journal homepage: www.intl.elsevierhealth.com/journals/cmpb

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

Guillermo Vallejo[a,*], *Joaquín Moris*[a], *Nélida M. Conejo*[b]

[a] *Methodology Area, Department of Psychology, University of Oviedo, Plaza Feijóo, s/n, E-33003 Oviedo, Spain*
[b] *Psychobiology Laboratory, Department of Psychology, University of Oviedo, Plaza Feijóo, s/n, E-33003 Oviedo, Spain*

ARTICLE INFO

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

ABSTRACT

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.

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 (ε) 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,

* *Corresponding author*. Tel.: +34 985 103274; fax: +34 985 104141.
E-mail address: gvallejo@uniovi.es (G. Vallejo).

0169-2607/$ – see front matter © 2006 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2006.06.006

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}:$\mathbf{A}_1\boldsymbol{\mu}=\mathbf{0}$ will not be different from those for testing H_{02}:$\mathbf{A}_2\boldsymbol{\mu}=\mathbf{0}$, by a change of scale of the elements of $\mathbf{A}$, where $\mathbf{A}$ is a known matrix of contrasts with appropriate size and $\boldsymbol{\mu}$ 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 tests 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}, $i=1,\ldots,n_j$; $j=1,\ldots,J$; $k=1,\ldots,K$, be the response for the ith subject in the jth group at the kth trial, and let $\mathbf{y}_{ij}=(y_{ij1},\ldots,y_{ijK})'$ be the random vector of responses associated with the ith subject in the jth group $\left(\sum_j n_j = N\right)$. Then, by stacking the subvectors $\mathbf{y}'_{11},\ldots,\mathbf{y}'_{NJ}$, a general linear model for univariate repeated measures can be written as

$$\mathbf{Y}=\mathbf{XB}+\mathbf{U}, \tag{1}$$

where $\mathbf{Y}$ is the $N\times K$ matrix of observed data, $\mathbf{B}$ is the $J\times K$ matrix that contains the unknown fixed effects to be estimated from the data with known design matrix $\mathbf{X}$ and $\mathbf{U}$ is the $N\times K$ matrix of unknown random errors. The model assumes that the random vector $\mathbf{y}_{ij}$ are normally and independently distributed within each level j, with mean vector $\boldsymbol{\mu}_j$ and variance–covariance matrix $\boldsymbol{\Sigma}_j$. The unbiased estimators of $\boldsymbol{\Sigma}_j$ are $\hat{\boldsymbol{\Sigma}}_j=(1/n_j-1)\mathbf{E}_j$, where $\mathbf{E}_j=\mathbf{Y}'_j\mathbf{Y}_j-\hat{\boldsymbol{\beta}}'_j\mathbf{X}'_j\mathbf{Y}_j$ are distributed independently as Wishart $W_K(n_j-1,\boldsymbol{\Sigma}_j)$ [21]. We also assume that $n_j-1\geq K$ so that $\hat{\boldsymbol{\Sigma}}_j^{-1}$ 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:$\mathbf{CBA}=0$ using the MBF approach, where $\mathbf{C}$ is the $(J-1)\times J$ matrix which defines a set of $(J-1)$ linearly independent contrasts for the between-groups factor and $\mathbf{A}$ 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 $\mathbf{HE}^{*-1}$, where the hypothesis matrix is

$$\mathbf{H}=(\mathbf{C}\hat{\mathbf{B}}\mathbf{A})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}']^{-1}(\mathbf{C}\hat{\mathbf{B}}\mathbf{A}), \tag{2}$$

and the error matrix is

$$\mathbf{E}^{*}=\left(\frac{\nu_e^*}{\nu_h^*}\right)\sum_{j=1}^{J}c_j^{\bullet}\mathbf{A}'(\boldsymbol{\Xi}^{1/2}\hat{\mathbf{Q}}_j\boldsymbol{\Xi}^{1/2})\mathbf{A}. \tag{3}$$

In Eq. (3), ν_e^* and ν_h^* denote the approximate d.f. for matrices $\mathbf{E}^*$ and $\mathbf{H}$, respectively; $c_j^{\bullet}=1-(n_j/N)$, $c_j=n_j/N$, $\hat{\mathbf{Q}}_j=(\boldsymbol{\Xi}^{-1/2}\hat{\boldsymbol{\Sigma}}_j\boldsymbol{\Xi}^{-1/2})$, $\boldsymbol{\Xi}=(c_1^{\bullet}\boldsymbol{\Sigma}_1+\cdots+c_J^{\bullet}\boldsymbol{\Sigma}_j)$, and $\boldsymbol{\Xi}^{1/2}$ denotes the square root of the matrix $\boldsymbol{\Xi}$. Notice that the error matrix in Eq. (3) is equivalent to Eq. (12) in Vallejo et al. [13] paper, since, by assumption, $\boldsymbol{\Xi}$ is a positive definite matrix. Then, exist $\boldsymbol{\Xi}^{1/2}$ and $\boldsymbol{\Xi}^{-1/2}$ such that $\boldsymbol{\Xi}^{1/2}\boldsymbol{\Xi}^{1/2}=\boldsymbol{\Xi}$ and $\boldsymbol{\Xi}^{-1/2}\boldsymbol{\Xi}^{-1/2}=\boldsymbol{\Xi}^{-1}$ and that $\boldsymbol{\Xi}^{1/2}\boldsymbol{\Xi}^{-1/2}=\mathbf{I}_K$, where $\mathbf{I}_K$ is the identity matrix. To find the d.f. associated with $\mathbf{E}^*$, first, the sum $\sum_{j=1}^{J}c_j^{\bullet}\mathbf{A}'\mathbf{Q}_j\mathbf{A}(=c_1^{\bullet}\mathbf{A}'\mathbf{Q}_1\mathbf{A}+\cdots+c_J^{\bullet}\mathbf{A}'\mathbf{Q}_j\mathbf{A})$ is approximated as

$$\sum_{j=1}^{J}c_j^{\bullet}\mathbf{A}'\mathbf{Q}_j\mathbf{A}\sim W_K\left(f_e^*,\frac{1}{f_e^*}\sum_{j=1}^{J}c_j^{\bullet}\mathbf{A}'\mathbf{Q}_j\mathbf{A}\right). \tag{4}$$

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 $\sum_{j=1}^{J}c_j^{\bullet}\mathbf{A}'\mathbf{Q}_j\mathbf{A}$, namely $c_j^{\bullet}\mathbf{A}'\mathbf{Q}_j\mathbf{A}$ and

$\sum_{j=1}^{J}(n_j-1)^{-1}(c_j^{\bullet}\mathbf{A}'\mathbf{Q}_j\mathbf{A})^2$, to those of $W_K(f^*, \mathbf{A}'\mathbf{Q}\mathbf{A})$. It follows that the quantity f_e^* is given by

$$f_e^* = \frac{\mathrm{tr}^2\left(\sum_{j=1}^{J}c_j^{\bullet}\mathbf{A}'\mathbf{Q}_j\mathbf{A}\right)+\mathrm{tr}\left(\sum_{j=1}^{J}c_j^{\bullet}\mathbf{A}'\mathbf{Q}_j\mathbf{A}\right)^2}{\sum_{j=1}^{J}1/(n_j-1)[\mathrm{tr}^2(c_j^{\bullet}\mathbf{A}'\mathbf{Q}_j\mathbf{A})+\mathrm{tr}(c_j^{\bullet}\mathbf{A}'\mathbf{Q}_j\mathbf{A})^2]}, \tag{5}$$

where tr(·) 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_h^* is given by

$$f_h^* = \frac{\mathrm{tr}^2\left(\sum_{j=1}^{J}c_j^{\bullet}\mathbf{A}'\mathbf{Q}_j\mathbf{A}\right)+\mathrm{tr}\left(\sum_{j=1}^{J}c_j^{\bullet}\mathbf{A}'\mathbf{Q}_j\mathbf{A}\right)^2}{\sum_{j=1}^{J}\{V\}+\mathrm{tr}^2\left(\sum_{j=1}^{J}c_j\mathbf{A}'\hat{\boldsymbol{\Sigma}}_j\hat{\boldsymbol{\Xi}}^{-1}\mathbf{A}\right)+\mathrm{tr}\left(\sum_{j=1}^{J}c_j\mathbf{A}'\hat{\boldsymbol{\Sigma}}_j\hat{\boldsymbol{\Xi}}^{-1}\mathbf{A}\right)^2}, \tag{6}$$

where $V = [\mathrm{tr}^2(\mathbf{A}'\hat{\boldsymbol{\Sigma}}_j\hat{\boldsymbol{\Xi}}^{-1}\mathbf{A})+\mathrm{tr}(\mathbf{A}'\hat{\boldsymbol{\Sigma}}_j\hat{\boldsymbol{\Xi}}^{-1}\mathbf{A})^2]-2c_j[\mathrm{tr}^2(\mathbf{A}'\hat{\boldsymbol{\Sigma}}_j\hat{\boldsymbol{\Xi}}^{-1}\mathbf{A})+\mathrm{tr}(\mathbf{A}'\hat{\boldsymbol{\Sigma}}_j\hat{\boldsymbol{\Xi}}^{-1}\mathbf{A})^2]$.

In Eqs. (5) and (6), the numerator can be simplified to $\mathrm{tr}(\mathbf{I}_{K-1}^2)+[\mathrm{tr}(\mathbf{I}_{K-1})]^2$, given that transforming $\boldsymbol{\Sigma}_j$ to $\boldsymbol{\Xi}^{-1/2}\boldsymbol{\Sigma}_j\boldsymbol{\Xi}^{-1/2}$, $\mathbf{A}'\mathbf{Q}\mathbf{A}\to\mathbf{I}_{K-1}$. Therefore, replacing $\mathbf{Q}_j$ in Eqs. (5) and (6) by its estimate $\hat{\boldsymbol{\Xi}}^{-1/2}\hat{\boldsymbol{\Sigma}}_j\hat{\boldsymbol{\Xi}}^{-1/2}$ and using the result that tr(AB) = tr(BA), the approximate d.f. simplifies to

$$\nu_e^* = \frac{(K-1)+(K-1)^2}{\sum_{j=1}^{J}1/(n_j-1)[\mathrm{tr}^2(c_j^{\bullet}\mathbf{A}'\hat{\boldsymbol{\Sigma}}_j\hat{\boldsymbol{\Xi}}^{-1}\mathbf{A})+\mathrm{tr}(c_j^{\bullet}\mathbf{A}'\hat{\boldsymbol{\Sigma}}_j\hat{\boldsymbol{\Xi}}^{-1}\mathbf{A})^2]}, \tag{7}$$

and

$$\nu_h^* = \frac{(K-1)+(K-1)^2}{\sum_{j=1}^{J}\{V\}+\mathrm{tr}^2\left(\sum_{j=1}^{J}c_j\mathbf{A}'\hat{\boldsymbol{\Sigma}}_j\hat{\boldsymbol{\Xi}}^{-1}\mathbf{A}\right)+\mathrm{tr}\left(\sum_{j=1}^{J}c_j\mathbf{A}'\hat{\boldsymbol{\Sigma}}_j\hat{\boldsymbol{\Xi}}^{-1}\mathbf{A}\right)^2}. \tag{8}$$

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 $c_j\mathbf{A}'\hat{\boldsymbol{\Sigma}}_j\mathbf{A}$ 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] Λ-criterion, the Hotelling–Lawley trace, and the Pillai–Bartlett trace statistics. Although their critical values have been widely tabled and charted (see [23]), in practice it is usual to obtain the level of significance by defining each of these statistics 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' Λ due to Rao [24] is chosen as the multivariate test statistic. According to this transformation, the interaction null hypothesis is rejected if

$$F_{\mathrm{MBF}_I}^* = \frac{1-\Lambda^{1/s^*}}{\Lambda^{1/s^*}}\left(\frac{\nu_2^*}{\nu_1^*}\right)\geq F^{1-\alpha}(\nu_1^*,\nu_2^*), \tag{9}$$

where $\Lambda = \det(\mathbf{E}^*)/\det(\mathbf{H}+\mathbf{E}^*)$, $s^* = [(l^2\nu_h^{*2}-4)/(l^2+\nu_h^{*2}-5)]^{1/2}$, $\nu_1^* = l\nu_h^*$, and $\nu_2^* = [\nu_e^*-(l-\nu_h^*+1)/2]s^*-(l\nu_h^*-2)/2$; here, det(·) denotes the determinant of the matrix.

If the null hypothesis of no interaction between groups and trials is rejected, the interpretation of tests of main effects is insufficient to understand the information of the data. When this occurs, test of possible 2×2 interactions (interaction contrasts or tetrad contrasts) can be used [28–31]. Under departures from the assumption of covariance homogeneity, tetrad contrasts results are obtained easily using the MBF procedure with appropriate $\mathbf{A}$ and $\mathbf{C}$ contrast matrices. Specifically, to test interaction contrasts using MBF approach, $\mathbf{C}=\mathbf{c}_{jj'}$ and $\mathbf{A}=\mathbf{a}_{kk'}$, where $\mathbf{c}_{jj'}$ is a $1\times J$ vector of coefficients that contrasts the jth and j'th between-subjects means and $\mathbf{a}_{kk'}$ is a $K\times 1$ vector of coefficients that contrasts the kth 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) tetrad 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 (FEW) for all possible 2×2 interactions, several post hoc procedures may be used. For instance, Lix and Keselman [31] found that the Hochberg [32] step-up Bonferroni, Schaffer [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 multisample 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 rank ordered, where $r = J^*\times K^*$, $J^* = J\times(J-1)$, and $K^* = K\times(K-1)$. Then, the largest probability is compared to α_{FEW}, where α_{FEW} is the family-wise error rate the researcher is willing to tolerate. If $p_r\leq\alpha_{\mathrm{FEW}}$, all hypotheses are rejected without further test; otherwise, the next largest probability is compared to $\alpha_{\mathrm{FEW}}/2$. If $p_{r-1}\leq\alpha_{\mathrm{FEW}}/2$, all hypotheses $H_1, \ldots, H_{r-1}$ are rejected. Continuing in this fashion, at any stage q, reject all H_q where $q'\leq q$, if $p_q\leq\alpha_{\mathrm{FEW}}/(r-q+1)$ 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 $\tilde{\mathbf{E}}(\tilde{\mathbf{H}}+\tilde{\mathbf{E}})^{-1}$, where

$$\tilde{\mathbf{H}} = (\mathbf{C}\tilde{\mathbf{B}}\mathbf{A})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}]^{-1}(\mathbf{C}\tilde{\mathbf{B}}\mathbf{A}), \tag{10}$$

and

$$\tilde{\mathbf{E}} = \left(\frac{\nu_e^{\bullet}}{\nu_h^{\bullet}}\right)\sum_{j=1}^{J}n_j^{-1}\mathbf{A}'(\boldsymbol{\Xi}^{1/2}\hat{\mathbf{Q}}_j\boldsymbol{\Xi}^{1/2})\mathbf{A}, \tag{11}$$

where $\mathbf{A}$ was defined before, $\tilde{\mathbf{B}} = \left[\sum_{j=1}^{J}(1/n_j)\right]^{1/2}\hat{\mathbf{B}}$, $\mathbf{C}\equiv\mathbf{c}$ is a $(J\times 1)$ vector consisting of all ones, $\boldsymbol{\Xi} = (\boldsymbol{\Sigma}_1/n_1+\cdots+\boldsymbol{\Sigma}_p/n_J)$,

and $\nu_h^\bullet = 1$. Extending the results reported by Krishnamoorthy and Yu [17] and Nel and van der Merwe [18], the distribution of $\sum_{j=1}^{J} n_j^{-1}\mathbf{A}'\mathbf{Q}_j\mathbf{A}$ can be approximated as a sum of Wishart distributions:

$$\sum_{j=1}^{J} n_j^{-1}\mathbf{A}'\mathbf{Q}_j\mathbf{A} \sim \mathrm{SW}_K\left(f_e^\bullet, \frac{1}{f_e^\bullet}\sum_{j=1}^{J} n_j^{-1}\mathbf{A}'\mathbf{Q}_j\mathbf{A}\right), \tag{12}$$

where the quantity $f_e^\bullet$ is given by

$$f_e^\bullet = \frac{\mathrm{tr}^2\left(\sum_{j=1}^{J} n_j^{-1}\mathbf{A}'\mathbf{Q}_j\mathbf{A}\right) + \mathrm{tr}\left(\sum_{j=1}^{J} n_j^{-1}\mathbf{A}'\mathbf{Q}_j\mathbf{A}\right)^2}{\sum_{j=1}^{J} 1/(n_j-1)[\mathrm{tr}^2(n_j^{-1}\mathbf{A}'\mathbf{Q}_j\mathbf{A}) + \mathrm{tr}(n_j^{-1}\mathbf{A}'\mathbf{Q}_j\mathbf{A})^2]}. \tag{13}$$

Replacing $\mathbf{Q}_j$ in Eq. (13) by its unbiased estimate $\hat{\boldsymbol{\Xi}}^{-1/2}\hat{\boldsymbol{\Sigma}}_j\hat{\boldsymbol{\Xi}}^{-1/2}$ the approximate d.f. can be written as

$$\nu_e^\bullet = \frac{(K-1)+(K-1)^2}{\sum_{j=1}^{J} 1/(n_j-1)\mathrm{tr}^2(n_j^{-1}\mathbf{A}'\hat{\boldsymbol{\Sigma}}_j\hat{\boldsymbol{\Xi}}^{-1}\mathbf{A}) + \mathrm{tr}\left(n_j^{-1}\mathbf{A}'\hat{\boldsymbol{\Sigma}}_j\hat{\boldsymbol{\Xi}}^{-1}\mathbf{A}\right)^2}. \tag{14}$$

The simplification at the Eq. (14) occurs because $\mathbf{A}'\mathbf{Q}\mathbf{A} \to \mathbf{I}_{K-1}$. 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

$$F^*_{\mathrm{MBF}_K} = \frac{1-\Lambda^{1/s}}{\Lambda^{1/s}}\left(\frac{\nu_2^\bullet}{\nu_1^\bullet}\right) \geq F^{1-\alpha}(\nu_1^\bullet, \nu_2^\bullet), \tag{15}$$

where $\Lambda = \det(\tilde{\mathbf{E}})/\det(\tilde{\mathbf{H}}+\tilde{\mathbf{E}})$, $s = [(l^2\nu_h^{\bullet 2}-4)/(l^2+\nu_h^{\bullet 2}-5)]^{1/2}$, $\nu_1^\bullet = l\nu_h^\bullet$, and $\nu_2^\bullet = [\nu_e^\bullet - (l-\nu_h^\bullet+1)/2]s - (l\nu_h^\bullet-2)/2$.

In turn, using the Mehrotra [35] 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 $\mathbf{E}^\bullet(\mathbf{H}+\mathbf{E}^\bullet)^{-1}$, where the hypothesis matrix is defined as in Eq. (2) with $\mathbf{A} \equiv \mathbf{a}$ (i.e., a $K \times 1$ vector with each element equal to one), the error matrix is

$$\mathbf{E}^\bullet = \left(\frac{\nu_e^\bullet}{\nu_h^\bullet}\right)\sum_{j=1}^{J} c_j\mathbf{A}'\Sigma_j\mathbf{A}, \tag{16}$$

and $\nu_e^\bullet$ and $\nu_h^\bullet$ are the approximate d.f. for $\mathbf{E}^\bullet$ and $\mathbf{H}$, 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

$$F_{\mathrm{MBF}_J} = \frac{\mathbf{H}}{\mathbf{E}^\bullet}\left(\frac{\nu_e^\bullet}{\nu_h^\bullet}\right) \geq F^{1-\alpha}(\nu_h^\bullet, \nu_e^\bullet). \tag{17}$$

It should be noted that the matrices $\mathbf{H}$ and $\mathbf{E}^\bullet$ are identical to the hypothesis and error sum of squares obtained employing a univariate Brown–Forsythe test with the numerator d.f. corrected.

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, $\mathbf{C} = \mathbf{c}_{jj'}$ and $\mathbf{A} \equiv \mathbf{a}$. The significance of the pair-wise comparisons for the within-subjects main effect can be probed in a similar manner, but with $\mathbf{C} \equiv \mathbf{c}$ (i.e., a $1 \times J$ vector with each element equal to one) and $\mathbf{A} = \mathbf{a}_{kk'}$.

At present there are numerous simultaneous or sequential multiple comparison procedures that maintain the FEW at or below its nominal α-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 Welsch'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.

3. Program description

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/docume/macrosas01.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 OMNIRESULTS, GROUPTEST, TIMETEST, INTERACTEST, and DEPVARTEST. 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

Table 1 – Data of CD4 cell count per mm^3 for 68 selected subjects from ACTG[a] study 193A

ID[b]	J[c]	K_1[d]	K_2[d]	K_3[d]	K_4[d]	K_5[d]	K_6[d]	ID[b]	J[c]	K_1[d]	K_2[d]	K_3[d]	K_4[d]	K_5[d]	K_6[d]	ID[b]	J[c]	K_1[d]	K_2[d]	K_3[d]	K_4[d]	K_5[d]	K_6[d]
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	01	0237	2	03	03	06	01	01	02	0084	4	11	05	13	14	19	09
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	14	23	14	16	0134	4	31	29	31	44	19	35
0175	1	22	25	17	16	32	23	0499	2	14	49	19	16	19	20	0148	4	72	102	63	36	22	50
0215	1	06	09	02	06	01	04	0557	2	34	56	32	21	14	09	0261	4	29	94	115	80	67	89
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	49	27	67	20	0325	4	10	30	10	10	10	10
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	08	20	13	10	0377	4	28	33	23	20	19	12
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	114	106	91	63	0484	4	39	70	52	54	40	25
0486	1	39	20	16	38	11	08	0247	3	42	21	110	51	30	23	0602	4	10	20	30	40	40	40
0597	1	20	10	10	10	10	10	0290	3	10	10	10	20	10	10	0609	4	15	160	130	150	100	90
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	10	10	30	10	0799	4	10	21	06	04	16	20
0876	1	20	20	20	10	10	10	0393	3	13	64	93	84	99	81	0811	4	89	129	229	166	123	91
0881	1	55	10	10	20	10	10	0440	3	13	15	21	20	11	12	0867	4	30	50	30	10	10	10
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	14	16	13	04	1015	4	30	161	213	145	138	162
0164	2	36	27	13	13	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	65	12	08	59	1131	4	06	20	30	53	06	21
0225	2	07	20	02	02	03	07	0953	3	30	100	146	97	63	26								

[a] AIDS clinical trial group.
[b] Variable that identifies the subject to which the record refers.
[c] Levels of the between-subjects factor.
[d] Levels of the within-subjects factor.

Table 2 – Summary of traditional univariate analysis

Hypotheses	MS_H[a]	MS_E[b]	F-value	d.f.[c]	p-Value
Groups	15989.81	5152.97	3.20	(3, 64)	0.0327
Trials	3316.53	416.47	7.95	(5, 320)	<0.0001
Groups × trials	1103.33	416.47	2.65	(15, 320)	0.0008

[a] Mean square for the hypothesis.
[b] Mean square for the error.
[c] Degrees of freedom.

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 Fitzmaurice et al. [20] from a study published 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 × trial interaction. A 0.05 significance level is assumed throughout the paper. For the between-subjects effect, $F = 2.89$, with 3 and 64 d.f. ($p = 0.0327$); 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 χ^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:

```
DATA DATARECORDED; INPUT GROUP Y1 Y2 Y3 Y4 Y5 Y6;
  CARDS;
PROC IML;
USE DATARECORDED;
NDV=1; /*NUMBER OF DEPENDENT VARIABLES*/
TESTOMNI=1; /*'1' OMNIBUS TESTS, '0' NO OMNIBUS TESTS*/
TESTGROUPS=1; /*'1' GROUPS PAIRWISE CONTRASTS, '0' NO
  GROUPS PAIRWISE*/
TESTTIME=1; /*'1' TRIALS PAIRWISE CONTRASTS, '0' NO TRI-
  ALS PAIRWISE*/
TESTINTERAC=1; /*'1' MULTIPLE INTERACTION CONTRASTS,
  '0' NO CONTRASTS*/
TESTDEPVAR=1; /*'1' UNIVARIATE TEST OF EACH DEPENDENT
  VAR, '0' NO TEST*/
RUN MBF;
```

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 38.82 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 no reject the null hypothesis at the 5% level of significance.

Table 3 – Summary of multivariate MBF analysis

Hypotheses	Wilks's Λ	d.f.$_W$[a]	F_{MBF}-value	d.f.$_R$[b]	p-Value
Groups	0.86	(1, 2.35, 47.51)	3.32	(2.35, 47.51)	0.0373
Trials	0.66	(5, 1.00, 38.82)	3.95	(5.00, 38.82)	0.0054
Groups × trials	0.62	(5, 13.16, 108.31)	1.72	(13.16, 108.31)	0.0666

[a] Degrees of freedom corresponding to Wilks's Λ criterion.
[b] Degrees of freedom corresponding to Rao's F-approximation.

Table 4 – Hochberg's adjusted p values for all possible pair-wise differences among the levels of the between-subjects and within-subjects factors

Test	d.f.$_1$	d.f.$_2$	F_{MBF}	p-Value	H-adjusted[a]	Decision
Pair-wise contrasts of the group						
J_1–J_4	1	23.84	8.56	0.0074	0.0445	Reject
J_1–J_3	1	14.82	7.16	0.0174	0.7279	Retain
J_1–J_2	1	21.31	5.04	0.0355	0.7279	Retain
J_2–J_4	1	33.15	1.35	0.2536	0.7279	Retain
J_2–J_3	1	23.08	0.66	0.4244	0.7279	Retain
J_3–J_4	1	31.13	0.12	0.7279	0.7279	Retain
Pair-wise contrasts of the trial						
K_1–K_2	1	56.02	16.37	1.6E−4	2.4E−3	Reject
K_1–K_3	1	43.72	13.11	7.5E−4	0.0106	Reject
K_3–K_6	1	24.17	13.18	1.3E−2	0.0172	Reject
K_2–K_6	1	33.49	10.80	2.3E−2	0.0286	Reject
K_3–K_5	1	34.57	8.69	5.7E−2	0.8240	Retain
K_1–K_4	1	44.34	7.42	9.1E−2	0.8240	Retain
K_2–K_5	1	48.77	6.95	0.0112	0.8240	Retain
K_3–K_4	1	33.90	7.18	0.0113	0.8240	Retain
K_4–K_6	1	25.70	6.40	0.0179	0.8240	Retain
K_2–K_4	1	53.23	4.23	0.0446	0.8240	Retain
K_1–K_5	1	45.90	2.75	0.1041	0.8240	Retain
K_4–K_5	1	45.60	2.69	0.1076	0.8240	Retain
K_5–K_6	1	32.56	2.21	0.1464	0.8240	Retain
K_1–K_6	1	40.06	0.56	0.4578	0.8240	Retain
K_2–K_3	1	51.54	0.50	0.8240	0.8240	Retain

J, levels of the between-subjects factor; K, levels of the within-subjects factor.
[a] Hochberg's adjusted p values.

After the overall null hypotheses referring to the groups and trials are rejected, the next step in the analysis is to decide which population means differ. As discussed above, both the pair-wise comparison tests for between-subjects marginal means and pair-wise comparison tests for measures repeated marginal means are affected by unequal covariance matrices across the grouping factor. However, it is possible to obtain robust tests for the pair-wise comparison hypotheses by using the MBF procedure and fitting the p values in step-up fashion for controlling the FWE. For pair-wise contrast and tetrad contrast, it can be verified that the MBF procedure and Johansen [9] test as given in Lix and Keselman [8] and Lix et al. [43] are equivalent; however, this would not be the case for $K > 2$.

Given that the null hypothesis of parallel profiles for groups is not rejected at 5% level significance, the researchers may average over trials and over groups, respectively, to test pair-wise contrast on group means and repeated measures means. The program generates automatically all possible pairs among the levels of the between-subjects and within-subjects factors. This part of the results has been included in Table 4.

For all possible contrast in between-subjects marginal means, the results reported in Table 4 show that applying Hochberg's sequentially rejective Bonferroni procedure one comparison is significant controlling FEW at a level no more 0.05: J_1 versus J_4. On the other hand, note that for the within-subjects marginal means, four comparisons are declared significant: K_1 versus K_2, K_1 versus K_3, K_3 versus K_6, and K_2 versus K_6. Because the interaction effect resulted no significant at 5%, the tetrad 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

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 lineal 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 assume 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 PR-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

[1] C. Ahn, J.E. Overall, S. Tonidandel, Sample size and power calculations in repeated measurement analysis, Comp. Meth. Prog. Biomed. 64 (2001) 121–124.
[2] R. Thiébaut, H. Jacqmin-Gadda, G. Chêne, C. Leport, D. Commenges, Bivariate linear mixed models using SAS proc MIXED, Comp. Meth. Prog. Biomed. 69 (2002) 249–256.
[3] R.K. Kowalchuk, H.J. Keselman, J. Algina, Repeated measures interaction test with aligned ranks, Multivariate Behav. Res. 38 (2003) 433–461.
[4] H. Huynh, Some approximate tests for repeated measurement designs, Psychometrika 43 (1978) 161–165.
[5] G.E.P. Box, Some theorems on quadratic forms applied in the study of analysis of variance problems. I. Effects of inequality of variance in the one-way classification, Ann. Math. Stat. 25 (1954) 290–403.
[6] C.L. Olson, Comparative robustness of six tests in multivariate analysis of variance, J. Am. Stat. Assoc. 69 (1974) 894–908.
[7] J. Algina, T.C. Oshima, An improved general approximation test for the main effect in a split-plot design, Br. J. Math. Stat. Psychol. 48 (1995) 149–160.
[8] L.M. Lix, H.J. Keselman, Approximate degrees of freedom tests: a unified perspective on testing for mean equality, Psychol. Bull. 117 (1995) 547–560.
[9] S. Johansen, The Welch–James approximation to the distribution of the residual sum of squares in a weighted linear regression, Biometrika 67 (1980) 85–92.
[10] J. Algina, H.J. Keselman, A power comparison of the Welch–James and improved general approximation tests in the split-plot design, J. Educ. Behav. Stat. 23 (1998) 152–169.
[11] G. Vallejo, P. Livacic-Rojas, A comparison of two procedures for analyzing small sets of repeated measures data, Multivariate Behav. Res. 40 (2005) 179–205.
[12] M.B. Brown, A.B. Forsythe, The small sample behavior of some statistics which test the equality of several means, Technometrics 16 (1974) 129–132.
[13] G. Vallejo, A.M. Fidalgo, P. Fernández, Effects of covariance heterogeneity on three procedures for analysing multivariate repeated measures designs, Multivariate Behav. Res. 36 (2001) 1–27.
[14] H. Akaike, A new look at the statistical model identification, IEEE Trans. Automat. Contr. 19 (1974) 716–723.
[15] G. Schwarz, Estimating the dimension of a model, Ann. Stat. 6 (1978) 461–464.
[16] G. Vallejo, M. Ato, Modified Brown–Forsythe test for analyzing repeated measures designs, Multivariate Behav., in press.
[17] K. Krishnamoorthy, J. Yu, Modified Nel and Van der Merwe test for the multivariate Behrens–Fisher problem, Stat. Probab. Lett. 66 (2004) 161–169.
[18] D.G. Nel, C.A. van der Merwe, A solution to the multivariate Behrens–Fisher problem, Commun. Stat. Theory Meth. 15 (1986) 3719–3735.
[19] SAS Institute, SAS/IML User's Guide, Version 9.1, SAS Institute Inc., Cary, NC, 2004.
[20] G.M. Fitzmaurice, N.M. Laird, J.H. Ware, Applied Longitudinal Analysis, Wiley, New Jersey, NJ, 2004 (Chapter 8).
[21] D.G. Nel, Tests for equality of parameter matrices in two multivariate linear models, J. Multivariate Anal. 61 (1997) 29–37.
[22] S.S. Wilks, Certain generalizations in the analysis of variance, Biometrika 24 (1932) 471–494.
[23] G.A.F. Seber, Multivariate Observations, Wiley, New York, NY, 1984 (Chapter 2).
[24] C.R. Rao, An asymptotic expansion of the distribution of Wilks's criterion, Bull. Int. Stat. Inst. 33 (1951) 177–180.
[25] K.C.S. Pillai, T.A. Mijares, On the moments of the trace of a matrix and approximations to its distribution, Ann. Math. Stat. 26 (1959) 117–121.
[26] J.T. Hughes, J.G. Saw, Approximating the percentage points of Hotelling's generalized T_0^2 statistic, Biometrika 59 (1972) 224–226.
[27] J.J. McKeon, F approximations to the distribution of Hotelling's T_0^2, Biometrika 61 (1974) 381–383.
[28] J.W. Tukey, The philosophy of multiple comparisons, Stat. Sci. 6 (1991) 100–116.
[29] R.J. Boik, The analysis of two-factor interactions in fixed effects linear models, J. Educ. Stat. 18 (1993) 1–40.

[30] N.H. Timm, Analysis of interactions: another look. Paper Presented at the Annual Meeting of the Am. Educ. Res. Assoc., New Orleans, 1994.
[31] L.M. Lix, H.J. Keselman, Interaction contrasts in repeated measures designs, Br. J. Math. Stat. Psychol. 49 (1996) 147–162.
[32] Y. Hochberg, A sharper Bonferroni procedure for multiple tests of significance, Biometrika 75 (1988) 800–802.
[33] J.P. Shaffer, Modified sequentially rejective multiple test procedures, J. Am. Stat. Assoc. 81 (1986) 826–831.
[34] Y. Hochberg, A.C. Tamhane, Multiple Comparison Procedures, Wiley, New York, NY, 1987 (Chapter 10).
[35] D.V. Mehrotra, Improving the Brown–Forsythe solution to the generalized Behrens–Fisher problem, Commun. Stat. Simul. 26 (1997) 1139–1145.
[36] J.P. Shaffer, Multiple hypothesis testing, Annu. Rev. Psychol. 46 (1995) 561–584.
[37] H.J. Keselman, Stepwise and simultaneous multiple comparison procedures of repeated measures' means, J. Educ. Behav. Stat. 19 (1994) 127–162.
[38] H.J. Keselman, L.M. Lix, Improved repeated measures stepwise multiple comparison procedures, J. Educ. Behav. Stat. 20 (1995) 83–99.
[39] R.K. Kowalchuk, H.J. Keselman, Mixed-model pairwise multiple comparisons of repeated measures means, Psychol. Meth. 6 (2001) 282–296.
[40] R.E. Welsch, Stepwise multiple comparison procedures, J. Am. Stat. Assoc. 72 (1977) 566–575.
[41] K. Henry, A. Erice, C. Tierney, H.H. Balfour, M.A. Fischl, A. Kmack, S.H. Liou, A. Kenton, M.S. Hirsch, J. Phair, A. Martinez, J.O. Kahn, A randomized, controlled, double-blind study comparing the survival benefit of four different reverse transcriptase inhibitor therapies (three-drug, two-drug, and alternating drug) for the treatment of advanced AIDS, J. Acq. Inmun. Def. Synd. 19 (1998) 339–349.
[42] N.H. Timm, Applied Multivariate Analysis, Springer-Verlag, New York, NY, 2002 (Chapter 3).
[43] L.M. Lix, H.J. Keselman, A.M. Hinds, Robust tests for the multivariate Behrens–Fisher problem, Comp. Meth. Prog. Biomed. 77 (2005) 129–139.
[44] D.B. Rubin, Multiple imputation after 18+ years, J. Am. Stat. Assoc. 91 (1996) 473–489 (with discussion).
[45] H.J. Keselman, J. Algina, R.K. Kowalchuk, R.D. Wolfinger, A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements, Commun. Stat. Simul. 27 (1998) 591–604.
[46] R.K. Kowalchuk, H.J. Keselman, J. Algina, R.D. Wolfinger, The analysis of repeated measurements with mixed-model adjusted *F* tests, Educ. Psychol. Meas. 64 (2004) 224–242.