THE MIXED MODEL FOR MULTIVARIATE REPEATED MEASURES: VALIDITY CONDITIONS AND AN APPROXIMATE TEST

ROBERT J. BOIK
DEPARTMENT OF MATHEMATICAL SCIENCES
MONTANA STATE UNIVERSITY

Repeated measures on multivariate responses can be analyzed according to either of two models: a doubly multivariate model (DMM) or a multivariate mixed model (MMM). This paper reviews both models and gives three new results concerning the MMM. The first result is, primarily, of theoretical interest; the second and third have implications for practice. First, it is shown that, given multivariate normality, a condition called multivariate sphericity of the covariance matrix is both necessary and sufficient for the validity of the MMM analysis. To test for departure from multivariate sphericity, the likelihood ratio test can be employed. The second result is an approximation to the null distribution of the likelihood ratio test statistic, useful for moderate sample sizes. Third, for situations satisfying multivariate normality, but not multivariate sphericity, a multivariate ε correction factor is derived. The ε correction factor generalizes Box's ε and can be used to construct an adjusted MMM test.

Key words: likelihood ratio test, matrix quadratic form, multivariate sphericity, Wishart distribution.

1. Introduction

In a repeated measures experiment having \( t \) time periods, each subject gives a \( t \)-dimensional response. As is well known, the resulting data can be analyzed according to Scheffé's (1956) univariate mixed model or according to a multivariate model. Barcikowski and Robey (1984) review some of the factors to consider when choosing between a univariate and a multivariate analysis. O'Brien and Kaiser (1985) provide a primer on multivariate analyses of repeated measures. Hand and Taylor's (1987) introductory text includes some examples of multivariate analysis of repeated measures on real data sets.

In a multivariate experiment having \( p \) dependent variables, each subject gives a \( p \)-dimensional response. In this case, \( p \) separate univariate analyses or one multivariate analysis can be performed. Separate analyses may be appropriate if the \( p \) dependent variables are uncorrelated or if they are measures of distinct theoretical constructs. A multivariate analysis is appropriate if the \( p \) dependent variables are functionally related or if the Type I error rate is to be controlled experimentwise.

In a multivariate repeated measures experiment, each subject gives a \( p \)-dimensional response on each of \( t \) occasions. The result is a \( pt \)-dimensional response from each subject. Here, the researcher must decide between \( p \) separate analyses and one joint analysis, and between a mixed model and a multivariate model. Thus, four distinct analysis strategies exist. The two models corresponding to a joint analysis are the multivariate mixed model (MMM) and the doubly multivariate model (DMM). Bock (1975) discusses both models and Timm (1980) illustrates each using a small data set. An excellent summary of the MMM, along with some new theoretical results, is given by Thomas (1983).

Section 2 of this paper briefly reviews the DMM and MMM analyses. The data set
from Timm (1980) is used to illustrate the analyses. Beginning in section 3, some new results on the MMM analysis are derived. Thomas (1983) showed that, given multivariate normality, a condition called multivariate sphericity is sufficient to insure validity of the MMM analysis. In section 3, multivariate sphericity is shown to be a necessary condition as well. The proof of necessity generalizes the univariate results of Huynh and Feldt (1970); Rouanet and Lépine (1970); and Mendoza, Toothaker, and Crain (1976).

The likelihood ratio (LR) test for departure from multivariate sphericity is discussed in section 4. The test is a generalization of Mauchly's (1940) sphericity test and was derived by Thomas (1983). Thomas also gave the asymptotic null distribution of the test statistic. In section 4, Box's (1949) expansion of the characteristic function of the test statistic is used to obtain a more accurate approximation to the null distribution for moderate sample sizes.

Based on empirical (Collier, Baker, Mandeville, & Hayes, 1967; Mendoza, Toothaker, and Nicewander, 1974) and theoretical (Huynh & Feldt, 1980) studies of the null behavior of the univariate mixed model test under nonsphericity, it is expected that the size of the MMM test will increase as the covariance matrix departs from multivariate sphericity. Thomas (1983) makes the same conjecture. Under nonsphericity, univariate mixed model tests can be adjusted according to an estimate of Box's (1954) $\varepsilon$. In section 5, a multivariate generalization of Box's $\varepsilon$ is derived. The generalized $\varepsilon$ can be used to study the robustness of the MMM tests to departure from multivariate sphericity and to construct adjusted MMM tests.

The new approximation to the null distribution of the LR test statistic for multivariate sphericity and the $\varepsilon$-adjusted MMM test are illustrated in section 6. Section 7 considers how the DMM, MMM, and adjusted MMM tests could be used in practice.

2. The Doubly Multivariate Model and the Multivariate Mixed Model Analyses

*The Doubly Multivariate Model Analysis*

The linear model for multivariate repeated measures on $n$ subjects can be written as

$$Y = \theta X + \zeta$$

where $Y: pt \times n$ is the response matrix, $\theta: pt \times k$ is a matrix of parameters, $X: k \times n$ is the between groups design matrix having rank $(X) = r \leq k$, and $\zeta: pt \times n$ is a matrix of random errors. The $j$-th column of $Y$ represents the $pt$ dimensional response from the $j$-th subject. The responses are ordered within each column according to time and within time according to dependent variable. That is, the first $p$ elements in the response vector represent the $p$ dependent measures at time 1. Denote the $pt \times 1$ vector of random errors for the $j$-th subject by $\zeta_j$. It is assumed that $\zeta_j \sim iid N(0, \Omega)$ where $\Omega: pt \times pt$ is positive definite. Jointly,

$$\text{vec} (\zeta) \sim N_{npt}([0, (I_p \otimes \Omega)])$$

where the vec operator stacks columns (Henderson & Searle, 1979).

Most hypotheses of interest can be put into the form

$$H_0: (C' \otimes I_p)\theta F = 0$$

where $C: t \times q$ has rank $(C) = q$ and $F: k \times s$ has rank $(F) = s$. The columns of $C$ consist of the coefficients of $q$ linear functions (e.g., contrasts) of the $t$ occasions (time periods). Without loss of generality, the occasions coefficient matrix is assumed to satisfy $C'C = I_q$. The columns of $F$ consist of the coefficients of $s$ estimable between group functions. For $\theta F$ to be estimable, $F$ must be contained in the column space of $X$. 