Hacettepe Journal of Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 52 (4) (2023), 1066 – 1081
DOI : 10.15672/hujms.1176840

Research Article

Two approaches to extend classical MANOVA tests to the unequal covariances case

Samaradasa Weerahandi[1], Malwane M. A. Ananda[2], Osman Dag*[3]

[1] *X-Techniques, New Jersey, USA*
[2] *Department of Mathematical Sciences, University of Nevada, Las Vegas, USA*
[3] *Department of Biostatistics, Hacettepe University, Turkey*

Abstract

This article presents two approaches to extend classical MANOVA tests to the unequal covariances case. The first approach is illustrated by extending the classical Wilks test, which is valid only when covariances are equal. Such tests will be based on exact probabilities of certain extreme regions. We will also show how tests numerically equivalent to the parametric bootstrap tests could be easily obtained without using any bootstrap sampling arguments, so that resulting p-values are also based on exact probabilities of well defined extreme regions. Being systematic approaches, by taking similar approaches, researchers should be able to derive generalized tests in MANCOVA, higher-way MANOVA, and in RM MANOVA under heteroscedasticity, in which the parametric bootstrap type approaches run into difficulties.

Mathematics Subject Classification (2020). 62H15, 62H12, 62J10

Keywords. Generalized inference, parametric bootstrap, generalized p-value, fiducial inference

1. Introduction

MANOVA problems arise in a variety of applications ranging from Biometrics to studies of Public Health. When there are number of populations/treatments involving more than one response variable, one needs to first conduct MANOVA before proceeding to pairwise comparisons. In this article we propose two sets of solutions to MANOVA under unequal covariances, which are based on exact probabilities of certain extreme regions. The first set deals with the problem extending classical tests valid under equal covariances, and the second set deals with tests involving Bootstrap samples, which provide no interpretation concerning extreme regions of underlying sample spaces. Moreover, the parametric bootstrap (PB) approach may run into difficulties when good point estimators are not available, as the case in dealing with lifetime distributions and mixed effects models.

Although we do not address the [5] test due to using Lawley-Hotelling's type chi-squared approximations, we do utilize their approach in the generalized Wilks test presented in

*Corresponding Author.
Email addresses: weerahandi@aol.com (S. Weerahandi), malwane.ananda@unlv.edu (M. M. A. Ananda), osman.dag@hacettepe.edu.tr (O. Dag)
Received: 19.09.2022; Accepted: 07.02.2023

the next section. We will also derive a generalized tests that are numerically equivalent to the parametric bootstrap tests proposed by [8] and [21]. We will also not directly address the multivariate Behrence-Fisher problem.

A reader interested in that particular case is referred to the generalized tests discussed by [5] and [15], which has the advantage of being numerically equivalent to the Bayesian solution derived by [6] under the widely used non-informative prior.

Although there is no clear winner in the case of MANOVA under equal covariances, it seems that the Wilks test is preferred by many practitioners. Moreover, Wilks test statistic has exact distributional results as discussed by [7]. Therefore, that is the first test we will extend to the unequal covariances case using [5] arguments. Although we provide details of generalized Wilks test, we will also briefly discuss the approach researchers can take to generalize other classical tests such as the Roy's largest root test, the Lawley-Hotelling's test, and the Bartlett-Nanda-Pillai test.

By taking the two approaches proposed in this article, researchers are encouraged to derive generalized tests in higher-way MANOVA problems under heteroscedasticity, in Mixed Effects applications, and applications involving non-normal distributions (e.g. Lifetime distributions such as Weibull and Gamma), where the PB approach run into difficulties.

1.1. About our approach

Each of the two systematic approaches we propose in this article should pave the way for researchers to extend the classical tests to higher-way MANOVA, MANCOVA, and repeated measuare (RM) MANOVA. For example, the approach we propose in extending the Wilks test (GW1 test and GW2 test) to the case of unequal covariances, researchers should find useful in extending their favorite MANOVA tests such as Roy's largest root test, Lawley-Hotelling's test, and Bartlett-Nanda-Pillai test to case of unequal covariances.

The generalized parametric bootstrap (GPB) approach we introduce in this article also has a number of advantages over the PB approach. For example, the GPB approach does not require one to have good point estimators such as MLEsmaximum likelihood estimates (MLEs) of parameters to start with. This will make inferences such as RM MANOVA in mixed effects models possible by employing distribution theory developed by [20]. Moreover, extensions to non-normal distributions such as the Weibull, Lognormal, and Gamma that arise in lifetime distributions will also become possible. Moreover, all generalized tests have the advantage that they are based on exact probabilities of well defined extreme regions of the underlying sample spaces.

1.2. About generalized inference

There are multiple solutions to the MANOVA problem, as is the case with classical ANOVA under unequal variances. Here we present a particular class of solutions, which are all based on the generalized inference approach introduced by [14,16,17]. In one-liner, generalized tests are based on random quantities called Generalized Test Variables (GTV) that are functions of (i) observable random variables, (ii) their observed values, and (iii) unknown parameters, defined in such a way that

(a) the distribution of GTV is free of unknown parameters, and

(b) at the observed sample points, the observed value of GTV will have no unknown parameters under the null hypothesis. If the GTV is also monotonic for deviations from the null hypothesis, then it can be employed to define extreme regions, on which generalized p-values can be based.

Weerahandi [17] defined the notion of Generalized Pivotal Quantities, abbreviated as GPQs. In one-liner, a GPQ of a certain parameter is also a function of (i) observable random variables, (ii) their observed values, and (iii) unknown parameters, defined in such a way that

(a) its distribution does not depend on nuisance parameters, and

(b) at the observed sample points, its observed value become equal to the parameter, or a function of the parameter.

Weerahandi [15,17] argued that the practitioners should do the best with data at hand like Bayesians do and that generalized p-values are exact probabilities of well defined extreme regions. For the benefits of practitioners who insist on frequent properties, Roy and Bose [11] studied the small and large sample coverage of generalized confidence intervals, and proposed improvements for the former.

In ANOVA and MANOVA type problems, the underlying GTVs and GPQs are not unique, and so there are multiple extreme regions yielding exact probabilities, on which one can base tests. As pointed out by [19], what the authors present as generalized fiducial tests (e.g. [10]) and PB tests (e.g. [8]), can be derived by GTV/GPQ approach, but the converse is not true. To be specific, in its numerical form, the fiducial tests are a subset of generalized tests based on regular probability arguments. Moreover, while the PB approach can solve many parametric problems, that approach runs into difficulty in such problems as Mixed Effects models and those involving such distributions as lifetime distributions such as the Weibull and Gamma, where there are not even good point estimators having known distributions for the PB approach to work.

1.3. Notations

In this article, we will use bold face letters such as $\mathbf{Y}$ to denote a multivariate random vector representing a population of interest. The observed value of $\mathbf{Y}$ is denoted as $\mathbf{y}$. If the random vector $\mathbf{Y}$ has p components, then its components are denoted as: $\mathbf{Y} = (Y_1\ Y_2\ \cdots\ Y_p)'$. The mean vector of a population will be denoted by lowercase Greek letters such as $\boldsymbol{\mu}$ and the covariance matrix by uppercase Greek letters such as $\boldsymbol{\Sigma}$. If $\boldsymbol{\mu}$ represents the mean vector of a random vector $\mathbf{Y}$, then it is defined as $\boldsymbol{\mu} = E(\mathbf{Y})$. The covariance matrix of $\mathbf{Y}$ is defined as:

$$\boldsymbol{\Sigma} = \mathbf{Var}(\mathbf{Y}) = \mathbf{E}(\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})',$$

respectively.

When we have a random sample of size n, we use the notation $\overline{\mathbf{Y}}$ to denote the unbiased estimator of $\boldsymbol{\mu}$ and $\mathbf{S}$ to denote the unbiased estimator of the covariance matrix defined as

$$\mathbf{S} = \mathbf{W}/(n-1), \text{ where } \mathbf{W} = \sum(\mathbf{Y}_j-\overline{\mathbf{Y}})(\mathbf{Y}_j-\overline{\mathbf{Y}})'. \tag{1.1}$$

In the sequel below, for convenience of presentation, derivations involving square roots, and reading we assume that covariance matrices we deal with are positive definite. We can easily relax that assumption using Cholesky decomposition in palace of the square roots.

2. Solutions to heteroscedastic MANOVA

Consider the heteroscedastic MANOVA problem involving I populations, each of which is p dimensional,

$$\mathbf{Y}_{ij} \sim N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i); \quad i = 1, \ldots, I;\ j = 1, \ldots, n_i,$$

and consider the null hypothesis

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \ldots = \boldsymbol{\mu}_I. \tag{2.1}$$

We will assume all population and sample covariance matrices to be positive definite. As in [5], consider the transformed data matrix $\mathbf{X}_{ij} = g(\mathbf{s})^{-1/2}\mathbf{Y}_{ij} \sim N_p(\boldsymbol{\theta}, n_i\boldsymbol{\Lambda}_i)$, where $\boldsymbol{\theta} = g(\mathbf{s})^{-1/2}\boldsymbol{\mu}$, $g(\mathbf{s})$ is a positive definite matrix such as the identity matrix $g = \mathbf{I}$, or $g(\mathbf{s}) = \mathbf{s}_1/n_1 + \mathbf{s}_2/n_2 + \cdots + \mathbf{s}_I/n_I$, the one used by [5] for the $p = 2$ case.

Next, we present two approaches to obtain generalized tests to validate H_0, namely

1. Use the form of the desired classical test when covariances are known and handle then them by their GPQs,
2. Apply a suitable (as described below) scale invariant transformation to the original data and obtain a GTV.

2.1. Approach 1 illustrated by two generalized Wilks tests

In this section we introduce an approach that should work in extending classical MANOVA tests such as Roy's largest root test, Lawley-Hotelling's test, Bartlett-Nanda-Pillai test, and Wilk's test to the case of unequal covariances. The approach one should be able to extend for higher-way MANOVA as well. This approach basically starts out with any classical test when the covariances are known, and then tackling the covariances by the GPQs of the type suggested by the following proposition:

Proposition 2.1. *Suppose the random vector* $\mathbf{V}$ *has a Wishart distribution of the form*

$$\mathbf{V} \sim W(m, \mathbf{\Lambda}),$$

and $\mathbf{v}$ *be its observed value. If* $\mathbf{\Lambda}$ *is positive definite, then,*

$$\tilde{\mathbf{\Lambda}} = \mathbf{v}^{1/2}\left((\mathbf{v}^{-1/2}\mathbf{\Lambda}\mathbf{v}^{-1/2})^{1/2}(\mathbf{v}^{1/2}\mathbf{V}^{-1}\mathbf{v}^{1/2})(\mathbf{v}^{-1/2}\mathbf{\Lambda}\mathbf{v}^{-1/2})^{1/2}\right)\mathbf{v}^{1/2} \tag{2.2}$$

is a GPQ for $\mathbf{\Lambda}$ *distributed as* $\tilde{\mathbf{\Lambda}} = \mathbf{v}^{1/2}\mathbf{U}^{-1}\mathbf{v}^{1/2}$*, where* $\mathbf{U} \sim \mathbf{W}(m, \mathbf{I})$.

Proof. From properties of the Wishart distribution it follows that

$$\mathbf{v}^{-1/2}\mathbf{V}\mathbf{v}^{-1/2} \sim W(m, \mathbf{v}^{-1/2}\mathbf{\Lambda}\mathbf{v}^{-1/2}) \text{ and hence} \tag{2.3}$$
$$\mathbf{U} = (\mathbf{v}^{-1/2}\mathbf{\Lambda}\mathbf{v}^{-1/2})^{-1/2}(\mathbf{v}^{-1/2}\mathbf{V}\mathbf{v}^{-1/2})(\mathbf{v}^{-1/2}\mathbf{\Lambda}\mathbf{v}^{-1/2})^{-1/2} \sim W(m, \mathbf{I})$$

thus implying the distribution of $\tilde{\mathbf{\Lambda}}$, which is distributed free of unknown parameters. Moreover, at the observed $\mathbf{v}$ of $\mathbf{V}$, $\tilde{\mathbf{\Lambda}}$ becomes equal to $\mathbf{\Lambda}$, thus proving the proposition. □

Now in order to utilize Proposition 2.1 in extending the classical tests such as the Wilks to the unequal covariances case, define

$$\mathbf{\Lambda}_i = g(\mathbf{s})^{-1/2}\frac{\mathbf{\Sigma}_i}{n_i}g(\mathbf{s})^{-1/2}, \qquad i = 1,\ 2\ldots, I.$$

Testing of the hypothesis (2.1) can be based on the independent random variables

$$\overline{\mathbf{X}}_i \sim N(\boldsymbol{\theta}, \mathbf{\Lambda}_i), \qquad \text{under } H_0 \tag{2.4}$$

and

$$\mathbf{V}_i = (n_i - 1)g(\mathbf{s})^{-1/2}\frac{\mathbf{S}_i}{n_i}g(\mathbf{s})^{-1/2} \sim W_p(n_i - 1, \mathbf{\Lambda}_i).$$

If covariances were known, one could establish a testing procedures by considering a between group of weighted sum of cross products of the form

$$\widetilde{\mathbf{T}}(\mathbf{X}; \mathbf{\Lambda}) = \sum_{i=1}^{I}(\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}})\mathbf{W}_i(\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}})', \tag{2.5}$$

where $\mathbf{W}_i$ is a suitable weight function and $\overline{\overline{\mathbf{X}}} = \widehat{\boldsymbol{\theta}}$ is an estimate the common mean vector $\boldsymbol{\theta}$ under H_0 obtained as

$$\overline{\overline{\mathbf{X}}} = \left(\sum_{i=1}^{I}\mathbf{\Lambda}_i^{-1}\right)^{-1}\left(\sum_{i=1}^{I}\mathbf{\Lambda}_i^{-1}\overline{\mathbf{X}}_i\right). \tag{2.6}$$

Two suitable weight functions are

$$\mathbf{W}_i = \mathbf{\Lambda}_i^{-1} \text{ and } \mathbf{W}_i = \mathbf{U}_i\mathbf{U}_i^{'}, \text{ where } \mathbf{U}_i = \left(\sum_{i=1}^{I}\mathbf{\Lambda}_i\right)\mathbf{\Lambda}_i^{-1}. \tag{2.7}$$

Gamage et al. [5] suggested using weights of the former set of weights, but that choice does not reduce to their generalized test for the Behrence-Fisher problem when $p = 2$. We will denote the generalized Wilks test developed below with that weight as $GW1$. The test based on the alternative second set of weights in (2.7) having that property is denoted as $GW2$.

We will denote the generalized Wilks test developed below with the second set of weights as $GW2$.

If covariances were known, we can indeed employ (2.5) to obtain classical type test statistics since the common mean term in the expression $\overline{\mathbf{X}_i} \sim \boldsymbol{\theta} + \mathbf{\Lambda}_i^{1/2}\mathbf{Z}_i$ cancels out under H_0, and since $\mathbf{\Delta}_i = (\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}})$ can then be expressed as

$$\begin{aligned} \mathbf{\Delta}(\mathbf{Z},\mathbf{\Lambda})_i &= \mathbf{X}(\mathbf{Z},\mathbf{\Lambda})_i - \overline{\mathbf{X}}(\mathbf{Z},\mathbf{\Lambda}), \text{where } \mathbf{X}(\mathbf{Z},\mathbf{\Lambda})_i = \mathbf{\Lambda}_i^{1/2}\mathbf{Z}_i, \\ \overline{\mathbf{X}}(Z,\mathbf{\Lambda}) &= \left(\sum_{i=1}^{I}\mathbf{\Lambda}_i^{-1}\right)^{-1}\left(\sum_{i=1}^{I}\mathbf{\Lambda}_i^{-1}\mathbf{X}(\mathbf{Z},\mathbf{\Lambda})_i\right), \text{ and } \mathbf{Z}_i \sim N(\mathbf{0},\mathbf{I}_p). \end{aligned} \tag{2.8}$$

When the covariances are unknown it is tempting to obtain a test by replacing $\mathbf{\Lambda}_i$ by an estimate, but that will lead to a test statistic having an unknown distribution. Nevertheless, any desired classical test, including the Wilks test, can be obtained by replacing $\mathbf{\Lambda}_i$ by its GPQ derived by Proposition 2.1, namely $\tilde{\mathbf{\Lambda}}_i$.

Corollary 2.2. *The random quantity*

$$\widetilde{\mathbf{T}}(\mathbf{X}(\mathbf{Z},\tilde{\mathbf{\Lambda}});\tilde{\mathbf{\Lambda}}) = \sum_{i=1}^{I}\mathbf{\Delta}(\mathbf{Z},\tilde{\mathbf{\Lambda}})_i\mathbf{W}_i(\tilde{\mathbf{\Lambda}})\mathbf{\Delta}(\mathbf{Z},\tilde{\mathbf{\Lambda}})_i', \tag{2.9}$$

is a GTV suitable for testing H_0.

Proof. Since $\tilde{\mathbf{\Lambda}}_i$ is GPQ for $\mathbf{\Lambda}_i$,

1. its distribution does not have any unknown parameters,
2. at the observed sample point it reduces to

$$\widetilde{\mathbf{t}} = \widetilde{\mathbf{T}}_{obs} = \sum_{i=1}^{I}(\overline{\mathbf{x}}_i - \overline{\overline{\mathbf{x}}})\mathbf{W}_i(\tilde{\mathbf{\Lambda}})(\overline{\mathbf{x}}_i - \overline{\overline{\mathbf{x}}})',$$

a quantity involving no unknown parameters. Moreover, it tends to take greater values for greater deviations from the null hypothesis. Therefore $\widetilde{\mathbf{T}}(\mathbf{X}(\mathbf{Z},\tilde{\mathbf{\Lambda}});\tilde{\mathbf{\Lambda}})$ is indeed generalized test variable. □

For example, if one is interested in extending the classical Lawley-Hostelings test, one would use the sum of Eigen values of $\widetilde{\mathbf{T}}$ and $\widetilde{\mathbf{t}}$. Of particular interest in this article is the generalized Wilks tests $GW1$ and $GW2$ tests based on the p-value

$$p = \Pr(|\mathbf{I}_p + \widetilde{\mathbf{T}}| \geq |\mathbf{I}_p + \widetilde{\mathbf{t}}|). \tag{2.10}$$

The p-value based on $\widetilde{\mathbf{T}}$ can be computed by the Monte Carlo simulation method. This is accomplished by simulating the Wishart random matrices using independent standard normal random variates and the well known result:

$$\text{If } \mathbf{Z}_j \sim N(\mathbf{0},\mathbf{I}_p), \quad j = 1,\ 2,\ldots,J, \quad \text{then } \sum \mathbf{Z}_j\mathbf{Z}_j' \sim W_p(J,\mathbf{I}_p).$$

The simulation is then carried out in the following steps:

1. Generate a large sample of random numbers from $\mathbf{Z}_i = \mathbf{\Lambda}_i^{-1/2}(\overline{\mathbf{X}}_i - \theta) \sim N(\mathbf{0},\mathbf{I})$ and

a large sample from each of the distributions of $\mathbf{R}_i \sim W_p(n_i - 1, \mathbf{I}_p)$ based on sets of independent standard normal random numbers.

2. Specify the weight function from alternatives in (2.7) depending on generalized Wilks test $GW1$ or $GW2$ is desired.

3. For each set of simulated samples from $\mathbf{R}_i$, $i = 1, 2, \ldots, I$, replace $\mathbf{\Lambda}_i$ by the simulated value of $\tilde{\mathbf{\Lambda}}_i = \mathbf{v}_i^{1/2}\mathbf{R}_i^{-1}\mathbf{v}_i^{1/2}$, and compute

$$\widetilde{\mathbf{t}} = \sum_{i=1}^{I}(\overline{\mathbf{x}}_i - \overline{\overline{\mathbf{x}}})\mathbf{W}(\tilde{\mathbf{\Lambda}})_i(\overline{\mathbf{x}}_i - \overline{\overline{\mathbf{x}}})', \tag{2.11}$$

where $\overline{\mathbf{x}}_i = g(\mathbf{s})^{-1/2}\overline{\mathbf{y}}_i$ is computed using the actual data.

4. For each set of simulated samples from $\mathbf{R}_i, \mathbf{Z}_i$; $i = 1, 2, \ldots, I$, also compute

$$\widetilde{\mathbf{T}} = \sum_{i=1}^{I}\mathbf{\Delta}(\mathbf{Z}, \tilde{\mathbf{\Lambda}})_i\mathbf{W}_i(\tilde{\mathbf{\Lambda}})\mathbf{\Delta}(\mathbf{Z}, \tilde{\mathbf{\Lambda}})_i'.$$

5. Repeat the above steps for a large set of simulated samples and compute the p-value based on the fraction of pairs for which $(|\mathbf{I}_p + \widetilde{\mathbf{T}}| \geq |\mathbf{I}_p + \widetilde{\mathbf{t}}|)$.

Then, one can also carry out multiple comparisons and compute confidence intervals for the desired differences in mean vectors as described by [5].

2.2. Approach 2 illustrated by generalized PB test for MANOVA

In this article, however we confine our attention to normal-theory based MANOVA, PB approach works, as derived by [9] and [22], followed by [19], who showed that bootstrap arguments are not needed to obtain such tests. In that regard, in this section we will derive a second generalized test without using any bootstrap notations or arguments. Although our derivation and formulas look different from that of [8], (i) they are numerically equivalent, (ii) one should find our approach easier in extending results to higher-way MANOVA. This approach simply apply an invariant scale transformation to raw data so that the resulting between group sum of cross products can be tackled by GPQs as we did in the previous section. In this section, for the ease of presentation, we will use matrix notations without subscript i to stand for the set of all terms involving the subscript.

To obtain such tests, consider the between group sum of cross products,

$$\begin{aligned} T(\mathbf{Y}; \tilde{\mathbf{\Sigma}}) &= \sum_{i=1}^{I}((\overline{\mathbf{Y}}_i - \boldsymbol{\mu}_i) - \mathbf{Y}_0(\overline{\mathbf{Y}}))'\tilde{\mathbf{\Sigma}}_i^{-1}((\overline{\mathbf{Y}}_i - \boldsymbol{\mu}_i) - \mathbf{Y}_0(\overline{\mathbf{Y}}) \qquad (2.12) \\ &= \sum_{i=1}^{I}g_i(\mathbf{Y}, \tilde{\mathbf{\Sigma}})'\tilde{\mathbf{\Sigma}}_i^{-1}g_i(\mathbf{Y}, \tilde{\mathbf{\Sigma}}), \end{aligned}$$

where $\tilde{\mathbf{\Sigma}}_i = \mathbf{\Sigma}_i/n_i$, $g_i(\mathbf{Y}, \tilde{\mathbf{\Sigma}}) = (\overline{\mathbf{Y}}_i - \boldsymbol{\mu}_i) - \mathbf{Y}_0(\overline{\mathbf{Y}}; \boldsymbol{\mu}, \tilde{\mathbf{\Sigma}})$, and

$$\mathbf{Y}_0(\overline{\mathbf{Y}}) = \mathbf{Y}_0(\overline{\mathbf{Y}}; \boldsymbol{\mu}, \tilde{\mathbf{\Sigma}}) = \left(\sum_{i=1}^{I}\tilde{\mathbf{\Sigma}}_i^{-1}\right)^{-1}\left(\sum_{i=1}^{I}\tilde{\mathbf{\Sigma}}_i^{-1}(\overline{\mathbf{Y}_i} - \boldsymbol{\mu}_i)\right), \tag{2.13}$$

a weighted difference between the population and sample means to play the role in the second term of the typical decomposition of between group sum of cross products. It is possible to decompose $T(\mathbf{Y}; \tilde{\mathbf{\Sigma}})$ as a difference of two terms, as in [8], but doing so will make the extension of results to higher-way MANOVA cumbersome.

Lemma 2.3. *Let $\mathbf{t}_i$ be the square root of the matrix $\tilde{\mathbf{s}}_i$ (more generally $\mathbf{t}_i$ is obtained from the Cholesky decomposition $\tilde{\mathbf{s}}_i = \mathbf{t}_i\mathbf{t}_i'$). Then, application of the scale transformation $\mathbf{t}_i\tilde{\mathbf{\Sigma}}_i^{-1/2}$ to summary statistics (or raw data) will transform*

1. $\overline{\mathbf{Y}}_i - \boldsymbol{\mu}_i$ *to* $\mathbf{t}_i\tilde{\boldsymbol{\Sigma}}_i^{-1/2}(\overline{\mathbf{Y}}_i - \boldsymbol{\mu}_i) = \mathbf{t}_i\mathbf{Z}_i$
2. $\tilde{\mathbf{S}}_i$ *to* $\mathbf{s}_{Ui} = \mathbf{t}_i\tilde{\boldsymbol{\Sigma}}_i^{-1/2}\tilde{\mathbf{S}}_i\tilde{\boldsymbol{\Sigma}}_i^{-1/2}\mathbf{t}_i' = \mathbf{t}_i\mathbf{U}_i\mathbf{t}_i'/(n_i-1)$,
3. the observed value $\tilde{\mathbf{s}}_i$ *of* $\tilde{\mathbf{S}}_i$ *to* $\tilde{\mathbf{s}}_i$ *itself, and*
4. the observed value $(\overline{\mathbf{y}}_i - \boldsymbol{\mu}_i)$ *of* $(\overline{\mathbf{Y}}_i - \boldsymbol{\mu}_i)$ *to* $(\overline{\mathbf{y}}_i - \boldsymbol{\mu}_i)$ *itself,*
where $\tilde{\mathbf{S}}_i = \mathbf{S}_i/n_i$, $\tilde{\boldsymbol{\Sigma}}_i = \boldsymbol{\Sigma}_i/n_i$, $\mathbf{Z}_i \sim N(\boldsymbol{0},\mathbf{I}_p)$, *and* $\mathbf{U}_i \sim W_p(n_i-1,\mathbf{I}_p)$.

Proof. The distribution $\overline{\mathbf{Y}}_i - \boldsymbol{\mu}_i \sim N(\mathbf{0},\tilde{\boldsymbol{\Sigma}}_i)$ implies $\tilde{\boldsymbol{\Sigma}}_i^{-1/2}(\overline{\mathbf{Y}}_i - \boldsymbol{\mu}_i) \sim N(\mathbf{0},\mathbf{I}_p)$, and hence assertion 1 of Lemma follows. The distribution $(n_i-1)\tilde{\mathbf{S}}_i \sim W_p\left(n_i-1,\tilde{\boldsymbol{\Sigma}}_i\right)$ implies $\tilde{\boldsymbol{\Sigma}}_i^{-1/2}\tilde{\mathbf{S}}_i\tilde{\boldsymbol{\Sigma}}_i^{-1/2} \sim W_p(n_i-1,\mathbf{I}_p)$, and hence assertion 2 follows.

Moreover, the above scale transformation transforms $\tilde{\mathbf{S}}_i$ to $\mathbf{t}_i\tilde{\boldsymbol{\Sigma}}_i^{-1/2}\tilde{\mathbf{S}}_i\tilde{\boldsymbol{\Sigma}}_i^{-1/2}\mathbf{t}_i'$ and so its observed value is $\mathbf{t}_i\tilde{\boldsymbol{\Sigma}}_i^{-1/2}\tilde{\mathbf{s}}_i\tilde{\boldsymbol{\Sigma}}_i^{-1/2}\mathbf{t}_i'$, thus implying that the transformation does not affect the observed value of $\mathbf{S}_i$, because the transformation does not involve random variables. The same argument works for the observed value $(\overline{\mathbf{y}_i}-\boldsymbol{\mu}_i)$ of $(\overline{\mathbf{Y}}_i-\boldsymbol{\mu}_i)$, and hence assertion 4 follows. □

Now we are in a position to derive a GTV for testing H_0. In the classical approach to inference, one may wish to replace $\tilde{\boldsymbol{\Sigma}}_i$ by $\tilde{\mathbf{s}}_i$, the observed value of $\tilde{\mathbf{S}}_i$, and $\boldsymbol{\mu}_i$ by $\overline{\mathbf{y}}_i$, the observed sample means, but the result would lead to an approximate test with an unknown distribution. In generalized fiducial approach, one also replaces $\tilde{\boldsymbol{\Sigma}}_i$ by $\tilde{\mathbf{s}}_i$, and replaces $\boldsymbol{\mu}_i$ by its GPQ, which indeed leads to a GTV having a distribution free of unknown parameters. In both cases, under H_0, the observed value of T reduces to

$$T_{obs} = \sum_{i=1}^{I}(\overline{\mathbf{y}}_i - \mathbf{y}_0)'\tilde{\mathbf{s}}_i^{-1}(\overline{\mathbf{y}}_i - \mathbf{y}_0) \tag{2.14}$$

at the observed $\overline{\mathbf{y}}$ of $\overline{\mathbf{Y}}$, where $\mathbf{y}_0 = \left(\sum_{i=1}^{I}\tilde{\mathbf{s}}_i^{-1}\right)^{-1}\left(\sum_{i=1}^{I}\tilde{\mathbf{s}}_i^{-1}\overline{\mathbf{y}_i}\right)$.

But in ANOVA type problems, there are multiple generalized tests satisfying the required conditions (cf. [15]) of a GTV. We can do better (cf. [1]), by considering the between group sum of cross products that does not replace $\tilde{\mathbf{S}}_i$ by $\tilde{\mathbf{s}}_i$ as the fiducial approach does, but rather tackling $\tilde{\boldsymbol{\Lambda}}_i$ by $\tilde{\mathbf{S}}_i$, and considering

$$\begin{aligned} T(\mathbf{Y};\mathbf{S}) &= \sum_{i=1}^{I}((\overline{\mathbf{Y}}_i - \boldsymbol{\mu}_i) - \mathbf{Y}_0(\overline{\mathbf{Y}}))'\tilde{\mathbf{S}}_i^{-1}((\overline{\mathbf{Y}}_i - \boldsymbol{\mu}_i) - \mathbf{Y}_0(\overline{\mathbf{Y}}) \\ &= \sum_{i=1}^{I} g_i(\mathbf{Y},\tilde{\mathbf{S}})'\tilde{\mathbf{S}}_i^{-1}g_i(\mathbf{Y},\tilde{\mathbf{S}}), \end{aligned} \tag{2.15}$$

where $g_i(\mathbf{Y},\tilde{\mathbf{S}}) = (\overline{\mathbf{Y}}_i - \boldsymbol{\mu}_i) - \mathbf{Y}_0(\overline{\mathbf{Y}};\boldsymbol{\mu},\tilde{\mathbf{S}})$, and

$$\mathbf{Y}_0(\overline{\mathbf{Y}}) = \mathbf{Y}_0(\overline{\mathbf{Y}};\boldsymbol{\mu},\tilde{\mathbf{S}}) = \left(\sum_{i=1}^{I}\tilde{\mathbf{S}}_i^{-1}\right)^{-1}\left(\sum_{i=1}^{I}\tilde{\mathbf{S}}_i^{-1}(\overline{\mathbf{Y}}_i - \boldsymbol{\mu}_i)\right), \tag{2.16}$$

and then by deriving a GPQ/GTV for $T(\mathbf{Y};\mathbf{S})$ itself, as follows.

Proposition 2.4. *The scale transformation of data as in Lemma 2.3 transforms* $T(\mathbf{Y};\mathbf{S})$ *to*

$$\tilde{T}(\mathbf{Y};\tilde{\mathbf{S}},\tilde{\mathbf{s}}) = \sum_{i=1}^{I}(n_i-1)\tilde{g}_i(\mathbf{Y},\tilde{\mathbf{S}},\tilde{\mathbf{s}})'\left(\mathbf{t}_i\mathbf{U}_i\mathbf{t}_i'\right)^{-1}\tilde{g}_i(\mathbf{Y},\tilde{\mathbf{S}},\tilde{\mathbf{s}})), \tag{2.17}$$

a GTV appropriate for testing H_0, *where* $\tilde{g}_i(\mathbf{Y},\tilde{\mathbf{S}},\tilde{\mathbf{s}})) = \mathbf{t}_i\mathbf{Z}_i - \mathbf{Y}_{0Z}$, *and*

$$\mathbf{Y}_{0Z} = \left(\sum_{i=1}^{I}\mathbf{s}_{Ui}^{-1}\right)^{-1}\left(\sum_{i=1}^{I}\mathbf{s}_{Ui}^{-1}\mathbf{t}_i\mathbf{Z}_i\right).$$

Proof. The scale transformation of data as in Lemma 2.3 transforms $\tilde{\mathbf{S}}_i$ terms appearing in $T(\mathbf{Y};\mathbf{S})$ to $\mathbf{s}_{Ui} = \mathbf{t}_i\mathbf{U}_i\mathbf{t}_i'/(n_i-1)$, and $(\overline{\mathbf{Y}}_i-\boldsymbol{\mu}_i)$ terms to $\mathbf{t}_i\mathbf{Z}_i$, which are both distributed free of unknown parameters. Hence, $\tilde{T}(\mathbf{Y};\tilde{\mathbf{S}},\tilde{\mathbf{s}})$ is distributed free of unknown parameters.

Moreover, under H_0, $\boldsymbol{\mu}_i$ terms in (2.15) and (2.16) cancel out, the observed value of $\mathbf{Y}_0(\overline{\mathbf{Y}};\boldsymbol{\mu},\tilde{\mathbf{S}})$ reduces to

$$\mathbf{Y}_0(\overline{\mathbf{y}};\mathbf{0},\tilde{\mathbf{s}}) = \left(\sum_{i=1}^{I}\tilde{\mathbf{s}}_i^{-1}\right)^{-1}\left(\sum_{i=1}^{I}\tilde{\mathbf{s}}_i^{-1}\overline{\mathbf{y}}_i\right),$$

and thus the observed value of $\tilde{T}(\mathbf{Y};\tilde{\mathbf{S}},\tilde{\mathbf{s}})$ becomes equal to the same t_{obs} given by (2.14). Since, deviations of individual population means from the null hypothesis of equal means tend to increase the between group sum of squares, $\tilde{T}(\mathbf{Y};\tilde{\mathbf{S}},\tilde{\mathbf{s}})$ is indeed a GTV appropriate for testing H_0. □

The generalized p-value based on the GTV suggested by Proposition 2.4 is computed as

$$p \;=\; Pr(\tilde{T}(\mathbf{Y};\tilde{\mathbf{S}},\tilde{\mathbf{s}}) \geq T_{obs}). \tag{2.18}$$

The p-value given by (2.18) can be computed by Monte Carlo method as follows:

1. Generate a large sample of size M random numbers for the set of random variables $\mathbf{Z}_i \sim N(\mathbf{0},\mathbf{I}_p)$ for $i = 1,\cdots,I$.

2. Generate a large sample of M random numbers from the set of standard Wishart distributions $\mathbf{U}_i$ appearing in the formula (2.17)

3. compute M simulated samples of $\tilde{T}(\mathbf{Y};\tilde{\mathbf{S}},\tilde{\mathbf{s}})$ using the formula (2.17)

4. Estimate the p-value of (2.18) by the fraction of times the inequality $\tilde{T}(\mathbf{Y};\tilde{\mathbf{S}},\tilde{\mathbf{s}}) \geq T_{obs}$ satisfied.

2.3. A case for extending GPV test to MANOVA

The main advantage of the GPB over PB, is that, being a systematic approach, the former approach one can easily extending results to the higher-way MANOVA, MANCOVA, and especially RM MANOVA, in which PB approach fails due to poor small sample properties of MLEs in Mixed Effects models. Moreover, the same drawback of PB arises with non-normal distributions such as those used in lifetime distributions.

Moreover, the approach we took in developing GPB should also suggest how researchers could extend the generalized p-value (GPV) method introduced by [12]. The GPV test currently available for ANOVA under heteroscedasticity is not only as good as the PB test, but also has the added advantage that it never exceeds the intended type-I error. The GPV test is also a test based on GTVs, but overcome the drawbacks of the GF test (Generalized F-test; cf. [15]) by reformulating the ANOVA problem as an equivalent problem involving pairwise contrasts.

3. Approach 2 illustrated by two-way MANOVA

In the context of multivariate analysis, it is beyond the scope of this article introducing generalized tests based on exact probabilities to develop tests for higher-way heteroscedastic two-way MANOVA. Nevertheless, we do so just for the fully cross classified two-way MANOVA to show how easy it is to do so. There are exact and approximate tests (unnecessarily) available for fully cross classified designs and nested designs. For example, the test developed in [23] is an approximate degrees of freedom solution of the Hotelling T-Squared type. Of course there is nothing wrong employing approximate tests if they do not have serious Type-I error issues. But the in-depth simulation study reported by [8] clearly demonstrated the serious type-I error issues of approximate tests even when the assumption of normality is satisfied. Therefore, with the aid of the second illustration we outline below, we encourage researchers to develop higher-way MANOVA and RM MANOVA under heteroscedasticity.

3.1. Dealing with weights in two-way MANOVA

In two-way ANOVA and beyond, one must use a set of suitable weights to make the parameters of the model identifiable, as pointed out by [2] and [4]. But for the mathematical tractability and for the convenience of handling main effects, the weights used by [23], such as the constant weights and weights that depend only on sample sizes, but do not depend on covariances.

Of course that does not make any difference in testing interactions of the model, but appropriate weights do matter in testing the main effects. Proper handling of weights is required to assure tests are affine invariant since the testing problem is affine invariant. The weights used by [23] does not make the sum of between group cross products scale invariant. Of course that is fine when we compare a test statistic against its observed value. However, if a researcher wishes to develop an approximate test to avoid Monte Carlo integration, it is desirable to make the test statistic itself scale invariant as the case with all classical tests involving mixtures of Normal, Chi-squared, and F random variables, which are all scale free.

3.2. Two-way MANOVA under unequal covariances

To further illustrate Approach 2 to develop tests in higher-way MANOVA, consider the two-way MANOVA problem

$$\mathbf{Y}_{ijk} = \boldsymbol{\theta}_{ij} + \boldsymbol{\epsilon}_{ijk} = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \boldsymbol{\gamma}_{ij} + \boldsymbol{\epsilon}_{ijk}, \tag{3.1}$$

where $\boldsymbol{\mu}_0$ is the grand mean vector, $\boldsymbol{\alpha}_i$ (i=1,2,...,a) is the effect vector of the i^{th} level of factor A, $\boldsymbol{\beta}_j$ (j=1,2,...,b) is the effect vector of the j^{th} level of factor B, and $\boldsymbol{\gamma}_{ij}$ is the interaction effect of the factor level A_i and the factor level B_j. Let a p-variare random sample of size n_{ij} is available from $(i,j)^{\text{th}}$ cell, $i = 1,\ldots,a;\ \ j = 1,\ldots,b.$ Then $\mathbf{Y}_{ijk}, i = 1,\ldots,a;\ \ j = 1,\ldots,b;\ k = 1,\ldots,n_{ij}$ represent these random vectors. The sample mean vector and the sample covariance matrix of the $(i,j)^{\text{th}}$ cell are denoted by $\overline{\mathbf{Y}}_{ij} = \Sigma_{k=1}^{n_{ij}}\mathbf{Y}_{ijk}/n_{ij}$ and $\mathbf{S}_{ij} = \Sigma_{k=1}^{n_{ij}}\left(\mathbf{Y}_{ijk} - \overline{\mathbf{Y}}_{ij}\right)\left(\mathbf{Y}_{ijk} - \overline{\mathbf{Y}}_{ij}\right)'/(n_{ij}-1)$, $i = 1,\ldots,a;\ \ j = 1,\ldots,b$. Furthermore, let $\overline{\mathbf{y}}_{ij}$ and $\mathbf{s}_{ij}$ be the observed values of $\overline{\mathbf{Y}}_{ij}$ and $\mathbf{S}_{ij},\ i = 1,\ldots,a;\ \ j = 1,\ldots,b.$

In order to assure $\boldsymbol{\mu}_0$, $\boldsymbol{\alpha}_i$, $\boldsymbol{\beta}_j$, and $\boldsymbol{\gamma}_{ij}$ are uniquely defined, additional constraints are necessary. Let $u_1,\ldots,u_a$, and $v_1,\ldots,v_b$, be non-negative weights imposing constraints

$$\sum_{i=1}^{a} u_i\boldsymbol{\alpha}_i = \mathbf{0},\ \ \sum_{j=1}^{b} v_j\boldsymbol{\beta}_j = \mathbf{0},\ \ \sum_{i=1}^{a} u_i\boldsymbol{\gamma}_{ij} = \mathbf{0},\ \ \sum_{j=1}^{b} v_j\boldsymbol{\gamma}_{ij} = \mathbf{0},\ \ i = 1,\ldots,a,\ \ j = 1,\ldots,b. \tag{3.2}$$

3.2.1. Testing interactions. Consider the problem of testing the interaction effects

$$H_{0AB}:\ \boldsymbol{\gamma}_{ij} = \mathbf{0}\ \text{ for } i = 1,\ldots,a,\ \ j = 1,\ldots,b. \tag{3.3}$$

In testing the interactions, the testing procedure does not depend on the chosen weights [21] . Under the null hypothesis H_{0AB}, Xu [21] showed that the minimum of $\sum\limits_{i=1}^{a}\sum\limits_{j=1}^{b}(\overline{\mathbf{Y}}_{ij} - \boldsymbol{\mu}_0 - \boldsymbol{\alpha}_i - \boldsymbol{\beta}_j)'\tilde{\boldsymbol{\Sigma}}_{ij}^{-1}(\overline{\mathbf{Y}}_{ij} - \boldsymbol{\mu}_0 - \boldsymbol{\alpha}_i - \boldsymbol{\beta}_j)$ occures at

$$\widehat{T}(\bar{\mathbf{Y}};\tilde{\mathbf{\Sigma}}) = \sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{\mathbf{Y}}_{ij}-\hat{\boldsymbol{\mu}}_0-\hat{\boldsymbol{\alpha}}_i-\hat{\boldsymbol{\beta}}_j)'\tilde{\mathbf{\Sigma}}_{ij}^{-1}(\overline{\mathbf{Y}}_{ij}-\hat{\boldsymbol{\mu}}_0-\hat{\boldsymbol{\alpha}}_i-\hat{\boldsymbol{\beta}}_j) \quad (3.4)$$
$$= (\bar{\mathbf{Y}}-\boldsymbol{X}\hat{\boldsymbol{\eta}})'\tilde{\mathbf{\Sigma}}^{-1}(\bar{\mathbf{Y}}-\boldsymbol{X}\hat{\boldsymbol{\eta}}),$$

where $n_{ij}\tilde{\mathbf{\Sigma}}_{ij}=\mathbf{\Sigma}_{ij}$, $\tilde{\mathbf{\Sigma}}=\text{diag}\Big(n_{11}^{-1}\mathbf{\Sigma}_{11}, n_{12}^{-1}\mathbf{\Sigma}_{12},...,n_{ab}^{-1}\mathbf{\Sigma}_{ab}\Big)$ with

$$\hat{\boldsymbol{\eta}} = (\hat{\boldsymbol{\mu}}_0', \hat{\boldsymbol{\alpha}}_1',\ldots,\boldsymbol{\alpha}_a', \hat{\boldsymbol{\beta}}_1',\ldots,\hat{\boldsymbol{\beta}}_b')' = \Big(\mathbf{X}'\tilde{\mathbf{\Sigma}}^{-1}\mathbf{X} + \mathbf{L}'\mathbf{L}\Big)^{-1}\mathbf{X}'\tilde{\mathbf{\Sigma}}^{-1}\bar{\mathbf{Y}}.$$

Here $\mathbf{X} = (\mathbf{1}_a\otimes\mathbf{1}_b, \mathbf{I}_a\otimes\mathbf{1}_b, \mathbf{1}_a\otimes\mathbf{I}_b)\otimes\mathbf{I}_p$, $\mathbf{L}=(\mathbf{L}_{*1}', \mathbf{L}_{*2}')'$, $\mathbf{L}_{*1}=(0,u_1,\ldots,u_I,0,...,0)\otimes\mathbf{I}_p$, $\mathbf{L}_{*2}=(0,0,...,0,v_1,\ldots,v_b)\otimes\mathbf{I}_p$, and $\bar{\mathbf{Y}} = \Big(\bar{\mathbf{Y}}_{11}', \bar{\mathbf{Y}}_{12}',...,\bar{\mathbf{Y}}_{1b}',...,\bar{\mathbf{Y}}_{a1}',...,\bar{\mathbf{Y}}_{ab}'\Big)'$.

Furthermore, Xu [21] showed that this interaction sum of squares given in (3.4) can be written as

$$\widehat{T}(\bar{\mathbf{Y}};\tilde{\mathbf{\Sigma}}) = \bar{\mathbf{Y}}'\tilde{\mathbf{\Sigma}}^{-1/2}\Big[\mathbf{I}_{abp}-\tilde{\mathbf{\Sigma}}^{-1/2}\mathbf{X}(\mathbf{X}'\tilde{\mathbf{\Sigma}}^{-1}\mathbf{X} + \mathbf{L}'\mathbf{L})^{-1}\mathbf{X}'\tilde{\mathbf{\Sigma}}^{-1/2}\Big]\tilde{\mathbf{\Sigma}}^{-1/2}\bar{\mathbf{Y}}. \quad (3.5)$$

In order to obtain a test taking Approach 2, here we first present a lemma similar to the one in the previous section, but in terms of the sufficient statistics in two-way MANOVA.

Lemma 3.1. *Let $\mathbf{t}_{ij}$ be the square root of the matrix $\tilde{\mathbf{s}}_{ij}$ (more generally $\mathbf{t}_{ij}$ is obtained from the Cholesky decomposition $\tilde{\mathbf{s}}_{ij}=\mathbf{t}_{ij}\mathbf{t}_{ij}'$. Then, application of the scale transformation $\mathbf{t}_{ij}\tilde{\mathbf{\Sigma}}_{ij}^{-1/2}$ to summary statistics (or raw data) will transform:*

1. *$\overline{\mathbf{Y}}_{ij}-\boldsymbol{\theta}_{ij}$ to $\mathbf{t}_{ij}\tilde{\mathbf{\Sigma}}_{ij}^{-1/2}(\overline{\mathbf{Y}}_{ij}-\boldsymbol{\theta}_{ij})=\mathbf{t}_{ij}\mathbf{Z}_{ij}$,*
2. *$\tilde{\mathbf{S}}_{ij}$ to $\mathbf{S}_{Uij}=\mathbf{t}_{ij}\tilde{\mathbf{\Sigma}}_{ij}^{-1/2}\tilde{\mathbf{S}}_{ij}\tilde{\mathbf{\Sigma}}_{ij}^{-1/2}\mathbf{t}_{ij}'=\mathbf{t}_{ij}\mathbf{U}_{ij}\mathbf{t}_{ij}'/(n_{ij}-1)$,*
3. *the observed value $(\bar{\mathbf{y}}_{ij}-\boldsymbol{\theta}_{ij},\tilde{\mathbf{s}}_{ij})$ of $\Big(\overline{\mathbf{Y}}_{ij}-\boldsymbol{\theta}_{ij},\tilde{\mathbf{S}}_{ij}\Big)$ to $(\bar{\mathbf{y}}_{ij}-\boldsymbol{\theta}_{ij},\tilde{\mathbf{s}}_{ij})$ itself, where $\tilde{\mathbf{S}}_{ij}=\mathbf{S}_{ij}/n_{ij}$, $\tilde{\mathbf{\Sigma}}_{ij}=\mathbf{\Sigma}_{ij}/n_{ij}$, $\mathbf{Z}_{ij}\sim N(\mathbf{0},\mathbf{I}_p)$, and $\mathbf{U}_{ij}\sim W_p(n_{ij}-1,\mathbf{I}_p)$.*

Proof. As before, the random variables appearing in the first two assertions of the above Lemma can be standardized as

$$\mathbf{Z}_{ij} = \tilde{\mathbf{\Sigma}}_{ij}^{-1/2}(\overline{\mathbf{Y}}_{ij}-\boldsymbol{\theta}_{ij})\sim N(\mathbf{0},\mathbf{I}_p),\text{ and} \quad (3.6)$$
$$\mathbf{U}_{ij} = (n_{ij}-1)\tilde{\mathbf{\Sigma}}_{ij}^{-1/2}\tilde{\mathbf{S}}_{ij}\tilde{\mathbf{\Sigma}}_{ij}^{-1/2}\sim W_p(n_{ij}-1,\mathbf{I}_p).$$

Hence, the transformation $\mathbf{t}_{ij}\tilde{\mathbf{\Sigma}}_{ij}^{-1/2}$ applied to the above random quantities will transform $\overline{\mathbf{Y}}_{ij}-\boldsymbol{\theta}_{ij}$ to $\mathbf{t}_{ij}\tilde{\mathbf{\Sigma}}_{ij}^{-1/2}(\overline{\mathbf{Y}}_{ij}-\boldsymbol{\theta}_{ij})=\mathbf{t}_{ij}\mathbf{Z}_{ij}$
and $\tilde{\mathbf{S}}_{ij}$ to $\mathbf{S}_{Uij}=\mathbf{t}_i\tilde{\mathbf{\Sigma}}_{ij}^{-1/2}\tilde{\mathbf{S}}_{ij}\tilde{\mathbf{\Sigma}}_{i}^{-1/2}\mathbf{t}_{ij}'=\mathbf{t}_{ij}\mathbf{U}_{ij}\mathbf{t}_{ij}'/(n_{ij}-1)$, thus proving the first two assertions of Lemma 3.1. The proof of assertion 3 of the lemma is similar to that in Lemma 2.3, because the transformation does not involve random variables. □

As in the one-way MANOVA, tackling $\tilde{\mathbf{\Sigma}}_{ij}$ by $\tilde{\mathbf{S}}_{ij}$, consider the potential test variable

$$T_1(\mathbf{Y};\mathbf{S}) = \sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{\mathbf{Y}}_{ij}-\boldsymbol{\mu}_0-\boldsymbol{\alpha}_i-\boldsymbol{\beta}_j-\mathbf{Y}_0(\overline{\mathbf{Y}}))'\tilde{\mathbf{S}}_{ij}^{-1}$$
$$(\overline{\mathbf{Y}}_{ij}-\boldsymbol{\mu}_0-\boldsymbol{\alpha}_i-\boldsymbol{\beta}_j-\mathbf{Y}_0(\overline{\mathbf{Y}}))$$
$$= \sum_{i=1}^{a}\sum_{j=1}^{b}g_{ij}(\mathbf{Y},\tilde{\mathbf{\Sigma}})'\tilde{\mathbf{S}}_{ij}^{-1}g_{ij}(\mathbf{Y},\tilde{\mathbf{\Sigma}}), \quad (3.7)$$

where $\tilde{\mathbf{S}}_{ij} = \boldsymbol{S}_{ij}/n_{ij}$, $g_{ij}(\mathbf{Y}, \tilde{\mathbf{S}}) = (\overline{\mathbf{Y}}_{ij} - \boldsymbol{\mu}_0 - \boldsymbol{\alpha}_i - \boldsymbol{\beta}_j) - \mathbf{Y}_0(\overline{\mathbf{Y}}; \boldsymbol{\eta}, \tilde{\mathbf{S}})$, and

$$\mathbf{Y}_0(\overline{\mathbf{Y}}) = \mathbf{Y}_0(\overline{\mathbf{Y}}; \boldsymbol{\eta}, \tilde{\mathbf{S}}) = \left(\sum_{i=1}^{a}\sum_{j=1}^{b} \tilde{\mathbf{S}}_{ij}^{-1}\right)^{-1} \left(\sum_{i=1}^{a}\sum_{j=1}^{b} \tilde{\mathbf{S}}_{ij}^{-1}(\overline{\mathbf{Y}}_{ij} - \boldsymbol{\mu}_0 - \boldsymbol{\alpha}_i - \boldsymbol{\beta}_j)\right). \quad (3.8)$$

Now we are in a position obtain a GPQ (a GTV for testing the hypothesis of no interactions) for $T_1(\mathbf{Y}; \mathbf{S})$ as follows.

Theorem 3.2. *The scale transformation of data in Lemma 3.1 transforms $T_1(\mathbf{Y}; \mathbf{S})$ to*

$$\tilde{T}(\mathbf{Y}; \mathbf{S}, \mathbf{s}) = (\mathbf{Y}_Z - \mathbf{X}\boldsymbol{\theta}_Z)'\mathbf{S}_U^{-1}(\mathbf{Y}_Z - \mathbf{X}\boldsymbol{\theta}_Z), \quad (3.9)$$

a GTV appropriate for testing H_{0AB}, where

$$\begin{aligned}
\mathbf{Y}_Z =\ & \left((\mathbf{t}_{11}\mathbf{Z}_{11})', (\mathbf{t}_{12}\mathbf{Z}_{12})', ..., (\mathbf{t}_{1b}\mathbf{Z}_{1b})', ..., (\mathbf{t}_{a1}\mathbf{Z}_{a1})', ..., (\mathbf{t}_{ab}\mathbf{Z}_{ab})'\right)' \\
\mathbf{S}_U =\ & diag(\mathbf{t}_{11}\mathbf{U}_{11}\mathbf{t}_{11}'/(n_{11}-1), \mathbf{t}_{12}\mathbf{U}_{12}\mathbf{t}_{12}'/(n_{12}-1), ..., \mathbf{t}_{1b}\mathbf{U}_{1b}\mathbf{t}_{1b}'/(n_{1b}-1), ..., \\
& \mathbf{t}_{a1}\mathbf{U}_{a1}\mathbf{t}_{a1}'/(n_{a1}-1), ..., \mathbf{t}_{ab}\mathbf{U}_{ab}\mathbf{t}_{ab}'/(n_{ab}-1)) \\
\boldsymbol{\theta}_Z =\ & \left(\mathbf{X}'\mathbf{S}_U^{-1}\mathbf{X} \ + \ \mathbf{L}'\mathbf{L}\right)^{-1} \mathbf{X}'\mathbf{S}_U^{-1}\mathbf{Y}_Z.
\end{aligned}$$

Proof. Clearly, the distribution of $\tilde{T}$ is free of unknown parameters. Furthermore, by design, $\tilde{T}$ is location and scale invariant, and under the null hypothesis specified in (3.3), the quantity $\tilde{T}$ will reduce to $\widehat{T}(\bar{\mathbf{Y}}; \tilde{\mathbf{\Sigma}})$ given in (3.5), which is free of unknown parameters.

Therefore, the p-value to test the interaction hypothesis (3.3) is given by

$$p \ = \ \Pr(\tilde{T} \geq T_{\text{obs}}), \quad (3.10)$$

where T_{obs} is the observed value of (3.5) which is given by

$$T_{\text{obs}} = \bar{\mathbf{y}}'\mathbf{s}^{-1/2}\left[\mathbf{I}_{abp} - \mathbf{s}^{-1/2}\mathbf{X}(\mathbf{X}'\mathbf{s}^{-1}\mathbf{X} \ + \ \mathbf{L}'\mathbf{L})^{-1}\mathbf{X}'\mathbf{s}^{-1/2}\right]\mathbf{s}^{-1/2}\bar{\mathbf{y}}. \quad (3.11)$$

Here $\bar{\mathbf{y}} \ = \ \left(\bar{\mathbf{y}}_{11}', \bar{\mathbf{y}}_{12}', ..., \bar{\mathbf{y}}_{1b}', ..., \bar{\mathbf{y}}_{a1}', ..., \bar{\mathbf{y}}_{ab}'\right)'$ and $\mathbf{s} = \text{diag}(\tilde{\mathbf{s}}_{11}, \tilde{\mathbf{s}}_{12}, ..., \tilde{\mathbf{s}}_{1b}, ..., \tilde{\mathbf{s}}_{a1}, ..., \tilde{\mathbf{s}}_{ab})$. □

As in one-way MANOVA, this p-value can be computed by generating large number of random samples from the normal and Wishart distributions given in Lemma 3.1 and using the Monte Carlo method to estimate the p-value.

Remark : Oversight in two-way MANOVA

Since the result is independent of the chosen weights, Xu [21] presented the above results using the generalized inverse notation $(\mathbf{X}'\tilde{\mathbf{\Sigma}}^{-1}\mathbf{X}\ )^{-}$ without the term $\mathbf{L}'\mathbf{L}$. However, in implementing the parametric bootstrap procedure, this is a misleading statement. Since bootstrapping is done numerically, if one ignores the term $\mathbf{L}'\mathbf{L}$ and replace it with zero, the computer will give a p-value which is not the intended p-value. In the case of univariate two-way ANOVA, this has been demonstrated in [1].

Actually dropping of the $\mathbf{L}'\mathbf{L}$ term is an oversight originated from [3]. Therefore, to help practitioners avoid wrong formulas, we have published an R package, named "twowaytests", on CRAN. The plans are underway to include tests for two-way MANOVA to facilitate the computations, while keeping the missing term $\mathbf{L}'\mathbf{L}$.

3.2.2. Testing the main effects. Testing the main effects with the two-way MANOVA can be extended in a similar fashion. Unlike the interaction effects, testing the main effects does depend on the chosen weights. Furthermore, there is no agreement on how to test the main effect in the presence of the interaction effects. Therefore, here we demonstrate the new approach for testing the hypothesis:

$$H_{0A}: \ \boldsymbol{\alpha}_i + \boldsymbol{\gamma}_{ij} \ = \ \mathbf{0} \ \text{ for } i = 1, \ldots, a, \quad j = 1, \ldots, b. \tag{3.12}$$

Xu (2015) showed that under the null hypothesis H_{0A}, the minimum of $\sum_{i=1}^{a} \sum_{j=1}^{b} (\overline{\mathbf{Y}}_{ij} - \boldsymbol{\mu}_0 - \boldsymbol{\beta}_j)'\tilde{\boldsymbol{\Sigma}}_{ij}^{-1}(\overline{\mathbf{Y}}_{ij} - \boldsymbol{\mu}_0 - \boldsymbol{\beta}_j)$ occurs at

$$T_A(\bar{\mathbf{Y}}; \tilde{\boldsymbol{\Sigma}}) = (\bar{\mathbf{Y}} - \boldsymbol{X}_1\hat{\boldsymbol{\theta}}_1)'\tilde{\boldsymbol{\Sigma}}^{-1}(\bar{\mathbf{Y}} - \boldsymbol{X}_1\hat{\boldsymbol{\theta}}_1), \tag{3.13}$$

where $n_{ij}\tilde{\boldsymbol{\Sigma}}_{ij} = \boldsymbol{\Sigma}_{ij}$, $\tilde{\boldsymbol{\Sigma}} = \text{diag}\Big(n_{11}^{-1}\boldsymbol{\Sigma}_{11}, n_{12}^{-1}\boldsymbol{\Sigma}_{12}, ..., n_{ab}^{-1}\boldsymbol{\Sigma}_{ab}\Big)$,

$$\hat{\boldsymbol{\theta}}_1 = (\hat{\boldsymbol{\mu}}_0^{'}, \hat{\boldsymbol{\beta}}_1^{'}, \ldots, \hat{\boldsymbol{\beta}}_b^{'})' = \Big(\mathbf{X}_1'\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_1 \ + \ \mathbf{L}_1'\mathbf{L}_1\Big)^{-1}\mathbf{X}_1'\tilde{\boldsymbol{\Sigma}}^{-1}\bar{\mathbf{Y}}., \tag{3.14}$$

where $\mathbf{X}_1 \ = \ (\mathbf{1}_a \otimes \mathbf{1}_b, \mathbf{1}_a \otimes \mathbf{I}_b) \otimes \mathbf{I}_p$, $\mathbf{L}_1 = (0, 0, ..., 0, v_1, \ldots, v_b) \otimes \mathbf{I}_p$,
and $\bar{\mathbf{Y}} \ = \ \Big(\bar{\mathbf{Y}}_{11}', \bar{\mathbf{Y}}_{12}', ..., \bar{\mathbf{Y}}_{1b}', ..., \bar{\mathbf{Y}}_{a1}', ..., \bar{\mathbf{Y}}_{ab}'\Big)'$.

As the case with two-way interaction, the between group sum of squares of cross product given by (3.13) is equal to

$$T_A(\bar{\mathbf{Y}}; \tilde{\boldsymbol{\Sigma}}) = \bar{\mathbf{Y}}'\tilde{\boldsymbol{\Sigma}}^{-1/2}\Big[\mathbf{I}_{abp} - \tilde{\boldsymbol{\Sigma}}^{-1/2}\mathbf{X}_1(\mathbf{X}_1'\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_1 \ + \ \mathbf{L}_1'\mathbf{L}_1)^{-1}\mathbf{X}_1'\tilde{\boldsymbol{\Sigma}}^{-1/2}\Big]\tilde{\boldsymbol{\Sigma}}^{-1/2}\bar{\mathbf{Y}}, \tag{3.15}$$

and under the null hypothesis H_{0A}, the observed value of the above test variable $T_A(\bar{\mathbf{Y}}; \tilde{\boldsymbol{\Sigma}})$ will reduce to

$$T_{\text{obs}}^{*} = \bar{\mathbf{y}}'\mathbf{s}^{-1/2}\Big[\mathbf{I}_{abp} - \mathbf{s}^{-1/2}\mathbf{X}_1(\mathbf{X}_1'\mathbf{s}^{-1}\mathbf{X}_1 \ + \ \mathbf{L}_1'\mathbf{L}_1)^{-1}\mathbf{X}_1'\mathbf{s}^{-1/2}\Big]\mathbf{s}^{-1/2}\bar{\mathbf{y}}, \tag{3.16}$$

where $\bar{\mathbf{y}} \ = \ \Big(\bar{\mathbf{y}}_{11}', \bar{\mathbf{y}}_{12}', ..., \bar{\mathbf{y}}_{1b}', ..., \bar{\mathbf{y}}_{a1}', ..., \bar{\mathbf{y}}_{ab}'\Big)'$ is the observed value of $\bar{\mathbf{Y}}$, and
$\mathbf{s} = \text{diag}(\tilde{\mathbf{s}}_{11}, \tilde{\mathbf{s}}_{12}, ..., \tilde{\mathbf{s}}_{1b}, ..., \tilde{\mathbf{s}}_{a1}, ..., \tilde{\mathbf{s}}_{ab})$.

In view of the above results, tackling $\tilde{\boldsymbol{\Sigma}}_{ij}$ by $\tilde{\mathbf{S}}_{ij}$,, as in previous sections consider the potential GTV

$$\begin{aligned} T_A(\bar{\mathbf{Y}}; \mathbf{S}) \ &= \ \sum_{i=1}^{a} \sum_{j=1}^{b} (\overline{\mathbf{Y}}_{ij} - \boldsymbol{\mu}_0 - \boldsymbol{\beta}_j - \mathbf{Y}_0(\overline{\mathbf{Y}}))'\tilde{\mathbf{S}}_{ij}^{-1} \\ &\quad (\overline{\mathbf{Y}}_{ij} - \boldsymbol{\mu}_0 - \boldsymbol{\beta}_j - \mathbf{Y}_0(\overline{\mathbf{Y}})) \\ &= \ \sum_{i=1}^{a} \sum_{j=1}^{b} g_{ij}(\mathbf{Y}, \tilde{\boldsymbol{\Sigma}})'\tilde{\mathbf{S}}_{ij}^{-1} g_{ij}(\mathbf{Y}, \tilde{\boldsymbol{\Sigma}}), \end{aligned} \tag{3.17}$$

where $\tilde{\mathbf{S}}_{ij} = \boldsymbol{S}_{ij}/n_{ij}$, $g_{ij}(\mathbf{Y}, \tilde{\mathbf{S}}) = (\overline{\mathbf{Y}}_{ij} - \boldsymbol{\mu}_0 - \boldsymbol{\beta}_j) - \mathbf{Y}_0(\overline{\mathbf{Y}}; \boldsymbol{\eta}, \tilde{\mathbf{S}})$, and

$$\mathbf{Y}_0(\overline{\mathbf{Y}}) = \mathbf{Y}_0(\overline{\mathbf{Y}}; \boldsymbol{\eta}, \tilde{\mathbf{S}}) = \left(\sum_{i=1}^{a} \sum_{j=1}^{b} \tilde{\mathbf{S}}_{ij}^{-1}\right)^{-1} \left(\sum_{i=1}^{a} \sum_{j=1}^{b} \tilde{\mathbf{S}}_{ij}^{-1}(\overline{\mathbf{Y}}_{ij} - \boldsymbol{\mu}_0 - \boldsymbol{\beta}_j)\right). \tag{3.18}$$

Then, the desired GPQ for $T_A(\mathbf{Y}; \mathbf{S})$ for testing the null hypothesis of equal main effects can be obtained with the aid of the following Proposition.

Theorem 3.3. *The scale transformation of data as in Lemma 3.1 transforms $T_A(\mathbf{Y};\mathbf{S})$ to*

$$\tilde{T}_A(\mathbf{Y};\mathbf{S},\mathbf{s}) = (\mathbf{Y}_Z - \mathbf{X}_1\boldsymbol{\theta}_{Z1})'\mathbf{S}_U^{-1}(\mathbf{Y}_Z - \mathbf{X}_1\boldsymbol{\theta}_{Z1}), \tag{3.19}$$

a GTV appropriate for testing H_{0A}. Here

$$\begin{aligned}
\mathbf{Y}_Z =& \left((\mathbf{t}_{11}\mathbf{Z}_{11})', (\mathbf{t}_{12}\mathbf{Z}_{12})', ..., (\mathbf{t}_{1b}\mathbf{Z}_{1b})', ..., (\mathbf{t}_{a1}\mathbf{Z}_{a1})', ..., (\mathbf{t}_{ab}\mathbf{Z}_{ab})'\right)' \\
\mathbf{S}_U =& \ diag(\mathbf{t}_{11}\mathbf{U}_{11}\mathbf{t}_{11}'/(n_{11}-1), \mathbf{t}_{12}\mathbf{U}_{12}\mathbf{t}_{12}'/(n_{12}-1), ..., \mathbf{t}_{1b}\mathbf{U}_{1b}\mathbf{t}_{1b}'/(n_{1b}-1), ..., \\
& \mathbf{t}_{a1}\mathbf{U}_{a1}\mathbf{t}_{a1}'/(n_{a1}-1), ..., \mathbf{t}_{ab}\mathbf{U}_{ab}\mathbf{t}_{ab}'/(n_{ab}-1)) \\
\boldsymbol{\theta}_{Z1} =& \left(\mathbf{X}_1'\mathbf{S}_U^{-1}\mathbf{X}_1 \ + \ \mathbf{L}_1'\mathbf{L}_1\right)^{-1}\mathbf{X}_1'\mathbf{S}_U^{-1}\mathbf{Y}_Z.
\end{aligned}$$

Proof. Clearly, the distribution of $\tilde{T}_A$ is free of unknown parameters, and under the null hypothesis specified in (3.12), the quantity $\tilde{T}_A$ will reduce to $T_A(\bar{\mathbf{Y}};\tilde{\boldsymbol{\Sigma}})$ given in (3.13), which is free of unknown parameters. □

Therefore, the p-value to test the hypothesis (3.12) is given by

$$p \ = \ \Pr(\tilde{T}_A \geq T^*_{\text{obs}}), \tag{3.20}$$

where T^*_{obs} is the observed value of $\tilde{T}_A$ which is given in (3.16). As with the case of interaction effects, this p-value can also be computed by generating large number of random samples from the normal and Wishart distributions given in Lemma 3.1 and employing the Monte Carlo method to estimate the p-value.

4. Numerical results

In this section, we use the same example that Krishnamoorthy and Lu [8] used to illustrate the parametric bootstrap procedure. The data set originally discussed in [13] contains five skull samples from the early pre-dynastic period, the late pre-dynastic period, the 12th and 13th dynasties period, the Ptolemaic period, and the Roman period; are abbreviated as C. 4000 BC, C. 3300 BC, C. 1850 BC, C. 200 BC, and C. AD 150 respectively. Each of the above samples contain measurements from thirty skulls. Four measurements are available for each skull.

Krishnamoorthy and Lu [8] used a subset of this data set, by picking only the first four groups, namely the periods C. 4000 BC, C. 3300 BC, C. 1850 BC, and C. 200 BC, and used only the first fifteen observations from each group. In our illustration we use exactly the same subset used by [8], which contains fifteen observations from each population, and each observation involves 4 measurements. For testing the equality of the mean vector of these 4 populations, Krishnamoorthy and Lu [8] showed that the parametric bootstrap approach yielded a p-value of 0.041. This method is analytically equivalent to the method we described in section 2.2.

We analyzed data by the classical Wilk's test, the two generalized Wilk's tests that were introduced in section 2.1 (GW1 and GW2 tests), and the PillaiBartlett test available from the CRAN package "manova" (abbreviated as MAN). The resulting p-values are given in Table 1.

Table 1. The p-values for testing the mean vector of skull measurements

Method	WL	GW1	GW2	GPB	MAN
p-values	0.028	0.016	0.030	0.041	0.033

It is evident from Table 1 that the p-values given by each of the five methods lead to the same conclusion that mean vector of measurements is not the same for the four populations. The p-value of the GW2 test is almost the same as that of the MAN test. Of course this is not the same in other scenarios, as we now study via simulated data sets.

4.1. Simulation study

Since the introduction of two approaches to handle heteroscedastc MANOVA models has been the primary purpose of this article, providing an in-depth simulation study is beyond the scope of this paper. Readers interested in such a study is referred to [8]. While undertaking a subset of parameter scenarios considered by [8], in our comparison of Type I Error rates of competing tests, in addition to the two generalized Wilks tests, abbreviated as GW1 and GW2, we also included the classical Wilks Lambda test (abbreviated as WL), since many practitioners simply apply the classical tests for convenience, despite that they are valid only when the covariances are equal. Also included in the comparison is the default PillaiBartlett test available from the CRAN package "manova" (abbreviated as MAN).

The scenarios on the covariance matrices and sample sizes, are as specified in Table 2 below. A preliminary study suggested that the computation of Type-I error requires a greater number of replications compared to the computation of p-values by the GW methods and the GPB method. Therefore, to evaluate the Type I error of competing tests, under each scenario we generated 10,000 samples of simulated data. For each simulated set of data, the p-values of the three generalized tests, GW1, GW2, and GPB were computed using 1,000 Monte Carlo samples. When estimated Type-I error we got was substantially different from what is reported by [8], Monte Carlo sample size is increased to 5,000. The classical tests, the Wilks test and the Pillai-Bartlett test do not require such Monte Carlo samples. At the intended size of 0.05, the estimated Type-I error rates are reported in Table 2 below.

Table 2. Monte Carlo estimates of Type-I error rates

$k=3,\ p=2,\ \sum_1=\mathbf{I}_2,\quad \sum_2=\mathrm{diag}(\lambda_1,\lambda_2)\,,\quad \sum_3=\begin{pmatrix}1 & \rho_3\\ \rho_3 & 1\end{pmatrix}$

$(n_1,\ n_2,\ n_3)$	$(\lambda_1,\ \lambda_2,\ \rho_3)$	WL	$GW1$	$GW2$	GPB	MAN
$(7,7,7)$	$(0.2,0.6,0.5)$	0.062	0.052	0.042	0.053	0.052
$(7,10,20)$	$(1,0.1,0.3)$	0.071	0.056	0.045	0.058	0.043
$(10,10,10)$	$(1,0.5,0.2)$	0.053	0.048	0.042	0.051	0.049
$(10,10,40)$	$(0.9,0.9,0.6)$	0.142	0.054	0.047	0.067	0.076
$(10,10,40)$	$(0.7,0.8,-0.2)$	0.088	0.051	0.041	0.059	0.036
$(25,20,20)$	$(1,0.5,0.2)$	0.049	0.051	0.049	0.048	0.045

$k=5,\ p=2,\ \sum_1=\mathbf{I}_2,\quad \sum_2=\mathrm{diag}(\lambda_1,\lambda_2)\,,\quad \sum_i=\begin{pmatrix}1 & \rho_i\\ \rho_i & 1\end{pmatrix},\ i=3,4,5$

$(n_1,\ n_2,\ n_3)$	$(\lambda_1,\ \lambda_2,\ \rho_1,\rho_2,\rho_3)$	WL	$GW1$	$GW2$	GPB	MAN
$(7,7,7,7,7)$	$(0.1,0.9,0.1,0.4,0.9)$	0.063	0.102	0.133	0.051	0.053
$(7,7,7,7,7)$	$(0.1,0.3,-0.1,0.1,0.9)$	0.067	0.104	0.128	0.051	0.059
$(12,12,12,12,12)$	$(0.1,0.7,0,0,0)$	0.060	0.079	0.095	0.055	0.056
$(15,20,10,32,7)$	$(0.1,0.9,0.1,0.4,0.9)$	0.091	0.078	0.088	0.057	0.066
$(15,20,10,32,7)$	$((0.9,0.9,-0.4,0.6,0.9)$	0.091	0.076	0.085	0.049	0.068
$(15,20,10,32,7)$	$(0.1,0.1,0.3,0.3,0.3)$	0.107	0.090	0.104	0.079	0.076

From the results of our simulation study, we can guide the practitioners as follows.

(1) When there are 3 populations of interest, GW1 is the preferred choice when one wishes to be close to the intended Type I error,

(2) When there are 3 populations of interest, GW2 is the preferred choice when one wishes to stay under the the intended Type I error,

(3) When there are a large number of populations, GPB is the preferred choice when one wishes to be close to the intended Type I error,

(4) The PillaiBartlett test, which is the default test available from the manova (MAN) available from CRAN, is better than the classical Wilks Lambda (WL) test preferred by many practitioners,

(5) When the models are balanced in terms of sample sizes, the MAN test is fairly good and easier to compute, though not as good as the GPB test.

In view of the last two findings, we encourage researchers to extend the PillaiBartlett test to the unequal covariances case by taking Approach 1 proposed in this article.

5. Discussion

In this article we proposed two approaches to extending classical MANOVA tests to avoid the classical assumption of equal error covariances and the assumption of equal cell frequencies. The approaches are illustrated by deriving two types of generalized Wilks tests and PB like tests without any bootstrap argument. The p-values of all four tests are exact probabilities of well defined extreme regions of the sample space, a property PB type tests lack. Weerahandi [17] argued that practitioners need to to the best with the sample at hand as Bayesians do, and that arguments of repeated sampling is not something practically useful. Nevertheless, until researchers develop alternative methods to compare competing tests, one needs to rely on repeated sampling based methods when there are alternative tests. Therefore, we compared each of the generalized test for MANOVA against each other and against widely used approximate tests.

The advantage of the two approaches proposed in this article is that one can easily apply them in extending results to higher-way MANOVA, MANCOVA, and RM MANOVA. Moreover, the approach we took in extending the Wilks type tests (GW tests) to the case of unequal covariances, researchers can take in extending their favorite tests such as Roy's largest root test, Lawley-Hotelling's test, and Bartlett-Nanda-Pillai test.

The GPB approach also has a number of advantages over the PB approach. For example, the former does not require good point estimators such as MLEs of parameters to start with, thus making inferences such as RM MANOVA in mixed effects models possible by employing distribution theory developed by [20].

Moreover, the approach we took in developing GPB should also suggest how researchers could extend the GPV method introduced by [12]. In the ANOVA case, while the GPV test is practically as good as the PB test, it has the added advantage that its size does not exceed the intended Type-I error.

References

[1] M.M.A. Ananda, O. Dag and S. Weerahandi, *Heteroscedastic two-way ANOVA under constraints*, Commun. Stat. Theory Methods, Doi:10.1080/03610926.2022.2059682, 2022.

[2] M.M.A. Ananda and S. Weerahandi, *Two-Way ANOVA with unequal cell frequencies and unequal variances*, Stat. Sin.**7** (3), 631-646, 1997.

[3] S.C. Chow and S.G. Wang, *Advanced Linear Models*, Marcel Dekker, New York, 1994.

[4] Y. Fujikoshi, *Two-way ANOVA models with unbalanced data*, Discrete Math. **116** (1-3), 315-334, 1993.

[5] J. Gamage, T. Mathew and S. Weerahandi, *Generalized p-values and generalized confidence regions for the multivariate BehrensFisher problem and MANOVA*, J. Multivar. Anal. **88** (1), 177189, 2004.

[6] R.A. Johnson and S. Weerahandi, *A Bayesian solution to the multivariate Behrens-Fisher problem*, J Am Stat Assoc **83** (401), 145-149, 1988.

[7] R.A. Johnson and D.W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice Hall Upper Saddle River, NJ, 2002.

[8] K. Krishnamoorthy and F. Lu, *A parametric bootstrap solution to the MANOVA under heteroscedasticity*, J Stat Comput Simul **80** (8), 873887, 2010.

[9] K. Krishnamoorthy, F. Lu and T. Mathew, *A parametric bootstrap approach for ANOVA with unequal variances: fixed and random models*, Comput Stat Data Anal **51** (12), 5731-5742, 2007.

[10] X. Li, J. Wang and H. Liang, *comparison of several means: a fiducial based approach*, Comput Stat Data Anal **55** (5), 1993-2002, 2011.

[11] A. Roy and A. Bose, *coverage of generalized confidence intervals*, J. Multivar. Anal. **100** (7), 1384-1397, 2009.

[12] S.M. Sadooghi-Alvandi, A.A. Jafari and H.A. Mardarni-Fard, *One-way ANOVA with unequal variances*, Commun. Stat. Theory Methods **41** (22), 42004221, 2012.

[13] A. Thomson and D. Randall-MacIver, *The Ancient Races of the Thebaid: Being an Anthropometrical Study of the Inhabitants of Upper Egypt from the Earliest Prehistoric Times to the Mohammedan Conquest, Based Upon the Examination of Over 1,500 Crania*, Clarendon Press, 1905.

[14] K. Tsui and S. Weerahandi, *Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters*, J Am Stat Assoc **84** (406), 602-607, 1989.

[15] S. Weerahandi, *Generalized Inference in Repeated Measures: Exact Methods in MANOVA and Mixed Models*, John Wiley & Sons, 2004.

[16] S. Weerahandi, *ANOVA under unequal error variances*, Biometrics **51** (2), 589-599, 1995.

[17] S. Weerahandi, *Generalized confidence intervals*, J Am Stat Assoc **88** (423), 899-905, 1993.

[18] S. Weerahandi, *Testing regression equality with unequal variances*, Econometrica **55** (5), 1211-1215, 1987.

[19] S. Weerahandi and K. Krishnamoorthy, *A note reconciling ANOVA tests under unequal error variances*, Commun. Stat. Theory Methods **48** (3), 689-693, 2019.

[20] S. Weerahandi and C.R. Yu, *Exact distributions of statistics for making inferences on mixed models under the default covariance structure*, J. Stat. Distrib. Appl. **7** (4), 1-14, 2020.

[21] L. Xu, *Parametric bootstrap approaches for two-way MANOVA with unequal cell sizes and unequal cell covariance matrices*, J. Multivar. Anal. **133**, 291-303, 2015.

[22] L. Xu, F. Yang, A. Abula and S. Qin, *A parametric bootstrap approach for two-way ANOVA in presence of possible interactions with unequal variances*, J. Multivar. Anal. **115**, 172-180, 2013.

[23] J. Zhang, *Two-way MANOVA with unequal cell sizes and unequal cell covariance matrices*, Technometrics **53** (4), 426-439, 2011.