypotheses. This supports the view of Dempster (1969) that the data-analytic approach to multivariate problems is often more sensible than a "solution" based on inference.

6. **Appendix**

Verification that the rank of the transformation from \((\mu, \nu, \Sigma, \Psi)\) to \((\Phi, \Lambda)\) is \((pq + p)\).

As was remarked in Section 4.1, the \(r \times k\) matrix of partial derivatives of \((\Phi, \Lambda)\) with respect to \((\mu, \nu, \Sigma, \Psi)\) must have rank of \(r = pq + p\) in order for the theory described in §2.2 to apply. We now verify that this is in fact the case.

It will suffice to show that the \(r \times r\) matrix of partial derivatives of \((\Phi, \Lambda)\) with respect to \((\Sigma_{12}, \mu)\) is nonsingular. Consider \(\nu, \Psi, \bar{\mu}, \Sigma_{11}\) and \(\Sigma_{22}\) to be held constant; it is sufficient to show that the Jacobian of the transformation from \((\Sigma_{12}, \mu)\) to \((\Phi, \Lambda)\) is non-zero.

Consider the matrix \(M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}\) where \(A\) is a \(pq \times pq\) matrix containing the partial derivatives of the components of \(\Phi\) with respect to the components of \(\Sigma_{12}\), denoted by \(\partial \Phi/\partial \Sigma_{12}\); similarly \(B = \partial \Lambda/\partial \Sigma_{12}\).
(pq x p), C = ∂M/∂μ (p x pq), and D = ∂M/∂μ (p x p). It is clear that, with the natural ordering of the components, D = I; it is also clear that C = 0. Hence |M| = |A| · |D - CA⁻¹B| = |A|, which is the Jacobian of the transformation from Σ₁₂ to Σ₁₂(Σ₂₂⁻¹I)⁻¹, whose magnitude is equal to |(Σ₂₂⁻¹I)⁻¹|ᵖ (see Deemer and Olkin (1951), Theorem 3.5).

7. **Acknowledgements**

The author is grateful for the encouragement and guidance of Professor Ingram Olkin.

Most of the results in this paper appeared in the author's Ph.D. dissertation at Stanford University, and were supported in part by the National Science Foundation, grant no. 6681.

The author also wishes to thank the Portland School District Administration for the use of their data.
**TABLE 1**

Values of the Test Statistic $J_N$ and corresponding probability point $P$ based on analysis of Composite Scores ($p=q=1$) and Vector Scores ($p=3, q=5$) (for sample sizes see Table 2).

<table>
<thead>
<tr>
<th></th>
<th>Comparing Conditional Covariance Matrices</th>
<th>Comparing Conditional Mean Vectors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Composite Scores</td>
<td>Vector Scores</td>
</tr>
<tr>
<td></td>
<td>$J_N$</td>
<td>$P$</td>
</tr>
<tr>
<td>Schools 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>versus 2-12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Girls)</td>
<td>1.04</td>
<td>.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schools 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>versus 2-12</td>
<td>1.00</td>
<td>.68</td>
</tr>
<tr>
<td>(Boys)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls</td>
<td></td>
<td></td>
</tr>
<tr>
<td>versus</td>
<td>6.89</td>
<td>.9913</td>
</tr>
<tr>
<td>Boys</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degrees of</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Freedom for $J_N$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

31
### TABLE 2

**SAMPLE SIZES AND MEAN SUBTEST SCORES**

(Scaled to Unit Error of Measurement)

<table>
<thead>
<tr>
<th></th>
<th>Girls 1</th>
<th>Girls 2-12</th>
<th>Boys 1</th>
<th>Boys 2-12</th>
<th>All Girls</th>
<th>All Boys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>195</td>
<td>1779</td>
<td>167</td>
<td>1794</td>
<td>1974</td>
<td>1921</td>
</tr>
<tr>
<td>Grade 9 ITED:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading (SS)</td>
<td>14.97</td>
<td>15.37</td>
<td>14.91</td>
<td>15.28</td>
<td>15.33</td>
<td>15.25</td>
</tr>
<tr>
<td>Reading (NS)</td>
<td>14.02</td>
<td>14.40</td>
<td>14.08</td>
<td>14.51</td>
<td>14.36</td>
<td>14.47</td>
</tr>
<tr>
<td>Grade 11 TAP:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Composition</td>
<td>15.02</td>
<td>15.91</td>
<td>13.57</td>
<td>14.28</td>
<td>15.82</td>
<td>14.21</td>
</tr>
<tr>
<td>Reading</td>
<td>18.06</td>
<td>19.01</td>
<td>18.23</td>
<td>18.75</td>
<td>18.92</td>
<td>18.71</td>
</tr>
<tr>
<td>Mathematics</td>
<td>17.98</td>
<td>18.90</td>
<td>19.62</td>
<td>20.50</td>
<td>18.81</td>
<td>20.42</td>
</tr>
</tbody>
</table>

32
REFERENCES


