The Canadian Journal of Statistics
Vol. 27, No. 2, 1999, Pages 395–407
La revue canadienne de statistique

Confidence estimation of a normal mean vector with incomplete data

K. KRISHNAMOORTHY and Maruthy K. PANNALA

University of Southwestern Louisiana and *United Guaranty Corporation*

Key words and phrases: Bonferroni intervals, block-monotone pattern, likelihood-ratio test, missing data, multiple comparison, repeated measurements, Scheffé intervals.
AMS 1991 subject classifications: Primary 62F25; secondary 62E15, 62C10.

ABSTRACT

The problem of confidence estimation of a normal mean vector when data on different subsets of response variables are missing is considered. A simple approximate confidence region is proposed when the data matrix is of monotone pattern. Simultaneous inferential procedures based on Scheffé's method and Bonferroni's method are outlined. Further, applications of the results to a repeated measurements model are given. The results are illustrated using a practical example.

RÉSUMÉ

Les auteurs s'intéressent à la construction de régions de confiance pour le vecteur moyenne d'une population normale dans la situation où certaines données sont manquantes pour des sous-ensembles particuliers de variables-réponses. Ils proposent une solution simple, quoiqu'approximative, à ce problème dans le cas où la matrice des observations possède une structure monotone. Ils présentent en outre des procédures d'inférence simultanée s'appuyant sur les méthodes de Scheffé et de Bonferroni. Ces résultats sont appliqués à un modèle de mesures répétées et illustrés au moyen d'un exemple concret.

1. INTRODUCTION

Statistical inference based on missing or extra data is one of the important applied problems because of its common occurrence in practice. Incomplete data arise, for example, during data recording, when some of the variables to be measured are too expensive, when responses are related to some sensitive questions or when the experiment is run on a group of individuals over a period of time as in clinical studies. To ignore the process that causes missing data while making inferences about the parameter of interest, it is assumed in this paper that data are missing completely at random. That is, missingness is completely independent of either the nature or the values of the variables in the data set under study. Formal definitions and tests for missing at random can be found in Rubin (1976) and Little (1988, 1995).

In this paper, we consider the problem of confidence estimation of a p-variate normal mean vector $\boldsymbol{\mu}$ based on incomplete data sets of triangular or monotone pattern. A monotone pattern is convenient for deriving the maximum-likelihood estimators (MLEs) and the likelihood-ratio tests (LRTs) for the mean $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$. Anderson (1957) gives an elegant approach to deriving MLEs and presents them in explicit form for the special case of a monotone data set and for data with some other patterns. Following his approach, Bhargava (1962) derived LRTs and their approximate null distributions for several problems. Mehta and Gurland (1969), Morrison (1973), and Naik (1975) considered the problem of testing equality of the components of $\boldsymbol{\mu}$ for the bivariate case.

Although several papers addressed the problems of testing for various patterns of missing data, an important problem of confidence estimation has received very little attention in the literature. The LRT statistic and its approximate null distribution are not useful for developing a confidence set for $\boldsymbol{\mu}$ in a simple-to-use form, and it is difficult to check whether such a confidence set is an ellipsoidal region. Another possible approach is to impute the missing data and then use standard methods available for complete data analysis. For the univariate case, interval estimation based on multiple imputation is given in Rubin and Schenker (1986). Li *et al.* (1991) give large-sample confidence estimation of a mean vector using multiple imputation. For a good exposition of imputation procedures and the validity of imputation inference in practice, we refer the reader to Meng (1994) and the references therein. The results based on the imputation method are satisfactory even for nonnormal data and valid when the data are missing at random. However, if normality is assumed, the computational task involved in the imputation procedure can be avoided and simple expressions for confidence regions for various patterns of data can be obtained. Furthermore, as shown in Section 2.2 of this paper, the results under the normality assumption are very satisfactory even for small samples.

In the following section we derive first an estimator of the covariance of the MLE $\hat{\boldsymbol{\mu}}$, denoted by $\widehat{Cov}\ \hat{\mu}$, for the data with missing observations from a single subset of components. We then approximate the distribution of the quadratic form $(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})^{\mathsf{T}}(\widehat{Cov}\ \hat{\boldsymbol{\mu}})^{-1}(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})$ by a constant times an F-distribution using the method of moments. Simulation studies and a numerical study indicate that the approximation is very satisfactory even for small samples. Numerical comparisons of the expected volumes of the proposed confidence region with those of the one based on complete data (which ignores the extra data) for various sample sizes indicate that the former has smaller volumes than the latter. Further, numerical study indicates that the power differences between the test based on the developed confidence region and the LRT are very small. A simultaneous inferential procedure based on the confidence region is outlined, with applications to a repeated-measurements model. Bonferroni intervals for the components of $\boldsymbol{\mu}$ are also given. In Section 3, these results are extended to a general monotone setup. In Section 4 the results are illustrated using an example. Finally, in Section 5, we point out the applicability of our method to other patterns of data and propose some future work in this area.

2. CONFIDENCE ESTIMATION OF $\boldsymbol{\mu}$

2.1. Missing Observations from a Subset of Components.

Let $\mathbf{X}_1,\ldots,\mathbf{X}_n$ be independent observations from $\mathrm{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$, and $\mathbf{Y}_1,\ldots,\mathbf{Y}_m$ be independent observations from $\mathrm{N}_{p_1}(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11})$, where $\boldsymbol{\mu}=(\boldsymbol{\mu}_1^{\mathsf{T}},\boldsymbol{\mu}_2^{\mathsf{T}})^{\mathsf{T}}$ and $\boldsymbol{\Sigma}_{ij}$ denotes the (i,j)th $p_i\times p_j$ submatrix of $\boldsymbol{\Sigma}$, $i,j=1,2$. It is assumed that the $\mathbf{X}_i$'s and $\mathbf{Y}_i$'s are independent. That is, we have m additional observations available on the first p_1 components, or equivalently m observations are missing on the last p_2 components. Let $\bar{\mathbf{X}}$ denote the sample mean vector based on the $\mathbf{X}$-observations, and define $\bar{\mathbf{Y}}$ similarly. Let $\mathbf{S}$ denote the sums-of-squares-and-products matrix based on the $\mathbf{X}$-observations, and let $\mathbf{V}$ denote the sums-of-squares-and-products matrix based on the $\mathbf{Y}$-observations. Note that $\bar{\mathbf{X}}$, $\bar{\mathbf{Y}}$, $\mathbf{S}$

and $\mathbf{V}$ are all statistically independent with

$$\bar{\mathbf{X}} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n), \qquad \bar{\mathbf{Y}} \sim \mathrm{N}_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}/m),$$
$$\mathbf{S} \sim W_p(n-1, \boldsymbol{\Sigma}) \quad \text{and} \quad \mathbf{V} \sim W_{p_1}(m-1, \boldsymbol{\Sigma}_{11}). \tag{2.1}$$

Let $\mathbf{S}_{ij} : p_i \times p_j$ denote the (i,j)th partitioned matrix of $\mathbf{S}$, and write $\bar{\mathbf{X}} = (\bar{\mathbf{X}}_1^{\mathsf{T}}, \bar{\mathbf{X}}_2^{\mathsf{T}})^{\mathsf{T}}$, so that $\bar{\mathbf{X}}_1$ is $p_1 \times 1$. The MLEs of the parameters are given by (see, for example, Anderson 1957)

$$\hat{\boldsymbol{\mu}}_1 = \frac{n\bar{\mathbf{X}}_1 + m\bar{\mathbf{Y}}}{m+n}, \qquad \hat{\boldsymbol{\mu}}_2 = \bar{\mathbf{X}}_2 + \frac{m\mathbf{S}_{21}\mathbf{S}_{11}^{-1}(\bar{\mathbf{Y}} - \bar{\mathbf{X}}_1)}{m+n},$$
$$\hat{\boldsymbol{\Sigma}}_{11} = \frac{\mathbf{S}_{11} + \mathbf{V} + mn(\bar{\mathbf{X}}_1 - \bar{\mathbf{Y}})(\bar{\mathbf{X}}_1 - \bar{\mathbf{Y}})^{\mathsf{T}}/(m+n)}{m+n}, \qquad \hat{\boldsymbol{\Sigma}}_{12} = \hat{\boldsymbol{\Sigma}}_{11}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$$

and

$$\hat{\boldsymbol{\Sigma}}_{22} = \frac{\mathbf{S}_{2.1}}{n} + \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\hat{\boldsymbol{\Sigma}}_{11}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}, \tag{2.2}$$

where $\mathbf{S}_{2.1} = \mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$. It is easy to show that

$$Cov\ \hat{\boldsymbol{\mu}}_1 = \frac{\boldsymbol{\Sigma}_{11}}{m+n}, \qquad Cov\,(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2) = \frac{\boldsymbol{\Sigma}_{12}}{m+n}$$

and

$$Cov\ \hat{\boldsymbol{\mu}}_2 = \frac{\boldsymbol{\Sigma}_{22}}{n} + \frac{c\mathcal{E}\,(\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\boldsymbol{\Sigma}_{11}\mathbf{S}_{11}^{-1}\mathbf{S}_{12} - \boldsymbol{\Sigma}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\boldsymbol{\Sigma}_{12})}{n}, \tag{2.3}$$

where $c = m/(m+n)$. To estimate $Cov\ \hat{\boldsymbol{\mu}}$, we replace the unknown parameters in (2.3) by their MLEs; after replacing and some simplification, we get

$$\widehat{Cov}\ \hat{\boldsymbol{\mu}} = (m+n)^{-1}\begin{pmatrix} \hat{\boldsymbol{\Sigma}}_{11} & \hat{\boldsymbol{\Sigma}}_{12} \\ \hat{\boldsymbol{\Sigma}}_{21} & (m+n)\mathbf{S}_{2.1}/n^2 + \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\hat{\boldsymbol{\Sigma}}_{11}\mathbf{S}_{11}^{-1}\mathbf{S}_{12} \end{pmatrix}.$$

Although the MLEs are valid under the assumption that the data are missing at random, a referee has pointed out that the expression for the $\mathcal{V}ar\ \hat{\mu}$ given in the (2,2) entry of the above matrix is valid only if the data are missing completely at random; for details, see Little and Rubin [1987, Equations (6.13), (6.14)].

Using the identity that, for a vector $\mathbf{X}$ and a nonsingular matrix $\mathbf{A}$,

$$\mathbf{X}^{\mathsf{T}}\mathbf{A}^{-1}\mathbf{X} = \mathbf{X}_1^{\mathsf{T}}\mathbf{A}_{11}^{-1}\mathbf{X}_1 + \mathbf{X}_{2.1}^{\mathsf{T}}\mathbf{A}_{2.1}^{-1}\mathbf{X}_{2.1}, \tag{2.4}$$

where $\mathbf{X}_{2.1} = \mathbf{X}_2 - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{X}_1$ and $\mathbf{A}_{2.1} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$, it is easy to show that

$$\begin{aligned} Q &= (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^{\mathsf{T}}(\widehat{Cov}\ \hat{\boldsymbol{\mu}})^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \\ &= (m+n)(\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)^{\mathsf{T}}\hat{\boldsymbol{\Sigma}}_{11}^{-1}(\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1) \\ &\quad + n^2\{\bar{\mathbf{X}}_{2.1} - (\boldsymbol{\mu}_2 - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\boldsymbol{\mu}_1)\}^{\mathsf{T}}\mathbf{S}_{2.1}^{-1}\{\bar{\mathbf{X}}_{2.1} - (\boldsymbol{\mu}_2 - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\boldsymbol{\mu}_1)\} \\ &= Q_1 + Q_2 \qquad \text{(say)}, \end{aligned} \tag{2.5}$$

where $\bar{\mathbf{X}}_{2.1} = \bar{\mathbf{X}}_2 - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\bar{\mathbf{X}}_1$. It is difficult to derive the exact distribution of Q in (2.5). Since Q is similar to Hotelling's T^2 statistic, and the distribution of Q is independent of the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ (see below), we approximate the distribution of Q by the distribution of $dF_{p,\nu}$, where $F_{p,\nu}$ denotes an F random variable with p and ν degrees of freedom. The unknown positive constants d and ν can be estimated by equating the first two moments of Q to those of $dF_{p,\nu}$. To find the moments of Q, we can assume without loss of generality that $\boldsymbol{\mu} = 0$. Let $Q_{2d} = n\bar{\mathbf{X}}_1^{\mathsf{T}}\mathbf{S}_{11}^{-1}\bar{\mathbf{X}}_1$ and $R_2 = Q_2/(1 + Q_{2d})$. It is known that $R_2 \sim np_2F_{p_2,n-p}/(n-p)$ independently of $Q_{2d} \sim p_1F_{p_1,n-p_1}/(n-p_1)$ and $Q_1 \sim (m+n)p_1F_{p_1,m+n-p_1}/(m+n-p_1)$ (see, for example, Seber 1984, p. 52). Since $Q_2 = (1+Q_{2d})R_2$, the distribution of Q_2 is independent of the covariance matrix $\boldsymbol{\Sigma}$. Thus the distribution of $Q = Q_1 + Q_2$ does not depend on $\boldsymbol{\Sigma}$. Using these results and noting that $\mathcal{E}\, Q_2 = \mathcal{E}\, R_2 \mathcal{E}\,(1 + Q_{2d})$, we find

$$\mathcal{E}\, Q = \mathcal{E}\,(Q_1 + Q_2) = \frac{(m+n)p_1}{m+n-p_1-2} + \frac{n(n-2)p_2}{(n-p_1-2)(n-p-2)} = M_1 \text{ (say)}, \quad (2.6)$$

where the first term on the right-hand side is $\mathcal{E}\, Q_1$ and the second term is $\mathcal{E}\, Q_2$.

Since Q_1 and Q_{2d} need not be independent, it is difficult to get the exact second moment of Q. To find an approximate expression for it, we note that $|\mathcal{E}\, Q_1Q_2 - \mathcal{E}\, Q_1 \mathcal{E}\, Q_2| = \mathcal{E}\, R_2 |\mathcal{E}\, Q_1Q_{2d} - \mathcal{E}\, Q_1 \mathcal{E}\, Q_{2d}|$; using this fact and applying the Cauchy-Schwarz inequality to $\mathcal{E}\, Q_1Q_{2d}$, it is easy to show that $|\mathcal{E}\, Q_1Q_2 - \mathcal{E}\, Q_1\mathcal{E}\, Q_2| < O(n^{-1})$. Hence for large n, $\mathcal{E}\, Q_1Q_2 \simeq \mathcal{E}\, Q_1 \mathcal{E}\, Q_2$. Also note that $\mathcal{E}\,(Q_2)^2 = \mathcal{E}\,(R_2)^2\mathcal{E}\,(1 + Q_{2d})^2$. Thus,

$$\begin{aligned}\mathcal{E}\,(Q)^2 &\simeq \mathcal{E}\,(Q_1)^2 + 2\mathcal{E}\, Q_1 \mathcal{E}\, Q_2 + \mathcal{E}\,(Q_2)^2 \\ &= \frac{(m+n)^2p_1(p_1+2)}{(m+n-p_1-2)(m+n-p_1-4)} + 2\mathcal{E}\, Q_1\mathcal{E}\, Q_2 \\ &\quad + \frac{n^2(n-2)(n-4)p_2(p_2+2)}{(n-p-2)(n-p-4)(n-p_1-2)(n-p_1-4)} \\ &= M_2 \quad \text{(say)}. \end{aligned} \quad (2.7)$$

If $n > p+4$, then equating M_1 and M_2 respectively to the first two moments of $dF_{p,\nu}$ and solving for d and ν, we get

$$\nu = \frac{4pM_2 - 2(p+2)M_1^2}{pM_2 - (p+2)M_1^2} \quad \text{and} \quad d = M_1\,\frac{\nu-2}{\nu}. \quad (2.8)$$

Thus, for given $0 < \alpha < 0.5$, an approximate $100(1-\alpha)\%$ confidence region for $\boldsymbol{\mu}$ is the set of values of $\boldsymbol{\mu}$ that satisfy

$$(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^{\mathsf{T}}(\widehat{Cov}\ \hat{\boldsymbol{\mu}})^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \le dF_{p,\nu}(\alpha), \quad (2.9)$$

where $F_{p,\nu}(\alpha)$ denotes the $100(1-\alpha)$th percentile point of an F-distribution with p and ν degrees of freedom.

2.2. Validity of the Approximation.

To understand the nature of the approximation, we estimated the coverage probabilities of (2.9) using simulation (100,000 runs) for $\alpha = 0.01$, 0.05 and different values of n and m. The IMSL subroutine RNMVN was used to generate normal variates, and a Fortran subroutine of Smith and Hocking (1972) was used to generate Wishart variates.

TABLE 1: Critical values and estimated coverage probabilities.

n	$m = 10$	15	20	25	30	50	100	1000
				$p_1 = p_2 = 2$, $\alpha = 0.01$				
10	55.51	55.07	54.79	54.61	54.47	54.19	53.94	53.69
	(0.989)	(0.989)	(0.989)	(0.989)	(0.989)	(0.989)	(0.989)	(0.990)
15	27.56	27.28	27.11	26.98	26.80	26.69	26.52	26.34
	(0.989)	(0.989)	(0.989)	(0.990)	(0.990)	(0.990)	(0.990)	(0.990)
20	21.68	21.46	21.30	21.19	21.10	20.90	20.71	20.49
	(0.988)	(0.989)	(0.989)	(0.989)	(0.990)	(0.989)	(0.989)	(0.990)
25	19.26	19.08	18.96	18.86	18.72	18.59	18.40	18.17
	(0.989)	(0.988)	(0.989)	(0.989)	(0.989)	(0.989)	(0.990)	(0.990)
30	17.94	17.80	17.69	17.61	17.55	17.38	17.19	16.95
	(0.988)	(0.988)	(0.988)	(0.988)	(0.989)	(0.989)	(0.989)	(0.989)
50	15.78	15.72	15.67	15.63	15.59	15.48	15.34	15.11
	(0.988)	(0.989)	(0.988)	(0.989)	(0.989)	(0.989)	(0.990)	(0.990)
100	14.45	14.43	14.42	14.40	14.39	14.34	14.27	14.09
	(0.989)	(0.989)	(0.989)	(0.989)	(0.989)	(0.989)	(0.989)	(0.990)
1000	13.39	13.38	13.38	13.38	13.38	13.38	13.38	13.36
	(0.990)	(0.990)	(0.990)	(0.990)	(0.990)	(0.990)	(0.990)	(0.990)
				$p_1 = p_2 = 2$, $\alpha = 0.05$				
10	27.65	27.33	27.13	27.00	26.91	26.69	26.51	26.33
	(0.952)	(0.952)	(0.952)	(0.952)	(0.953)	(0.952)	(0.953)	(0.953)
15	16.93	16.75	16.62	16.54	16.47	16.31	16.18	16.04
	(0.950)	(0.949)	(0.952)	(0.948)	(0.949)	(0.952)	(0.952)	(0.953)
20	14.15	14.01	13.92	13.85	13.79	13.66	13.53	13.38
	(0.948)	(0.950)	(0.950)	(0.949)	(0.952)	(0.952)	(0.952)	(0.953)
25	12.90	12.80	12.72	12.66	12.62	12.50	12.37	12.22
	(0.949)	(0.948)	(0.949)	(0.950)	(0.949)	(0.950)	(0.951)	(0.951)
30	12.19	12.11	12.05	12.00	11.96	11.86	11.74	11.59
	(0.947)	(0.949)	(0.948)	(0.948)	(0.949)	(0.951)	(0.950)	(0.949)
50	10.97	10.94	10.91	10.88	10.86	10.80	10.71	10.58
	(0.949)	(0.950)	(0.948)	(0.948)	(0.949)	(0.949)	(0.949)	(0.949)
100	10.19	10.18	10.17	10.17	10.16	10.13	10.09	9.99
	(0.948)	(0.949)	(0.950)	(0.948)	(0.949)	(0.949)	(0.948)	(0.950)
1000	9.56	9.56	9.56	9.56	9.56	9.55	9.55	9.54
	(0.950)	(0.949)	(0.951)	(0.950)	(0.950)	(0.950)	(0.950)	(0.950)

The critical values and the corresponding estimated coverage probabilities are given in Table 1. It is clear from this table that the approximation is very good for all values of n and m.

Remark 2.1. A referee suggested approximating the distribution of Q by the distribution of $dF_{p,n}$. The unknown constant d can be estimated as $d = \mathcal{E}\, Q/\mathcal{E}\, F_{p,n}$. We also evaluated the validity of this one-moment approximation using simulation. The simulated coverage probabilities are more or less equal to the specified confidence level as long as $n \geq 20$ and $m \geq 10$, although they are less satisfactory than the results of the two-moment approximation. For example, when $\alpha = 0.01$ and $(m, n) = (10, 10)$, the simulated coverage probability is 0.982; at (10,20) it is 0.989; at (20,10) it is 0.982; at (10,25) it is 0.989.

TABLE 2: Powers of the LRT and the T^2 test: $n = 20$, $m = 10$, $p_1 = p_2 = 2$, $\alpha = 0.05$.

δ_1	δ_2	LRT	T^2-test	δ_1	δ_2	LRT	T^2-test	δ_1	δ_2	LRT	T^2-test
0	0	0.049	0.051	0.13	0.03	0.284	0.271	0.06	0.15	0.294	0.301
0.01	0.18	0.244	0.256	0.18	0.07	0.432	0.429	0.13	0.10	0.356	0.357
0.01	0.28	0.362	0.381	0.23	0.04	0.465	0.457	0.13	0.18	0.460	0.467
0.01	0.35	0.446	0.466	0.36	0.04	0.679	0.665	0.25	0.10	0.585	0.578
0.01	0.41	0.506	0.527	0.50	0.04	0.822	0.804	0.25	0.20	0.671	0.670
0.01	0.50	0.588	0.612	0.69	0.04	0.936	0.926	0.30	0.18	0.720	0.716
0.01	0.89	0.852	0.871	0.84	0.04	0.968	0.964	0.36	0.19	0.789	0.784
0.01	1.32	0.962	0.969	0.88	0.03	0.976	0.973	0.36	0.30	0.849	0.847
0.01	1.93	0.995	0.997	0.94	0.06	0.986	0.984	0.38	0.61	0.951	0.953

2.3. Power Comparison.

We compute the power of the test based on (2.9) (we call it the T^2 test) to understand the probability that (2.9) covers false values of $\boldsymbol{\mu}$. Further, we compare the powers of the T^2 test with those of the LRT. The LRT statistic (Bhargava 1962) for the present setup is given by

$$\Lambda = \left(1 + \frac{Q_1}{m+n}\right)^{-(m+n)/2} \left(1 + \frac{R_2}{n}\right)^{-n/2}. \tag{2.10}$$

Its approximate distribution using a Box series approximation [for details about Box series approximation see, for example, Anderson (1984, p. 311)] is given by

$$P(-2\rho \ln \Lambda \le x) = (1 - w_2)P(\chi^2_p \le x) + w_2 P(\chi^2_{p+4} \le x) + O(n^{-3}), \tag{2.11}$$

where

$$\rho = 1 - \frac{(m+n)p(p+2) - mp_1(p_1+2)}{2np(m+n)},$$

and

$$w_2 = p((m+n)^2p^2(p^2-4) - mp_1(p_1+2)[3m(p_1+1)^2$$
$$- 6(p+1)^2(m+n) + (m+2n)\{4p(p_1+1)+3\}])$$
$$\div\ 12[(m+n)p\{2(n-1)-p\} + mp_1(p_1+2)]^2.$$

For given α and an observed value Λ_o of Λ, the LRT rejects $\mathcal{H}_0 : \boldsymbol{\mu} = 0$ when $P(-2\rho \ln \Lambda > -2\rho \ln \Lambda_0) < \alpha$. We computed the powers of the LRT and the T^2 test for $n = 20$, $m = 10$, and various values of $\delta_1 = \boldsymbol{\mu}_1^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1$ and $\delta_2 = (\boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_{2.1}^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1)$ using simulation with 100,000 runs. The maximum standard error of the simulation estimates is $(0.25/100,000)^{0.5} = 0.0016$, and hence it is expected that the actual powers are within 0.0032 of the values given in Table 2.

We observe from Table 2 that the T^2 test has slightly more power than the LRT test when δ_2 is large compared to δ_1, and vice versa when δ_1 is large compared to δ_2; otherwise the differences between the powers are negligible.

2.4. Some Features of the Confidence Region and an Advantage of Using Extra Data.

The confidence region (2.9) is an ellipsoid centered at $\hat{\boldsymbol{\mu}}$ and invariant under the transformation $(\mathbf{X}, \mathbf{Y}) \longrightarrow (\mathbf{AX}, \mathbf{A}_{11}\mathbf{Y})$, where $\mathbf{A}$ is a nonsingular matrix whose (1,2) partitioned $p_1 \times p_2$ part is a zero matrix. To understand the advantage of using the additional data, we want to compute the ratio of the expected volume of the ellipsoid (2.9) to that of the one based on complete data (additional data are ignored). This ratio is given by

$$\frac{n^{p_1}\left(\prod_{i=1}^{p_1}\frac{\Gamma((m+n-i+1)/2)}{\Gamma((m+n-i)/2)}\right)\{dF_{p,\nu}(\alpha)\}^{p/2}}{(m+n)^{p_1}\left(\prod_{i=1}^{p_1}\frac{\Gamma((n-i+1)/2)}{\Gamma((n-i)/2)}\right)\left(\frac{npF_{p,n-p}(\alpha)}{n-p}\right)^{p/2}}. \tag{2.12}$$

Numerical values of (2.12) when $\alpha = 0.05$ and $p = 2$ are 0.990 for $n = 10$ and $m = 0$, 0.785 for 10 and 5, 0.991 for 20 and 0, 0.876 for 20 and 5, and 0.795 for 20 and 10; when $p_1 = 2$ and $p_2 = 1$, they are 0.717 for $n = 30$ and $m = 10$, and 0.566 for 30 and 20. These values indicate that extra data are useful for better confidence estimation.

2.5. Simultaneous Inferences.

Simultaneous confidence intervals can be developed using Scheffé's S-method as follows: Let $c_o = \{dF_{p,\nu}(\alpha)\}^{0.5}$ and $n > p+4$. Note that the inequality (2.9) holds if and only if

$$\frac{\{\mathbf{a}^{\mathrm{T}}(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})\}^2}{\mathbf{a}^{\mathrm{T}}(\widehat{Cov}\ \hat{\boldsymbol{\mu}})\mathbf{a}} \leq dF_{p,\nu}(\alpha)$$

for every nonzero vector $\mathbf{a} \in \mathbb{R}^p$. This implies that

$$P\left(\mathbf{a}^{\mathrm{T}}\hat{\boldsymbol{\mu}} - c_o\sqrt{\mathbf{a}^{\mathrm{T}}(\widehat{Cov}\ \hat{\boldsymbol{\mu}})\mathbf{a}} \leq \mathbf{a}^{\mathrm{T}}\boldsymbol{\mu} \leq \mathbf{a}^{\mathrm{T}}\hat{\boldsymbol{\mu}} + c_o\sqrt{\mathbf{a}^{\mathrm{T}}(\widehat{Cov}\ \hat{\boldsymbol{\mu}})\mathbf{a}} \text{ for all } \mathbf{a}\right) = 1-\alpha. \tag{2.13}$$

Thus for particular finite choices of $\mathbf{a}$, the left-hand side of (2.13) is at least $1-\alpha$. In particular, we have

$$P\left(\hat{\mu}_i - c_o\sqrt{\hat{\Sigma}_{ii}} \leq \mu_i \leq \hat{\mu}_i + c_o\sqrt{\hat{\Sigma}_{ii}} \text{ for } i = 1,\ldots,p\right) \geq 1-\alpha, \tag{2.14}$$

where $\hat{\Sigma}_{ii}$ denotes the (i,i)th element of $\widehat{Cov}\ \hat{\boldsymbol{\mu}}$.

It seems to be difficult to derive Bonferroni intervals for any arbitrary linear functions of $\boldsymbol{\mu}$; however, a simple form of Bonferroni intervals for the components of $\boldsymbol{\mu}$ can be obtained. Let $\mathbf{S}_{1u}$ denote $(m+n-1)^{-1}$ times the sums-of-squares-and-products matrix based on all the observations on the first p_1 components, and $\mathbf{S}_{2u} = \mathbf{S}_{22}/(n-1)$. Further, let $t_l(q)$ denote the $100(1-q)$th percentile point of Student's t-distribution with l degrees of freedom, let $\mathbf{a}_{1i}$ be the $p_1 \times 1$ vector with the ith element 1 and the rest 0, $i = 1,\ldots,p_1$, and define $\mathbf{a}_{2j}$ similarly for $j = 1,\ldots,p_2$. Then simultaneous confidence intervals for $\mu_1,\ldots,\mu_p$ are given by

$$P\left(\frac{|\mathbf{a}_{1i}^{\mathrm{T}}(\hat{\boldsymbol{\mu}}_1-\boldsymbol{\mu}_1)|}{(\mathbf{a}_{1i}^{\mathrm{T}}\mathbf{S}_{1u}\mathbf{a}_{1i})^{0.5}} \leq t_1,\ \frac{|\mathbf{a}_{2j}^{\mathrm{T}}(\bar{\mathbf{X}}_2-\boldsymbol{\mu}_2)|}{(\mathbf{a}_{2j}^{\mathrm{T}}\mathbf{S}_{2u}\mathbf{a}_{2j})^{0.5}} \leq t_2,\ i = 1,\ldots,p_1,\ j = 1,\ldots,p_2\right) \geq 1-\alpha, \tag{2.15}$$

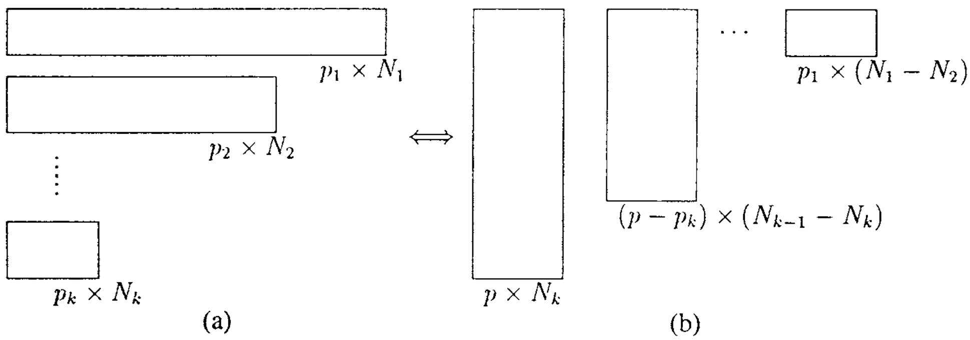

FIGURE 1: General monotone pattern.

where $t_1 = t_{m+n-1}(\alpha/(2p))$ and $t_2 = t_{n-1}(\alpha/(2p))$.

Remark 2.2. We now give an application of the above results to a repeated-measurements model. Suppose that the pretreatment data have m more observations than the posttreatment data and we want to test the equality of the pre- and post-treatment mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. In our present setup, this is equivalent to testing $\mathcal{H}_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = 0$, where $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T)^T$. Therefore, (2.13) with appropriate choices of $\mathbf{a}$ gives simultaneous conservative confidence intervals for the differences between the components of the pre- and posttreatment mean vectors. These simultaneous confidence intervals can be used as a conservative test for the repeated-measurements model.

3. GENERAL MONOTONE CASE

The results of the previous section can be extended to data having a block-monotone pattern. By block-monotone we mean that there are N_i observations available on the ith p_i components, $i = 1, \ldots, k$, $p_1 + \cdots + p_k = p$, and $N_1 \geq \cdots \geq N_k$, as shown in Figure 1(a). In other words, as shown in Figure 1(b), we have N_k observations on all the p components, $N_{k-1} - N_k$ observations on the first $p - p_k$ components, and so on. Note that the blocks in Figure 1(b) are independent and the observations within each block are also independent.

Although it is not difficult to see that the idea of Section 2 works for a general monotone pattern, notational complexities prevent us from giving confidence regions in explicit form as in (2.9). Therefore, we give expressions for the MLEs of $\boldsymbol{\mu}_i$'s, where $(\boldsymbol{\mu}_1^T, \ldots, \boldsymbol{\mu}_k^T)^T = \boldsymbol{\mu}$, and $\widehat{Cov}\ \hat{\boldsymbol{\mu}}$ for $k = 3$. Explicit expressions for the MLEs of μ and $\boldsymbol{\Sigma}$ for a general k and the details of the moment calculations can be obtained from the authors upon request.

Let $\bar{\mathbf{X}}_{i,l}$ be the sample mean of the ith block of p_i components, and $\mathbf{S}_{ij,l}$ be the sums-of-squares-and-products matrix corresponding to the ith block of p_i components and the jth block of p_j components that are obtained by pooling the first l samples [first l blocks in Figure 1(b)], $i, j = 1, \ldots, k - l + 1$ and $l = 1, \ldots, k$. For $k = 3$, let

$$\mathbf{B}_{21} = \mathbf{S}_{21,2}\mathbf{S}_{11,2}^{-1} \quad \text{and} \quad (\mathbf{B}_{31}, \mathbf{B}_{32}) = (\mathbf{S}_{31,1}\mathbf{S}_{32,1}) \begin{pmatrix} \mathbf{S}_{11,1} & \mathbf{S}_{12,1} \\ \mathbf{S}_{21,1} & \mathbf{S}_{22,1} \end{pmatrix}^{-1}.$$

The MLEs are

$$\hat{\boldsymbol{\mu}}_1 = \bar{\mathbf{X}}_{1,3}, \qquad \hat{\boldsymbol{\mu}}_2 = \bar{\mathbf{X}}_{2,2} - \mathbf{B}_{21}(\bar{\mathbf{X}}_{1,2} - \hat{\boldsymbol{\mu}}_1),$$

$$\hat{\boldsymbol{\mu}}_3 = \bar{\mathbf{X}}_{3,1} - \mathbf{B}_{31}(\bar{\mathbf{X}}_{1,1} - \hat{\boldsymbol{\mu}}_1) - \mathbf{B}_{32}(\bar{\mathbf{X}}_{2,1} - \hat{\boldsymbol{\mu}}_2),$$

$$\hat{\boldsymbol{\Sigma}}_{11} = \frac{\mathbf{S}_{11,3}}{N_1}, \qquad \hat{\boldsymbol{\Sigma}}_{21} = \mathbf{B}_{21}\hat{\boldsymbol{\Sigma}}_{11}, \qquad \hat{\boldsymbol{\Sigma}}_{22} = \frac{\mathbf{S}_{2.1,2}}{N_2} + \mathbf{B}_{21}\hat{\boldsymbol{\Sigma}}_{12},$$

$$\hat{\boldsymbol{\Sigma}}_{31} = \mathbf{B}_{31}\hat{\boldsymbol{\Sigma}}_{11} + \mathbf{B}_{32}\hat{\boldsymbol{\Sigma}}_{21}, \qquad \hat{\boldsymbol{\Sigma}}_{32} = \mathbf{B}_{31}\hat{\boldsymbol{\Sigma}}_{12} + \mathbf{B}_{32}\hat{\boldsymbol{\Sigma}}_{22},$$

$$\hat{\boldsymbol{\Sigma}}_{33} = \frac{\mathbf{S}_{3.21,1}}{N_3} + \mathbf{B}_{31}\hat{\boldsymbol{\Sigma}}_{13} + \mathbf{B}_{32}\hat{\boldsymbol{\Sigma}}_{23},$$

where

$$\mathbf{S}_{2.1,2} = \mathbf{S}_{22,2} - \mathbf{S}_{21,2}\mathbf{S}_{11,2}^{-1}\mathbf{S}_{12,2}$$

and

$$\mathbf{S}_{3.21,1} = \mathbf{S}_{33,1} - (\mathbf{S}_{31,1}\mathbf{S}_{32,1}) \begin{pmatrix} \mathbf{S}_{11,1} & \mathbf{S}_{12,1} \\ \mathbf{S}_{21,1} & \mathbf{S}_{22,1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{S}_{13,1} \\ \mathbf{S}_{23,1} \end{pmatrix}.$$

Let $\bar{\mathbf{X}}_* = (\bar{\mathbf{X}}_{1,1}^{\mathrm{T}}, \bar{\mathbf{X}}_{2,1}^{\mathrm{T}}, \bar{\mathbf{X}}_{3,1}^{\mathrm{T}}, \bar{\mathbf{X}}_{1,2}^{\mathrm{T}}, \bar{\mathbf{X}}_{2,2}^{\mathrm{T}}, \bar{\mathbf{X}}_{1,3}^{\mathrm{T}})^{\mathrm{T}}$ and

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & 0 & \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & 0 & -\mathbf{B}_{21} & \mathbf{1} & \mathbf{B}_{21} \\ -\mathbf{B}_{31} & -\mathbf{B}_{32} & 1 & -\mathbf{B}_{32}\mathbf{B}_{21} & \mathbf{B}_{32} & (\mathbf{B}_{31} + \mathbf{B}_{32}\mathbf{B}_{21}) \end{pmatrix},$$

so that $\mathbf{A}\bar{\mathbf{X}}_* = \hat{\boldsymbol{\mu}}$. Let $[\widehat{Cov}\ \hat{\boldsymbol{\mu}}]_{ij} : p_i \times p_j$ denote the (i,j)th partitioned matrix of $\widehat{Cov}\ \hat{\boldsymbol{\mu}} = \mathbf{A}[\widehat{Cov}(\bar{\mathbf{X}}_*)]\mathbf{A}'$ for $i,j = 1,2,3$. Then

$$[\widehat{Cov}\ \hat{\boldsymbol{\mu}}]_{11} = \frac{\hat{\boldsymbol{\Sigma}}_{11}}{N_1}, \qquad [\widehat{Cov}\ \hat{\boldsymbol{\mu}}]_{12} = \frac{\hat{\boldsymbol{\Sigma}}_{12}}{N_1}, \qquad [\widehat{Cov}\ \hat{\boldsymbol{\mu}}]_{13} = \frac{\hat{\boldsymbol{\Sigma}}_{13}}{N_1},$$

$$[\widehat{Cov}\ \hat{\boldsymbol{\mu}}]_{22} = \frac{\hat{\boldsymbol{\Sigma}}_{22}}{N_2} - \frac{N_1 - N_2}{N_1 N_2}\mathbf{B}_{21}\hat{\boldsymbol{\Sigma}}_{12},$$

$$[\widehat{Cov}\ \hat{\boldsymbol{\mu}}]_{23} = \frac{\hat{\boldsymbol{\Sigma}}_{23}}{N_2} - \frac{N_1 - N_2}{N_1 N_2}\mathbf{B}_{21}\hat{\boldsymbol{\Sigma}}_{13},$$

$$[\widehat{Cov}\ \hat{\boldsymbol{\mu}}]_{33} = \frac{\hat{\boldsymbol{\Sigma}}_{33}}{N_3} - \frac{N_1 - N_3}{N_1 N_3}(\mathbf{B}_{31}\hat{\boldsymbol{\Sigma}}_{13} + \mathbf{B}_{32}\hat{\boldsymbol{\Sigma}}_{23}) + \frac{N_1 - N_2}{N_1 N_2}\mathbf{B}_{32}\hat{\boldsymbol{\Sigma}}_{2.1}\mathbf{B}_{23},$$

where $\hat{\boldsymbol{\Sigma}}_{2.1} = \mathbf{S}_{2.1,2}/N_2$.

The moments of the quadratic form $Q = (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^{\mathrm{T}}(\widehat{Cov}\ \hat{\boldsymbol{\mu}})^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})$ are (details can be obtained from the authors)

$$M_1 = \sum_{l=1}^{k} \frac{N_l(N_l - 2)p_l}{(N_l - p_{(l-1)} - 2)(N_l - p_{(l)} - 2)} \tag{3.1a}$$

and

$$M_2 \simeq \sum_{l=1}^{k} \frac{N_l^2(N_l-2)(N_l-4)p_l(p_l+2)}{(N_l-p_{(l-1)}-2)(N_l-p_{(l)}-2)(N_l-p_{(l-1)}-4)(N_l-p_{(l)}-4)}$$

$$+2\sum_{q=l+1}^{k}\sum_{l=1}^{k-1} \frac{N_q(N_q-2)N_l(N_l-2)p_q p_l}{(N_q-p_{(q-1)}-2)(N_q-p_{(q)}-2)(N_l-p_{(l-1)}-2)(N_l-p_{(l)}-2)}. \tag{3.1b}$$

where $p_{(l)} = \sum_{j=1}^{l} p_j$ for $l = 1, \ldots, k$ and $p_{(0)} = 0$. The moments in (3.1) are valid for a general k. If $N_k > p+4$, then using (2.8) one can find d and ν so that Q is approximately distributed as $dF_{p,\nu}$. We checked the accuracy of this approximation using simulation as in Section 2 when $p_1 = p_2 = p_3 = 1$, when $p_1 = 2, p_2 = 1$, and $p_3 = 1$, and for different values of the N_i's. It is found that the approximations are as satisfactory as those in Section 2, and so they are not reported here.

4. AN EXAMPLE

We now illustrate the results of this paper using an example given in Johnson and Wichern (1992, p. 183). The data set consists of measurements on perspiration from 20 healthy females, and satisfies the normality assumption. Each observation has three components, namely, X_1 = sweat rate, X_2 = sodium content and X_3 = potassium content. We first consider a monotone pattern with two blocks by deleting the last five observations on the second and third components. In the notation of Section 3, we have $p_1 = 1$, $p_2 = 2$, $N_1 = 20$, and $N_2 = 15$. The hypotheses considered in the example are $\mathcal{H}_0 : \boldsymbol{\mu}^{\mathrm{T}} = (4, 50, 10)$ and $\mathcal{H}_1 : \boldsymbol{\mu}^{\mathrm{T}} \neq (4, 50, 10)$. Using incomplete data, we computed

$$\hat{\boldsymbol{\mu}} = \begin{pmatrix} 4.640 \\ 44.311 \\ 9.893 \end{pmatrix}, \quad \widehat{Cov}(\hat{\boldsymbol{\mu}}) = \begin{pmatrix} 0.137 & 0.521 & -0.095 \\ 0.521 & 12.354 & -0.105 \\ -0.095 & -0.105 & 0.237 \end{pmatrix}$$

which give $T^2 = 11.14$ with p-value 0.070. Now, ignoring the additional data, that is, using only the first 15 complete observations (partially complete), we computed

$$\hat{\boldsymbol{\mu}}_{\mathrm{pc}} = \begin{pmatrix} 4.247 \\ 42.813 \\ 10.167 \end{pmatrix}, \quad \widehat{Cov}(\hat{\boldsymbol{\mu}}_{\mathrm{pc}}) = \begin{pmatrix} 0.157 & 0.598 & -0.109 \\ 0.598 & 13.388 & -0.140 \\ -0.109 & -0.140 & 0.258 \end{pmatrix}$$

which give $T^2_{\mathrm{pc}} = 7.99$ with p-value 0.131. Note that T^2 provides more evidence against the null hypothesis than T^2_{pc}. Further, using all the 20 complete observations, the observed value of Hotelling's T^2 is $T^2_{\mathrm{c}} = 9.74$ with p-value 0.0650, which is quite close to the p-value 0.070 of T^2. Further, it is evident from Figure 2 that the confidence ellipsoid based on incomplete data is closer (in terms of location and volume) to the one based on complete data than to the one based on partially complete data. Thus, in a sense, our approach is able to recover the information lost due to deletion of the five bivariate observations.

We next compute 90% simultaneous confidence intervals based on partially complete data (T^2_{pc}), incomplete data (T^2) and all the 20 complete observations (T^2_{c}). These intervals along with Bonferroni intervals are given in Table 3.

The intervals based on incomplete data are narrower than the corresponding ones based on partially complete data in all the cases. Further, observe that even though both T^2

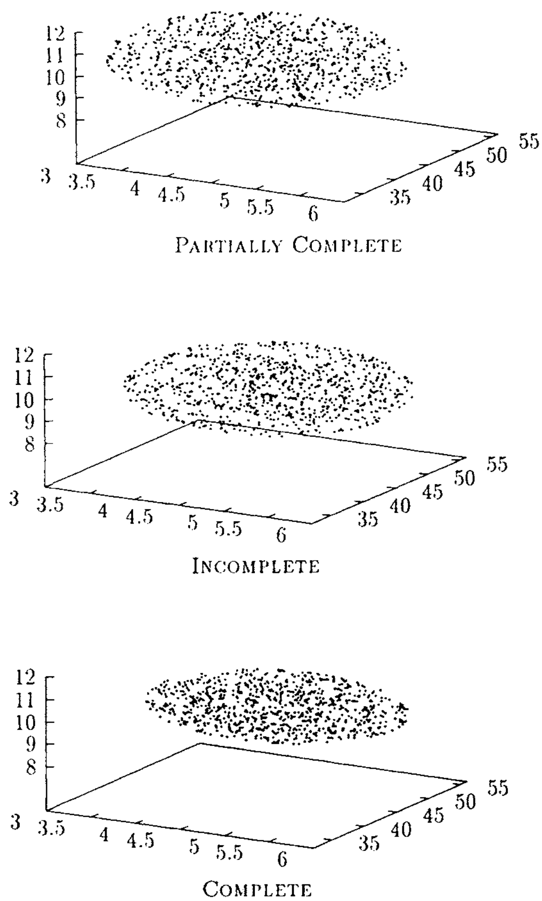

FIGURE 2: Confidence ellipsoids for sweat data.

(when $N_1 = 20$ and $N_2 = 15$) and T_c^2 tests reject $\mathcal{H}_0 : \boldsymbol{\mu}' = (4, 50, 10)$ at the 10% level, the confidence intervals based on them and Bonferroni intervals are not able to detect the means that are significantly different from the specified ones in $\mathcal{H}_0$. This is not surprising, because all these confidence intervals are conservative.

5. CONCLUSIONS AND DISCUSSION

In this article, we have developed approximate confidence regions for a normal mean vector based on incomplete data sets with a monotone pattern and a nonmonotone pattern. The confidence regions are easy to use, and as tests they are as powerful as the LRTs. In general, the method used to develop confidence regions is applicable for different patterns of data for which Anderson's (1957) approach can be used to derive the MLEs—that is, patterns for which the MLEs can be obtained from the conditional and marginal likelihood functions. For confidence estimation of normal mean vectors based on data sets with a nonmonotone pattern (Lord 1955), we refer to Krishnamoorthy and Pannala (1996). Further, the results of this article can be extended to the two-sample case and multivariate analysis of variance. We are currently investigating these problems and plan to report the results elsewhere.

TABLE 3: 90% simultaneous confidence intervals.

(N_1,N_2)	Statistic	μ_1	μ_2	μ_3
(20, 10)	T^2_{pc}	4.58 ± 1.76	46.50 ± 13.05	9.92 ± 2.28
	T^2	4.64 ± 1.31	46.54 ± 12.75	9.86 ± 1.98
	BF	4.64 ± 0.87	46.50 ± 9.52	9.92 ± 1.66
(20, 12)	T^2_{pc}	4.52 ± 1.37	46.73 ± 10.96	10.35 ± 1.99
	T^2	4.64 ± 1.22	46.84 ± 10.76	10.23 ± 1.88
	BF	4.64 ± 0.87	46.72 ± 8.29	10.35 ± 1.51
(20, 15)	T^2_{pc}	4.25 ± 1.20	42.81 ± 11.05	10.17 ± 1.54
	T^2	4.66 ± 1.14	44.31 ± 10.87	9.89 ± 1.50
	BF	4.64 ± 0.87	42.81 ± 8.64	10.17 ± 1.20
(20, 17)	T^2_{pc}	4.51 ± 1.26	45.31 ± 10.87	9.87 ± 1.45
	T^2	4.64 ± 1.11	45.77 ± 10.60	9.78 ± 1.38
	BF	4.64 ± 0.87	45.31 ± 8.62	9.87 ± 1.15
(20, 20)	T^2_c	4.64 ± 1.09	45.40 ± 9.04	9.97 ± 1.22
	BF	4.64 ± 0.87	45.40 ± 7.24	9.97 ± 0.98

ACKNOWLEDGEMENTS

The authors are grateful to two referees, an associate editor and the editor for many helpful comments and providing useful references.

REFERENCES

Anderson, T.W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *J. Amer. Statist. Assoc.*, 52, 200–203.

Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

Bhargava, R.P. (1962). Multivariate tests of hypotheses with incomplete data. Ph.D. Thesis, Department of Statistics, Stanford University.

Johnson R.A., and Wichern, D.W. (1992). *Applied Multivariate Statistical Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey.

Krishnamoorthy, K., and Pannala, M.K. (1996). Confidence estimation of a normal mean vector with incomplete data. Technical Report 96-1, Department of Mathematics, University of Southwestern Louisiana.

Li, K.H., Raghunathan, T.E., and Rubin, D.B. (1991). Large-sample significance levels from multiply imputed data using moment-based statistics and an F reference distribution. *J. Amer. Statist. Assoc.*, 86, 1065–1073.

Little, R.J.A. (1988). A test of missing completely at random for multivariate data with missing values. *J. Amer. Statist. Assoc.*, 83, 1198–1202.

Little, R.J.A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *J. Amer. Statist. Assoc.*, 90, 1112–1121.

Little, R.J.A., and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. Wiley, New York.

Lord, F.M. (1955). Estimation of parameters from incomplete data. *J. Amer. Statist. Assoc.*, 50, 870–876.

Mehta, J.S., and Gurland, J. (1969). A test of equality of means in the presence of correlation and missing values. *Biometrika*, 60, 211–213.

Meng, X.L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statist. Sci.*, 9, 538–573.

Morrison, D.F. (1973). A test for equality of means of correlated variates with missing data on one response. *Biometrika*, 60, 101–105.

Naik, U.D. (1975). On testing equality of means of correlated variables with incomplete data. *Biometrika*, 62, 615–622.

Rubin, D.B. (1976). Inference and missing data (with discussion). *Biometrika*, 63, 581–592.

Rubin, D.B., and Schenker, N. (1986). Multiple imputation for interval estimation from simple random sample with ignorable nonresponses. *J. Amer. Statist. Assoc.*, 81, 366–374.

Seber, G.A.F. (1984). *Multivariate Observations*. Wiley, New York.

Smith, W.B., and Hocking, R.R. (1972). Wishart variates generator (algorithm AS 53). *Appl. Statist.*, 21, 341–345.

Received 29 May 1996
Revised 27 November 1996
Accepted 15 October 1997

Department of Mathematics
University of Southwestern Louisiana
Lafayette, Louisiana 70504
U.S.A.
e-mail: kxk4695@usl.edu

Actuarial Department
United Guaranty Corporation
Greensboro, North Carolina 27420
U.S.A.