A SAS/IML program for implementing the modified Brown–Forsythe procedure in repeated measures designs

Guillermo Vallejo\textsuperscript{a},*, Joaquín Moris\textsuperscript{a}, Néilda M. Conejo\textsuperscript{b}

\textsuperscript{a} Methodology Area, Department of Psychology, University of Oviedo, Plaza Feijóo, s/n, E-33003 Oviedo, Spain
\textsuperscript{b} Psychobiology Laboratory, Department of Psychology, University of Oviedo, Plaza Feijóo, s/n, E-33003 Oviedo, Spain

\textbf{A R T I C L E  I N F O}

Article history:
Received 23 July 2005
Received in revised form 29 May 2006
Accepted 1 June 2006

Keywords:
Univariate repeated measures
Covariance heterogeneity
Modified Brown–Forsythe approximation
SAS/IML
Multiple post hoc contrasts

\textbf{A B S T R A C T}

In this paper, we present a computer program written in version 9.1 of SAS' interactive matrix language in order to implement a new approach for analyzing repeated measures data. Previous studies reported that the new procedure is as powerful as conventional solutions and generally more robust (i.e., insensitive) to violations of assumptions that underlie conventional solutions. The program also included a step-wise procedure based on the Bonferroni inequality to test comparisons among the repeated measurements. Both univariate and multivariate repeated measures data can be analyzed. Finally, the application of the SAS/IML program is illustrated with a numeric example.

© 2006 Elsevier Ireland Ltd. All rights reserved.

\textbf{1. Introduction}

Repeated measures designs are used extensively in epidemiology, psychology, neuropsychology, psychopharmacology, and other research areas [1–3]. In many repeated measures designs, especially those employed in clinical studies, the data are collected from \( N \) subjects forming \( J \) independent groups over \( K \) occasions or trials or under different experimental conditions. Several methods have been proposed to analyze such designs, many of which can be implemented using widely available standard statistical packages such as SAS, S-PLUS, or SPSS. When the variances of all pair-wise differences among levels of the repeated measures factor are equal (i.e., sphericity) and this constant variance is the same for all levels of the between-subjects grouping factor (jointly, these two assumptions have been referred to as multisample sphericity; see [4] for details), it is well known that the conventional univariate (or mixed-model of Scheffé) approach provides the most powerful tests. If sphericity fails to hold but the covariance matrices are homogeneous, the resulting data can be analyzed by either a univariate model with Box's epsilon (\( \varepsilon \)) correction for degrees of freedom (d.f.; [5]) or a full multivariate model. The empirical literature indicates, however, that both tests are sensitive to departures from the assumptions of multivariate normality and multisample sphericity, particularly when group sizes are unequal [6].

To counteract the negative impact of the violation of multisample sphericity on the type I error rates, diverse solutions have been proposed. Algina and Oshima [7] suggest using the improved general approximation (IGA) test developed initially by Huynh [4], whereas Lix and Keselman [8] proposed a Welch–James (WJ) type test derived by Johansen [9]. Based on the power results presented by Algina and Keselman [10], the WJ test may be preferred over the IGA test,
provided that sample size are sufficiently large to obtain a robust WJ test. Nevertheless, when the sample size is small and the data are not extracted from normal distributions, the WJ approach does not adequately control the type I error rate for a test of the groups × trials interaction, commonly the most interest test for researchers. For their part, Vallejo and Livacic-Rojas [11] compared the behaviour of a multivariate extension of the Brown and Forsythe (BF; [12]) procedure [13] with that of a mixed model based on covariance structures selected by means of information criteria such as Akaike’s information criterion [14] and Bayesian information criterion [15] when normality and covariance assumptions were violated. Their simulation studies showed that the BF approach performed as well or better than its competitor, in terms of control of type I error rates, particularly with small sample sizes and unstructured covariance matrices. For tests of the interaction, however, the BF rates were conservative for the negatively paired conditions of group size and covariance matrices—particularly when the sample sizes was small and the degree of group size inequality was substantial.

However, in a recent article, Vallejo and Ato [16] showed that a typically robust test for the interaction effect may be obtained by modifying BF’s approach (referred to as modified BF test hereafter, MBF). In order to obtain better type I error control with the BF procedure, Vallejo and Ato, used the d.f. correction proposed by Krishnamoorthy and Yu [17] instead of the correction suggested by Nel and van der Merwe [18], which was used in the previous d.f. formulation provided by Vallejo et al. [13]. The procedure due to Vallejo and Ato, in addition to being more robust than that provided by Vallejo et al., is improved in other two aspects. First, the approximate error d.f. are invariant to linear transformations of outcome variables. Consequently, the p values for testing $H_{01}: A_1 \mu = 0$ will not be different from those for testing $H_{02}: A_2 \mu = 0$, by a change of scale of the elements of $A$, where $A$ is a known matrix of contrasts with appropriate size and $\mu$ is the $K$-variate location parameter. Second, the new approximate error d.f. always are positive, which is not so clearly the case on applying Nel and van der Merwe’s solution.

Therefore, the purpose of this article is to extend the MBF procedure to produce focused test statistics and to make available a program written in the SAS/IML language [19] in order to obtain numerical results. Besides to facilitate the access to this robust method, developing this program within SAS provides an opportunity to utilize other statistical procedures of this system widely used nowadays. In Sections 2 and 3, we present the modified BF method and code required for the program, along with instructions on its use, for testing omnibus effects and multiple contrast hypotheses related to these effects. In Section 4 we use the data from a study reported by Fitzmaurice et al. [20] to illustrate the application of the computer program with MBF procedure. Finally, some concluding remarks are given in Section 5.

2. Definition of test statistic

Let $y_{ijk}$, $i = 1, \ldots, n_j; j = 1, \ldots, j_k; k = 1, \ldots, K$, be the response for the $i$th subject in the $j$th group at the $k$th trial, and let $y_{ij} = (y_{i1j}, \ldots, y_{ijn_j})'$ be the random vector of responses associated with the $i$th subject in the $j$th group ($\sum n_j = N$). Then, by stacking the subvectors $y_{1ij}^\prime, \ldots, y_{n_jij}^\prime$, a general linear model for univariate repeated measures can be written as

$$Y = XB + U,$$

where $Y$ is the $N \times K$ matrix of observed data, $B$ is the $J \times K$ matrix that contains the unknown fixed effects to be estimated from the data with known design matrix $X$ and $U$ is the $N \times K$ matrix of unknown random errors. The model assumes that the random vector $y_{ij}$ are normally and independently distributed within each level $j$, with mean vector $\mu_j$ and variance–covariance matrix $\Sigma_j$. The unbiased estimators of $\Sigma_j$ are $\hat{\Sigma}_j = (1/n_j - 1)E_j$, where $E_j = y_j'y_j - \hat{\beta}_j^\prime X_j y_j$ are distributed independently as Wishart $W_K(n_j - 1, 1, \Sigma_j)$ [21]. We also assume that $n_j - 1 \geq K$ so that $\hat{\Sigma}_j^{-1}$ exists with probability one.

2.1. Modified BF procedure

Let us consider the problem of finding a transformation $F$ for the common multivariate criteria when homogeneity of covariance matrices is not a tenable assumption, with the aim of testing hypotheses of the form $H_0: CB\bar{A} = 0$ using the MBF approach, where $C$ is the $(J - 1) \times J$ matrix which defines a set of $(J - 1)$ linearly independent contrasts for the between-groups factor and $A$ is a $K \times (K - 1)$ matrix which defines a set of $(K - 1)$ linearly independent contrasts for the within-subjects factor. The new invariant solution is obtained by modifying the approximate d.f. proposed by Vallejo et al. [13]. The statistics used to test the hypothesis concerning to the interaction effect using the MBF approach, are functions of the eigenvalues of $H^\prime E^{-1}H$, where the hypothesis matrix is

$$H = (C\bar{A})[C(X'X)^{-1}C]^\prime (C\bar{A}),$$

and the error matrix is

$$E' = \left(\begin{smallmatrix} V_1^c \\ V_h^c \end{smallmatrix}\right) \sum_{j=1} J c_j A(\Sigma^{1/2} Q_j \Sigma^{1/2}) A'.$$

In Eq. (3), $V_1^c$ and $V_h^c$ denote the approximate d.f. for matrices $E'$ and $H$, respectively; $c_1 = 1 - (n_j/N)$. $c_q = (\Sigma - c_j A_1 A_1^\prime)$. $Q_j = (\Sigma - c_j A_1 A_1^\prime)$. $\Sigma = (c_1 E_1 + \cdots + c_j E_j)$, and $\Sigma^{1/2}$ denotes the square root of the matrix $\Sigma$. Notice that the error matrix in Eq. (3) is equivalent to Eq. (12) in Vallejo et al. [13] paper, since, by assumption, $\Sigma$ is a positive definite matrix. Then, exist $\Sigma^{1/2}$ and $\Sigma^{-1/2}$ such that $\Sigma^{1/2} \Sigma^{-1/2} = \Sigma$ and $\Sigma^{-1/2} \Sigma^{1/2} = \Sigma^{-1}$ and that $\Sigma^{1/2} \Sigma^{-1/2} = I_k$, where $I_k$ is the identity matrix. To find the d.f. associated with $E'$, first, the sum $\sum_{j=1} J c_j A_1 Q_j A = c_1 A_1 Q_1 A + \cdots + c_j A_1 Q_j A$ is approximated as

$$\sum_{j=1} J c_j A_1 Q_j A \sim W_k \left( f_k' \frac{1}{f_k} \sum_{j=1} J c_j A_1 Q_j A \right).$$

Then, proceeding in a fashion similar to Nel [21] and Nel and van der Merwe [18], the parameter will be found equating the first two moments of $\sum_{j=1} J c_j A_1 Q_j A$, namely $c_1 A_1 Q_1 A$ and
\[ \sum_{j=1}^{J} (t_j - 1)^{-1} (c_j^T A Q_j A)^2 \], to those of \( \text{W}_k(f^{*}, A^* Q^* A^*) \). It follows that the quantity \( f^*_n \) is given by

\[
f^*_n = \frac{\text{tr}^2 \left( \sum_{j=1}^{J} c_j^T A Q_j A \right) + \text{tr} \left( \sum_{j=1}^{J} c_j^T A Q_j A \right)^2}{\sum_{j=1}^{J} 1/(t_j - 1) \text{tr}^2 (c_j^T A Q_j A) + \text{tr} (c_j^T A Q_j A)^2}. \tag{5}
\]

where \( \text{tr}(\cdot) \) stands for the trace of the matrix. In turn, using the so-called multivariate Satterthwaite’s approximation described by Vallejo and Ato [16] in the Appendix, the quantity \( f^*_n \) is given by

\[
f^*_n = \frac{\text{tr}^2 \left( \sum_{j=1}^{J} c_j^T A Q_j A \right) + \text{tr} \left( \sum_{j=1}^{J} c_j^T A Q_j A \right)^2}{\sum_{j=1}^{J} (V_j + \text{tr}^2 (\sum_{j=1}^{J} c_j^T A^j E_j^T E_j A) + \text{tr} (\sum_{j=1}^{J} c_j^T A^j E_j^T E_j A)^2). \tag{6}
\]

where \( V = \{ \text{tr}^2 (A^j E_j^T E_j A) + \text{tr} (A^j E_j^T E_j A)^2 \} = 2c_j [\text{tr}^2 (A^j E_j^T E_j A) + \text{tr} (A^j E_j^T E_j A)^2] \). In Eqs. (5) and (6), the numerator can be simplified to \( \text{tr}(\hat{T}_{J - 1}^2) + \text{tr}(\hat{T}_{K - 1})^2 \), given that transforming \( \Sigma \) to \( E_j^{-1/2} E_j E_j^{-1/2} \), \( A Q_j \rightarrow I_{K - 1} \). Therefore, replacing \( Q_j \) in Eqs. (5) and (6) by its estimate \( E_j^{-1/2} E_j E_j^{-1/2} \) and using the result that \( \text{tr} (A B) = \text{tr} (B A) \), the approximate d.f. simplifies to

\[
f^*_n = \frac{(K - 1) + (K - 1)^2}{\sum_{j=1}^{J} 1/(t_j - 1) [\text{tr}^2 (c_j^T A^j E_j^T E_j A) + \text{tr} (c_j^T A^j E_j^T E_j A)^2]}. \tag{7}
\]

and

\[
f^*_n = \frac{(K - 1) + (K - 1)^2}{\sum_{j=1}^{J} (V_j + \text{tr}^2 (\sum_{j=1}^{J} c_j^T A^j E_j^T E_j A) + \text{tr} (\sum_{j=1}^{J} c_j^T A^j E_j^T E_j A)^2). \tag{8}
\]

The result in Eqs. (7) and (8) has considerable theoretical appeal because, as Krishnamoorthy and Yu [17] show, in addition to being invariant under any nonsingular transformation, lies between \( \min(n_j - 1) \) and \( N - J \) for all \( c_j^T A^j E_j A \) and \( n_j - 1 \geq K - 1 \). Therefore, the approximate d.f. never could be negative, while this not the case with Nel and van der Merwe’s solution.

There are several multivariate test statistics for testing the null hypothesis of no interaction between groups by trials. The most common ones are the Wilks’ [22] \( \Lambda \)-criterion, the Hotelling-Lawley trace, and the Pillai-Bartlett trace statistics. Although their critical values have been widely tabulated and charted (see [23]), in practice it is usual to obtain the level of significance by defining each of these approximations in terms of an \( F \)-variable [24-27]. In the two-group case, all the \( F \)-test approximations are interchangeable. For our purpose, the \( F \)-approximation of Wilks’ \( \Lambda \) due to Rao [24] is chosen as the multivariate test statistic. According to this transformation, the interaction null hypothesis is rejected if

\[
F_{\text{MBF}} = \frac{1 - A^{1/2} (v^*_1)^2}{\frac{1}{2} (v^*_2)^2} \geq F(1 - \alpha, v^*_1, v^*_2). \tag{9}
\]

where \( A = \det(E^T)/\det(H + E^T) \), \( \alpha = [(2v_1^2 - 4)/(2^2 + v_1^2 - 5)]^{1/2} \), \( v_1^* = v_1 - 1 \), and \( v_2^* = (v_1^* - (v_1^* + 1)/2) \). If \( A = 1 \), the null hypothesis is rejected, the interpretation of tests of main effects is insufficient to understand the information of the data. When this occurs, test of possible 2 \( \times \) 2 interactions (interaction contrasts or tetrads contrasts) can be used [28-31]. Under departures from the assumption of covariance homogeneity, tetrads contrasts results are obtained easily using the MBF procedure with appropriate \( A \) and \( C \) contrast matrices. Specifically, to test interaction contrasts using MBF approach, \( C = \Sigma_{1}, A = \Sigma_{2} \), where \( c_j \) is a \( 1 \times J \) vector of coefficients that contrasts the \( j \)th and \( j \)th subjects means and \( \Sigma_{2} \) is a \( K \times 1 \) vector of coefficients that contrasts the \( k \)th and \( k \)th within-subjects means. It is important to mention, as a reviewer indicates, that it is not necessary that the omnibus \( F \)-test be significant prior to testing planned (i.e., the contrasts are determined on before the data are collected) tetrads contrasts, provided the type I error rate is controlled. Contemporary practice favors adopting the family of contrasts as the conceptual unit for control of the type I error rate.

To control the family-wise error rate (FWER) for all possible 2 \( \times \) 2 interactions, several post hoc procedures may be used. For instance, Lix and Keselman [31] found that the Hochberg [32] step-up Bonferroni, Shaffer [33] modified sequentially rejective Bonferroni, and Studentized maximum modulus critical value [34] procedures used in combination with Johansen’s [9] procedure are largely robust to departures from multi-sample sphericity. Nevertheless, only the Hochberg procedure will be considered in this article. We selected the Hochberg step-up Bonferroni procedure over the Shaffer and studentized maximum modulus approaches because Lix and Keselman [31] found minimal power differences between them and because is very simple to apply. With Hochberg’s [32] method, the \( p \) values corresponding to the \( r \) tests statistics for testing the hypotheses \( H_1, \ldots, H_r \) are ranked, where \( r = J \times J \), \( J = \sum_{1}^{r} \) and \( K = K \). Then, the largest probability is compared to \( \alpha_{/}\text{FWER} \), where \( \alpha_{/}\text{FWER} \) is the family-wise error rate the researcher is willing to tolerate. If \( p_{/}\text{FWER} \leq \alpha_{/}\text{FWER} \), all hypotheses are rejected without further test; otherwise, the next largest probability is compared to \( \alpha_{/}\text{FWER}/2 \). If \( p_{/}\text{FWER} \leq \alpha_{/}\text{FWER}/2 \), all hypotheses \( H_1, \ldots, H_{K - 1} \) are rejected. Continuing in this fashion, at any stage \( q \), reject all \( H_q \) where \( q \leq q' \), if \( p_{/}\text{FWER}/(r - q + 1) \) for any \( q = r, r - 1, \ldots, 1 \).

As before, the test used for checking the effect of the trials with unweighted means is given by the determinant of the matrix.

\[
\tilde{H} = (C^T A)^{-1} \mid C^T (A X^T)^{-1} C \mid^{-1} (C^T A)^{-1}. \tag{10}
\]

and

\[
\tilde{E} = \left( \frac{1}{\sum_{j=1}^{J} n_j^{-1}} \sum_{j=1}^{J} n_j^{-1} A^j (X_j E_j^{1/2} X_j^{1/2}) A \right), \tag{11}
\]

where \( A \) was defined before, \( \tilde{B} = \left[ \sum_{j=1}^{J} (1/n_j) \right]^{1/2} \). C = c is a \((J \times 1)\) vector consisting of all ones, \( E = (\Sigma_{1}/n_1 + \cdots + \Sigma_{J}/n_J) \),
and \( \psi^* = 1 \). Extending the results reported by Krishnamoorthy and Yu [17] and Nel and van der Merwe [18], the distribution of \( \sum_{j=1}^{J} n_j^{-1} A^\prime Q_j A \) can be approximated as a sum of Wishart distributions:

\[
\sum_{j=1}^{J} n_j^{-1} A^\prime Q_j A \sim SW_k \left( f^*_k, \frac{1}{f^*_k} \sum_{j=1}^{J} n_j^{-1} A^\prime Q_j A \right),
\]

where the quantity \( f^*_k \) is given by

\[
f^*_k = \frac{\text{tr}^2 \left( \sum_{j=1}^{J} n_j^{-1} A^\prime Q_j A \right) + \text{tr} \left( \sum_{j=1}^{J} n_j^{-1} A^\prime Q_j A \right)^2}{\sum_{j=1}^{J} 1/(n_j - 1) \text{tr}^2(n_j^{-1} A^\prime Q_j A) + \text{tr}(n_j^{-1} A^\prime Q_j A)^2}.
\]

Replacing \( Q_j \) in Eq. (13) by its unbiased estimate \( \hat{\Sigma}_j \) the approximate d.f. can be written as

\[
v^*_k = \frac{(K - 1) + (K - 1)^2}{\sum_{j=1}^{J} 1/(n_j - 1) \text{tr}^2(n_j^{-1} A^\prime \hat{\Sigma}_j \hat{\Sigma}_j^{-1} A) + \text{tr}(n_j^{-1} A^\prime \hat{\Sigma}_j \hat{\Sigma}_j^{-1} A)^2}.
\]

The simplification at the Eq. (14) occurs because \( A^\prime Q_j A \rightarrow I_{K-1} \). For the main effect of trials averaged over the groups, all the F-test approximations are interchangeable. According to the adaptation of the Rao [24] transformation, the main effect of trials null hypothesis is rejected if

\[
F_{\text{MBF}} = \frac{1 - A^{1/2}}{A^{1/2}} \left( \frac{\psi^*_k}{\psi^*_k} \right) \geq F^{1-\alpha}(v^*_1, v^*_2).
\]

where \( A = \text{det}(\hat{E})/\text{det}(\hat{H} + \hat{E}) \), \( s = [(I^2 + \nu^2 - 4)/(I^2 + \nu^2 - 5)]^{1/2} \), \( v^*_1 = h^*_k, \) and \( v^*_2 = \left( \psi^*_k - (1 - v^*_1 + 1)/2 - (h^*_k - 2)/2 \right) \).

In turn, using the Mehrotra [33] extension of the univariate Brown–Forsythe test applied to the sum of the within-subject variables, the test statistic for testing the effect of the groups is given by determinant of \( E^*(H + E^*)^{-1} \), where the hypothesis matrix is defined as in Eq. (2) with \( A \equiv a \) (i.e., a \( K \times 1 \) vector with each element equal to one), the error matrix is

\[
E^* = \left( \frac{\psi^*_k}{\psi^*_k} \right) \sum_{j=1}^{J} c_j \hat{A}^\prime \hat{\Sigma}_j A.
\]

and \( \psi^*_k \) and \( \psi^*_k \) are the approximate d.f. for \( E^* \) and \( H \), respectively. The definition of the estimators referring to the d.f. as it applies to the analysis of repeated measures can be found in Vallejo and Ato [16]. Lastly, the null hypothesis referring to the equality of the groups, weighted by means of the trials is rejected if

\[
F_{\text{MBF}} = \frac{H}{E^*} \left( \frac{\psi^*_k}{\psi^*_k} \right) \geq F^{1-\alpha}(v^*_h, v^*_k).
\]

When there is no interaction and the assumption of multisample sphericity is not satisfied, the MBF approach also may be applied to obtain robust multiple comparison procedures for examining all possible pair of groups and trials comparison marginal means. To test between-subjects pairwise comparison hypotheses using MBF procedure, \( C = c_{j'} \) and \( A = a. \) The significance of the pair-wise comparisons for the within-subjects main effect can be probed in a similar manner, but with \( C = c \) (i.e., a \( 1 \times J \) vector with each element equal to one) and \( A = a_{j,j'}. \)

At present there are numerous simultaneous or sequential multiple comparison procedures that maintain the FWE at or below its nominal \( \alpha \)-level when the validity assumptions of traditional statistics are satisfied (see [36]). However, when normality and covariance homogeneity are not satisfied, the number of procedures that remain relatively unaffected by assumption violations it diminishes considerably. Results of Keselman [37], Keselman and Lix [38], and Kowalchuk and Keselman [39] suggest that the Welch’s [40] step-up range, Schaffer’s [33] sequentially rejective step-down Bonferroni, and Hochberg’s [32] sequentially rejective step-up Bonferroni procedures performed well in terms of control of type I error rates and power to detect true pair-wise differences. Therefore, the method used in the preceding paragraphs can be applied to control the FWE.

To obtain numerical results for the MBF procedure described in the previous section we developed a computational program (available for download at http://gip.uniovi.es/gdiyad/documen/macrossas01.pdf) written in the SAS/IML programming language [19]. The program is presented as a set of subroutines or modules and a driver. The subroutines are OMNIRETEST, GROUPTEST, TIMETEST, INTERACTEST, and DEVPARTEST. They are run sequentially, and each of them checks the conditions of application. The program calculates the MBF approximate solution for tests of the main and interaction effects in repeated measures designs. In addition, contrasts among marginal means or all possible interaction contrasts (i.e., tetrad contrasts) can be obtained. When we have a set of multivariate repeated measures data, the program also can be used to test omnibus effects and multiple comparison hypotheses related to these effects; both separately for each dependent variable and simultaneously. All of the \( F \) tests and the Hochberg adjusted \( p \) values are calculate automatically by the program.

To implement the program it is assumed that the data is entered in a SAS data set named DATARECORDED with multivariate format. The hallmark feature of a univariate format is that each subject has multiple rows (or records)—one for each measurement occasion, whereas the hallmark features of a multivariate format is that each subject has only one row (or record), regardless of the number of measurements made. The program only requires that the user specifies the number of dependent variables (NVD). A run statement of the program generates as output \( F \)-statistics, along with degree of freedom and significance levels for hypothesis testing. The program also provides as output by default a step-wise
| ID<sup>b</sup> | J<sup>c</sup> | K<sub>1</sub><sup>d</sup> | K<sub>2</sub><sup>d</sup> | K<sub>3</sub><sup>d</sup> | K<sub>4</sub><sup>d</sup> | K<sub>5</sub><sup>d</sup> | K<sub>6</sub><sup>d</sup> | ID<sup>b</sup> | J<sup>c</sup> | K<sub>1</sub><sup>d</sup> | K<sub>2</sub><sup>d</sup> | K<sub>3</sub><sup>d</sup> | K<sub>4</sub><sup>d</sup> | K<sub>5</sub><sup>d</sup> | K<sub>6</sub><sup>d</sup> | ID<sup>b</sup> | J<sup>c</sup> | K<sub>1</sub><sup>d</sup> | K<sub>2</sub><sup>d</sup> | K<sub>3</sub><sup>d</sup> | K<sub>4</sub><sup>d</sup> | K<sub>5</sub><sup>d</sup> | K<sub>6</sub><sup>d</sup> | ID<sup>b</sup> | J<sup>c</sup> | K<sub>1</sub><sup>d</sup> | K<sub>2</sub><sup>d</sup> | K<sub>3</sub><sup>d</sup> | K<sub>4</sub><sup>d</sup> | K<sub>5</sub><sup>d</sup> | K<sub>6</sub><sup>d</sup> | ID<sup>b</sup> | J<sup>c</sup> | K<sub>1</sub><sup>d</sup> | K<sub>2</sub><sup>d</sup> | K<sub>3</sub><sup>d</sup> | K<sub>4</sub><sup>d</sup> | K<sub>5</sub><sup>d</sup> | K<sub>6</sub><sup>d</sup> |
|------|-----|------|------|------|------|------|------|------|-----|-----|------|------|------|------|------|------|------|-----|-----|------|------|------|------|------|------|-----|-----|------|------|------|------|------|
| 0035 | 1   | 39   | 43   | 19   | 29   | 26   | 16   | 0229 | 2   | 31   | 31   | 38   | 38   | 26   | 19   | 1162 | 3   | 04   | 17   | 06   | 05   | 16   | 08   | 0056 | 1   | 05   | 05   | 01   | 02   | 01   | 02   | 0237 | 2   | 03   | 03   |
| 0142 | 1   | 29   | 13   | 26   | 11   | 17   | 14   | 0259 | 2   | 48   | 122  | 65   | 50   | 22   | 16   | 0123 | 4   | 07   | 19   | 09   | 14   | 15   | 04   | 0150 | 1   | 44   | 17   | 18   | 30   | 14   | 05   | 0470 | 2   | 32   | 30   |
| 0175 | 1   | 22   | 25   | 17   | 16   | 32   | 23   | 0499 | 2   | 14   | 49   | 19   | 16   | 19   | 19   | 0148 | 4   | 72   | 102  | 63   | 36   | 22   | 50   | 0215 | 1   | 06   | 09   | 02   | 06   | 01   | 04   | 0557 | 2   | 34   | 56   |
| 0226 | 1   | 17   | 28   | 04   | 09   | 02   | 04   | 0596 | 2   | 80   | 60   | 50   | 70   | 60   | 40   | 0291 | 4   | 30   | 30   | 20   | 20   | 10   | 10   | 0227 | 1   | 55   | 37   | 39   | 27   | 24   | 31   | 0631 | 2   | 49   | 46   |
| 0248 | 1   | 11   | 11   | 04   | 18   | 34   | 20   | 0658 | 2   | 36   | 45   | 32   | 35   | 15   | 17   | 0348 | 4   | 16   | 29   | 07   | 11   | 06   | 06   | 0264 | 1   | 32   | 169  | 49   | 33   | 18   | 26   | 0659 | 2   | 44   | 31   |
| 0344 | 1   | 09   | 12   | 10   | 16   | 21   | 29   | 0882 | 2   | 50   | 90   | 130  | 120  | 80   | 90   | 0466 | 4   | 20   | 05   | 40   | 40   | 32   | 24   | 0482 | 1   | 10   | 19   | 07   | 07   | 06   | 10   | 0177 | 3   | 40   | 109  |
| 0486 | 1   | 39   | 20   | 16   | 38   | 11   | 08   | 0247 | 3   | 42   | 21   | 110  | 51   | 30   | 23   | 0602 | 4   | 10   | 20   | 30   | 40   | 40   | 034 | 0597 | 1   | 20   | 10   | 10   | 10   | 10   | 10   | 0290 | 3   | 10   | 10   |
| 0598 | 1   | 20   | 10   | 40   | 10   | 20   | 20   | 0324 | 3   | 02   | 10   | 40   | 50   | 50   | 80   | 0784 | 4   | 15   | 50   | 40   | 40   | 50   | 30   | 0603 | 1   | 20   | 30   | 20   | 20   | 30   | 20   | 0332 | 3   | 20   | 20   |
| 0876 | 1   | 20   | 20   | 10   | 10   | 10   | 0393 | 3   | 13   | 64   | 93   | 84   | 89   | 81   | 0811 | 4   | 89   | 129  | 229  | 166  | 123  | 91   | 0881 | 1   | 55   | 10   | 10   | 20   | 10   | 0440 | 3   | 13   | 15   | 21   | 20   | 11   | 12   | 0867 | 4   | 30   | 50   |
| 0133 | 2   | 21   | 23   | 15   | 08   | 06   | 03   | 0456 | 3   | 16   | 108  | 80   | 29   | 20   | 26   | 0884 | 4   | 40   | 130  | 90   | 90   | 90   | 110  | 0161 | 2   | 28   | 62   | 72   | 65   | 121  | 62   | 0483 | 3   | 15   | 19   |
| 0164 | 2   | 36   | 27   | 13   | 19   | 39   | 08   | 0570 | 3   | 33   | 53   | 107  | 86   | 111  | 130  | 1072 | 4   | 58   | 92   | 37   | 31   | 43   | 25   | 0191 | 2   | 37   | 28   | 33   | 46   | 33   | 27   | 0570 | 3   | 30   | 30   |
| 0225 | 2   | 07   | 20   | 02   | 02   | 03   | 07   | 0953 | 3   | 30   | 100  | 146  | 97   | 63   | 26   | 0225 | 2   | 07   | 20   | 02   | 02   | 03   | 07   | 0953 | 3   | 30   | 100  |

* AIDS clinical trial group.

<sup>b</sup> Variable that identifies the subject to which the record refers.

<sup>c</sup> Levels of the between-subjects factor.

<sup>d</sup> Levels of the within-subjects factor.
procedure based on the Bonferroni inequality (i.e., Hochberg method) to discover which linear combination of the means or interaction contrasts have significant differences. However, the user program can use several optional scalars (TESTOMNI, TESTGROUPS, TESTTIME, TESTINTERAC, TESTDEPVAR), which assume values of 0 or 1, for print the interest information.

4. Example and comparison with traditional analysis of variance

The application of this procedure is illustrated using data reported by Henry et al. [41] in the Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology. These authors discuss a randomized, doubly blind, study to determine the relative clinical efficacy of four different reverse transcriptase inhibitor therapies in AIDS patients with advanced immune suppression (CD4 counts of less than or equal to 50 cells per mm$^3$). Specifically, 1313 HIV-infected patients were randomized to one of four daily regimens containing 600 mg of zidovudine: zidovudine alternating monthly with 400 mg didanosine; zidovudine plus 2.25 mg of zalcitabine; zidovudine plus 400 mg of didanosine; or zidovudine plus 400 mg of didanosine plus 400 mg of nevirapine (triple therapy). The time to new HIV disease progression or death, toxicities, the change in CD4 cells, and plasma HIV-1 RNA concentrations in a subset of study subjects were evaluated. Measurements of CD4 counts at baseline (prior to the initiation of treatment) and at 8-week intervals during a 40-week follow-up period for 60 and 8 selected subjects are displayed in Table 1 in a multivariate format. Each subject has his or her own row of data containing the values of outcome variable on each of the six levels of the within-subjects factor (0, 8, 16, 24, 32, and 40 weeks, which are denoted by $K_1$, $K_2$, $K_3$, $K_4$, $K_5$, and $K_6$, respectively). Each record also contains two identifying variables: ID, which identifies the subject to which the record refers; $J$, which identifies the levels of the between-subjects factor. The categorical variable treatment is coded: 1 = zidovudine alternating monthly with 400 mg didanosine, 2 = zidovudine plus 2.25 mg of zalcitabine, 3 = zidovudine plus 400 mg of didanosine, and 4 = zidovudine plus 400 mg of didanosine plus 400 mg of nevirapine.

For the data shown in Table 1, first we will carry out a conventional repeated measures analysis of variance, which is summarized in Table 2. According to this analysis, the classical F-test statistic gives stronger evidence for effects of treatment group, trials, and treatment $\times$ trial interaction. A 0.05 significance level is assumed throughout the paper. For the within-subjects main effect, $F = 8.11$, with 5 and 320 d.f. ($p < 0.0001$). Finally, for the interaction effect, results contained in Table 2 show that $F$-value is 2.32, with 15 and 320 d.f. ($p = 0.0008$), so that also is highly significant. Consequently, the classical statistic indicates that the shapes of the profiles are not the same across the four groups.

The three univariate tests we have just considered have assumed normality and equal dispersion matrices for the four groups under study. However, using Box's M-test, as given in Timm [42], p. 134), the hypothesis of equal covariance matrices is untenable. The $\chi^2$-approximation criterion is 180.29 with 61 d.f. ($p < 0.0001$). When multisample sphericity is violated, the mixed-model of Scheffé's approach suffers from inflated nominal levels and thus should be used with caution. In order to circumvent the problems caused for the lack of homogeneity of dispersion matrices, the MBF procedure is a good choice, since it becomes more conservative in these cases. A part of the results generated by SAS/IML program appears in Table 3.

To produce the previous results the following program statements were specified:

```plaintext
data datarecorded; input group y1 y2 y3 y4 y5 y6; data datarecorded; input group y1 y2 y3 y4 y5 y6;
```

Table 2 – Summary of traditional univariate analysis

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>$MS_H$ a</th>
<th>$MS_E$ b</th>
<th>F-value</th>
<th>d.f.</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groups</td>
<td>15989.81</td>
<td>5152.97</td>
<td>3.20</td>
<td>3, 64</td>
<td>0.0327</td>
</tr>
<tr>
<td>Trials</td>
<td>3316.53</td>
<td>416.47</td>
<td>7.95</td>
<td>5, 320</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Groups $\times$ trials</td>
<td>1103.33</td>
<td>416.47</td>
<td>2.65</td>
<td>15, 320</td>
<td>0.0008</td>
</tr>
</tbody>
</table>

a Mean square for the hypothesis.

b Mean square for the error.

c Degrees of freedom.

According to the results included in Table 3, it can be appreciated that the subject’s mean levels are significantly different among the four treatments ($F = 3.32$ with 2.35 and 47.51 d.f., $p = 0.0373$). Also is evident that the within-subjects main effect is highly significant ($F = 3.95$ with 5 and 38.44 d.f., $p = 0.0054$). However, it is important to note that there is not a significant difference between the response patterns for the groups over time; in other words, there is some evidence that the groups do not respond differently during the first 40 weeks of follow-up ($F = 1.72$ with 13.16 and 108.31 d.f., $p = 0.0666$). Consequently, one could not reject the null hypothesis at the 5% level of significance.
The basic purpose of this paper was to extend the MBF procedure for testing omnibus effects and multiple contrast hypotheses related to these effects, to make available a program written in the SAS/IML language in order to implement this procedure, and to illustrate the application of the computer program using data for a design grouped measures repeated. Previous studies had revealed that the MBF procedure was generally robust (i.e., insensitive) to violations of assumptions.

Hochberg’s sequentially rejective Bonferroni procedure one comparison is significant controlling FWE at a level no more 0.05. J1 versus J4. On the other hand, note that for the within-subjects marginal means, four comparisons are declared significant: K1 versus K2, K1 versus K3, K1 versus K6, and K2 versus K6. Because the interaction effect resulted no significant at 5%, the tetrads contrasts involving pairs of levels of two factors has not been printed in Table 4. However, a significant result would not need additional code lines to produce all possible interaction contrasts.

5. Conclusion

The basic purpose of this paper was to extend the MBF procedure for testing omnibus effects and multiple contrast hypotheses related to these effects, to make available a program written in the SAS/IML language in order to implement this procedure, and to illustrate the application of the computer program using data for a design grouped measures repeated. Previous studies had revealed that the MBF procedure was generally robust (i.e., insensitive) to violations of assumptions.
multisample sphericity and to lack of normality of the data in unbalanced designs similar to that employed in the current work [16]. To date, the performance of that approach had been restricted to the examination of robustness in a between by within subjects repeated measures design and in a two-group multivariate design. Nonetheless, this approach may also be applied to a variety of research designs using a general linear model to define the hypotheses of interest. In particular, independent and correlated groups designs containing one or more dependent variables.

Adopting the approach presented in this paper, one must keep in mind the limits of the procedure. Specifically, it should be noted that the MBF procedure assumes complete measurements for all subjects, which represents the main limitation of this procedure in longitudinal settings. In many studies, however, those researchers who do not have complete measurements on all subjects across time can use the procedure confining their attention to those complete vectors or by using the idea of multiple imputation, which has been incorporated into widely available software. Multiple imputation can be used with any kind of data and any kind of analytic procedure, however, some assumptions must be satisfied to have unbiased and efficient estimators (see [44]). Other limitations to be noted are that the MBF procedure it is not implemented in the major statistical packages and does not allows users to accommodate time-dependent covariates.

In conclusion, in spite of the fact that these limitations might dissuade potential users from using this method, we believe that the researchers should be comfortable using the MBF procedure to analyze longitudinal data, specially under conditions that are not optimal for the mixed model (e.g., when fitting a correct model requires many parameters and sample sizes are small), since they need neither to model their data nor to rely on methods that typically selected an incorrect covariance structure, as noted by Keselman et al. [45] and Kowalchuk et al. [46].

Acknowledgments

This research was supported by grant MEC-SEJ2005-01883 (Ministry of Education and Science, Spain) to Guillermo Vallejo, and by grant FICYT FR-01-GE-2 (Institute for Promotion of Scientific and Technical Research) to Nélida M. Conejo. We thank the editor and anonymous referee for their helpful comments and very constructive suggestions for improving an initial version of this article.

References

[16] G. Vallejo, M. Ato, Modified Brown–Forsythe test for analyzing repeated measures designs, Multivariate Behav. in press.


