How to Compare Small Multivariate Samples Using Nonparametric Tests

Arne C. Bathke[1], Solomon W. Harrar[2], Laurence V. Madden[3]

Department of Statistics, University of Kentucky[1]

Department of Mathematical Sciences, University of Montana[2]

Department of Plant Pathology, Ohio State University[3]

June 8, 2007

Abstract

In plant pathology, in particular, and plant science, in general, experiments are often conducted to determine disease and related responses of plants to various treatments. Typically, such data are multivariate, where different variables may be measured on different scales that can be quantitative, ordinal, or mixed. To analyze these data, we propose different nonparametric (rank-based) tests for multivariate observations in balanced and unbalanced one-way layouts. Previous work has led to the development of tests based on asymptotic theory, either for large numbers of samples or groups; however, most experiments comprise only small or moderate numbers of groups and samples. Here, we investigate several tests based on small-sample approximations, and compare their performance in terms of α levels and power for different simulated situations, with continuous and discrete observations. For positively correlated responses, an approximation based on Brunner et al. (1997) ANOVA-Type statistic performed best; for responses with negative correlations, in general, an approximation based on the Lawley-Hotelling type test performed best. We demonstrate the use of the tests based on the approximations for a plant pathology experiment.

Keywords: Rank Test, Small Samples, ANOVA-Type Test, Lawley-Hotelling Test, Bartlett-Nanda-Pillai Test.

1 Introduction

In this manuscript, we describe nonparametric methods to test hypotheses regarding multivariate data from several samples. *Nonparametric* methods are in general more versatile than parametric procedures in the sense that they are not restricted to specific population distributions. In particular, the nonparametric methods proposed here can be used for continuous responses as well as for discrete ordinal data.

Since these procedures are based on ranks, they are robust to outliers. *Multivariate* data are data in which more than one response is of interest. In clinical trials, this situation is referred to as *multiple endpoints*. The approach of this manuscript allows for a mix of response variables where some variables are quantitative, and others are ordinal, or even binary.

In the field of plant pathology, multivariate data consisting of continuous responses, binary observations or percentages, and ordinal measurements are common. For instance, Rowe et al. (1985) assessed the effects of inoculum density (dose) of two pathogens that collectively cause early dying disease of potatoes on the plant response to infection. In particular, in a randomized field study, there were 12 treatments and four response variables, an ordinal measurement of the severity of the disease, plant yield, and biomass of foliage and of the roots. Chatfield et al. (2000) assessed the visible symptoms of a disease of ornamental crapapples at four different months using an ordinal measurement scale. Sixty-three varieties that varied in resistance were compared for their response over the four months. As a third example, in a greenhouse study, the effects of pesticide (fungicide), application of a biocontrol agent, and use of a suppressive soil mixture on a foliar and flower disease of ornamental begonia were assessed (Horst et al., 2005). The percentage of the total plant surface with disease symptoms, the shoot dry weight (grams), and an ordinal measurement of the overall quality of plant appearance (salability) were determined for each of six treatments. Although the authors of these studies originally used univariate methods for analyzing each of the response variables, it would be more effective to test for treatment effects for variables simultaneously.

Factorial designs involve more than one factor to make inference about. In this manuscript, however, we will only consider the situation involving one factor with a different levels. The a levels may be thought of as a populations from which we sample, or as applying a different treatments to subjects. Typical statistical hypotheses to be tested are the following. "Are the a samples from the same population (multivariate distribution)?" or "Do the a treatments have the same effect?" A schematic layout for the type of data considered in this manuscript is provided in Table 1.

Table 1: General schematic layout for multivariate observations from several samples.

Sample 1				*Sample 2*				$\dots$	*Sample a*			
$X_{11}^{(1)}$	$X_{12}^{(1)}$	$\dots$	$X_{1n_1}^{(1)}$	$X_{21}^{(1)}$	$X_{22}^{(1)}$	$\dots$	$X_{2n_2}^{(1)}$	$\dots$	$X_{a1}^{(1)}$	$X_{a2}^{(1)}$	$\dots$	$X_{a,n_a}^{(1)}$
$X_{11}^{(2)}$	$X_{12}^{(2)}$	$\dots$	$X_{1n_1}^{(2)}$	$X_{21}^{(2)}$	$X_{22}^{(2)}$	$\dots$	$X_{2n_2}^{(2)}$	$\dots$	$X_{a1}^{(2)}$	$X_{a2}^{(2)}$	$\dots$	$X_{a,n_a}^{(2)}$
	$\dots$				$\dots$			$\dots$		$\dots$		
$X_{11}^{(p)}$	$X_{12}^{(p)}$	$\dots$	$X_{1n_1}^{(p)}$	$X_{21}^{(p)}$	$X_{22}^{(p)}$	$\dots$	$X_{2n_2}^{(p)}$	$\dots$	$X_{a1}^{(p)}$	$X_{a2}^{(p)}$	$\dots$	$X_{a,n_a}^{(p)}$

Early work on nonparametric, multivariate statistics has been done by Puri and Sen (1966; Sen 1969, 1970), assuming continuous population distributions and semiparametric location models. Thompson (1990, 1991) suggested a nonparametric approach involving rankings over all coordinates. However, such an overall ranking has the effect that the test statistics are not invariant under separate monotone transformations (*e.g.*, scale changes) of the coordinate variables. Munzel and Brunner (2002a,b) used separate rankings for the different variables, thus allowing for possible separate transformations of the individual response variables. Munzel and Brunner (2002a,b), following Munzel (1999) for univariate data, developed their methodology based on the normalized distribution (Lévy 1925; Kruskal 1952; Ruymgaart 1980), rather than assuming a continuous distribution, to provide a theory that directly allows for ties. These authors, as well as Bathke and Harrar (2007) and Harrar and Bathke (2006, 2007), formulated the hypotheses in terms of the normalized distributions instead of in terms of a location shift. This approach has become quite common in nonparametric univariate analyses (see Brunner and Puri, 2001).

Munzel and Brunner (2002a,b) and Harrar and Bathke (2007) derived asymptotic theory for large sample sizes n_i per treatment level. More recently, Bathke and Harrar (2007) and Harrar and Bathke (2006) have derived asymptotic theory mainly for the situation in which a is large and the sample sizes n_i per treatment level are small. However, often in plant pathology and related plant sciences, neither a nor n_i is large. Here, we focus on finite approximations for the tests based on asymptotic theory for the situations in which both, a and n_i are moderate to small, and on demonstrating the applicability of the newly derived procedures for data from experimental designs similar to those described above. We focus on nonparametric tests that are based on three parametric multivariate test statistics known to perform well under normality. These are the ANOVA-type, Lawley-Hotelling, and Bartlett-Nanda-Pillai tests (Munzel and Brunner 2002a,b; Bathke and Harrar 2007; Harrar and Bathke 2006, 2007). The approximation procedures described in this manuscript are mainly designed to improve performance of the asymptotic methods proposed in Harrar and Bathke (2006).

The procedures introduced or examined in this manuscript are based on ranks, but they could in principle also be employed using the original observations in the calculations instead of the ranks. In that case, the correct null hypothesis would be equality of mean vectors instead of equality of normalized distributions (see below). However, unless the data justifies the normal distribution assumption, we recommend using the nonparametric procedures, that is, the tests based on ranks. Main advantages of the nonparametric techniques are robustness, invariance under monotone transformations of the scales on which the individual variables are measured, and, as confirmed in several simulations, a much higher power of the nonparametric tests when the data contains outliers (Munzel and Brunner 2002a,b; Bathke and Harrar 2007; Harrar and Bathke 2006, 2007).

In the following sections, we first review asymptotic nonparametric tests for multivariate balanced single-factor designs when the number of treatments is large and replication sizes are small ($a \to \infty,\ n_i$ bounded), and also when the number of treatments is small and replication sizes are large. The nonparametric tests for unbalanced designs are then reviewed. We do *not* assume normality. In particular for the nonparametric tests, the population distributions can be almost arbitrary, it is only required that the distributions are not degenerate. We present several finite approximations for the ANOVA-type, Lawley-Hotelling, and Bartlett-Nanda-Pillai test statistics for balanced or unbalanced designs. A simulation study is presented to determine type I error rate and power for all the finite approximations for a range of design situations. Then, an example analysis is presented for a small data set consisting of ordinal and continuous measurements.

Notation: We use the letter p for the number of variables that are measured, a for the number of factor levels, treatments, or populations from which subjects are sampled, and n or n_i denotes the sample size per treatment. Vectors or matrices of random variables are indicated by bold letters.

2 Theory and Asymptotic Results

2.1 Model and Statistical Hypotheses

Let $\mathbf{X}_{ij} = (X_{ij}^{(1)}, \ldots, X_{ij}^{(p)})' \sim F_i,\ i = 1, \ldots, a,\ j = 1, \ldots, n_i$ be independent random vectors with multivariate distributions F_i. The vectors $\mathbf{X}_{ij}$ have possibly dependent components $X_{ij}^{(k)}$ with marginal distributions

$$F_i^{(k)}(x) = \frac{1}{2}\left(P(X_{ij}^{(k)} \le x) + P(X_{ij}^{(k)} < x)\right), \qquad k = 1, \ldots, p,$$

and the marginal distributions $F_i^{(k)}$ are assumed non-degenerate. We call this version of the distribution function the *normalized version*. It is simply the average of the left-continuous and the right-continuous versions. Even though most introductory statistics textbooks prefer to employ the right-continuous version, using the normalized version of the distribution function has advantages in the context of non-parametric tests because is allows for discrete ordinal data and accommodates ties in a natural way. The normalized version of the distribution function dates back to Lévy (1925) and has been used in the context of rank statistics by Kruskal (1952), Ruymgaart (1980), and Munzel (1999), among others.

The nonparametric test statistics considered in this paper use separate rankings for the p different variables. Let $R_{ij}^{(k)}$ denote the (mid-)rank of $X_{ij}^{(k)}$ among all $N = \sum_{i=1}^{a} n_i$ random variables $X_{11}^{(k)}, \ldots, X_{an_a}^{(k)}$. Then, the column vector $\mathbf{R}_{ij} = (R_{ij}^{(1)}, \ldots, R_{ij}^{(p)})'$ contains all the ranks of one multivariate observation, and the $p \times N$ matrix $\mathbf{R} = (\mathbf{R}_{11}, \ldots, \mathbf{R}_{1n_1}, \mathbf{R}_{21}, \ldots, \mathbf{R}_{an_a})$ contains the ranks for all variables and all observations. A variable corresponds to a row, and a multivariate observation corresponds to a column of the matrix $\mathbf{R}$.

Hypotheses are stated in terms of distribution functions. We focus on testing the *multivariate hypothesis*

$$H_0 : F_1 = \cdots = F_a\ ,$$

whose name relates to the fact that it is formulated using the joint multivariate distributions F_i. This hypothesis is stronger than the so-called *marginal hypothesis*

$$H_0^* :\ F_1^{(k)} = \cdots = F_a^{(k)},\ k = 1, \ldots, p,$$

which is formulated in terms of the marginal distributions of the individual response variables. When H_0 is true, H_0^* must also be true.

In the following sections, we summarize the asymptotic results for nonparametric multivariate tests that have been obtained recently. For exact conditions and detailed proofs, refer to the corresponding original manuscripts whose citations are provided below.

2.2 Asymptotic Results for the Balanced Design

In a balanced design, $n_1 = \ldots = n_a = n$, and the total number of multivariate observations is therefore $N = n \cdot a$. Define the matrix-valued quadratic forms

$$\mathbf{H} = \frac{1}{a-1} \sum_{i=1}^{a} n(\bar{\mathbf{R}}_{i.} - \bar{\mathbf{R}}_{..})(\bar{\mathbf{R}}_{i.} - \bar{\mathbf{R}}_{..})' \quad \text{and} \quad \mathbf{G} = \frac{1}{N-a} \sum_{i=1}^{a} \sum_{j=1}^{n} (\mathbf{R}_{ij} - \bar{\mathbf{R}}_{i.})(\mathbf{R}_{ij} - \bar{\mathbf{R}}_{i.})',$$

where $\bar{\mathbf{R}}_{i.}$ and $\bar{\mathbf{R}}_{..}$ are $p \times 1$ vectors defined similar to $\mathbf{R}_{ij}$. The dot-overbar notation indicates averaging over the index corresponding to the dot. In the univariate case ($p = 1$), $\mathbf{H}$ and $\mathbf{G}$ would be the mean squares due to treatment (hypothesis mean sum of squares) and the mean squares due to error, respectively. In the multivariate case, $\mathbf{H}$ and $\mathbf{G}$ are both $p \times p$ matrices, but because of the similarity to the univariate case, we will still refer to them as hypothesis and error sum of squares matrices. The kth elements on the diagonal of both matrices are equivalent to the mean squares that would be obtained in a univariate analysis of the kth variable, and the off-diagonal entries estimate covariances among the p different variables.

One of the main issues in multivariate statistical inference is how to obtain a one-dimensional test statistic based upon the hypothesis and error sum of squares matrices. In univariate inference, the F test statistic is simply the ratio of $\mathbf{H}$ and $\mathbf{G}$, but in multivariate inference, this ratio, if it exists, is a $p \times p$ matrix itself. Based on the assumption of multivariate normality, several approaches to obtain a univariate test statistic have been proposed in the literature, and none of these is uniformly better than the others (*cf.* Rencher 2002, pp.176–178). We consider here three nonparametric test statistics that are all directly or indirectly motivated by some of the most popular parametric tests for multivariate data: ANOVA-type (Brunner, Dette, and Munk 1997; Brunner, Munzel, and Puri 1999; following ideas by Satterthwaite 1946, Welch 1951, Box 1954, and Dempster 1958, 1960), Lawley-Hotelling (Lawley 1938; Bartlett 1939; Hotelling 1947, 1951; see also Anderson 1984, pp.323–326), and Bartlett-Nanda-Pillai (Bartlett 1939; Nanda 1950; Pillai 1955; see also Anderson 1984, pp.326–328).

2.2.1 Results for a Large Number of Samples a

The following standardized multivariate nonparametric test statistics have under $H_0 : F_1 = \cdots = F_a$, as $a \to \infty$, n, p fixed, asymptotically a standard normal distribution (derived in Bathke and Harrar, 2007).

ANOVA-Type Statistic.

$$\sqrt{\frac{a(n-1)}{2n}} \cdot \frac{\text{tr}(\mathbf{H}) - \text{tr}(\mathbf{G})}{\sqrt{\text{tr}(\mathbf{G}^2)}} \tag{1}$$

Lawley-Hotelling-Type Statistic.

$$\sqrt{\frac{a}{\tau}} \cdot \left[\text{tr}\left(\mathbf{H}\mathbf{G}^-\right) - r_1\right] \ , \ \text{ where } \tau = \frac{2n\rho}{n-1}, \ r_1 = \text{rank}(\mathbf{G}), \text{ and } \rho = \text{rank}(\Sigma_Y) \leq p. \tag{2}$$

Bartlett-Nanda-Pillai-Type Statistic.

$$\sqrt{\frac{a}{\tau}} \cdot \left(\frac{N-1}{N-a}\right) \cdot \Big((N-1)\text{tr}\Big(\mathbf{H}\big((a-1)\mathbf{H} + (N-a)\mathbf{G}\big)^-\Big) - r_2\Big)\,, \tag{3}$$

where τ and ρ are defined as in (2), and $r_2 = \text{rank}\left[(a-1)\mathbf{H} + (N-a)\mathbf{G}\right]$.

Note that A^- denotes the Moore-Penrose generalized inverse of A. The matrix Σ_Y denotes here the covariance matrix of the so-called *asymptotic rank transforms (ART)* of the observations. The concept of asymptotic rank transforms was introduced by Akritas (1990), in order to facilitate the theoretical derivation of asymptotic sampling distributions of rank-based tests. If we define the *average marginal distribution* for variable k as $\bar{F}_{\cdot\cdot}^{(k)}(x) = \frac{1}{N}\sum_{i=1}^{a} n_i F_i^{(k)}(x)$, then the ART of an observation $X_{ij}^{(k)}$ is $Y_{ij}^{(k)} = \bar{F}_{\cdot\cdot}^{(k)}(X_{ij}^{(k)})$. Σ_Y can then be defined as the $p \times p$ covariance matrix with elements

$$\Sigma_Y(k,k') = \text{Cov}(Y_{11}^{(k)}, Y_{11}^{(k')})\ . \tag{4}$$

The asymptotic results stated above for the nonparametric ANOVA-Type, Lawley-Hotelling, and Bartlett-Nanda-Pillai tests are, under straightforward modifications, also valid for the corresponding test statistics that are defined analogously in terms of the original observations instead of the ranks if we assume that the covariance matrix of the original observations, Σ_X, with elements $\Sigma_X(k,k') = \text{Cov}(X_{11}^{(k)}, X_{11}^{(k')})$, exists.

2.2.2 Results for Large Sample Size n

ANOVA-Type Statistic.

$$(a-1) \cdot \frac{(\text{tr}\mathbf{G})(\text{tr}\mathbf{H})}{\text{tr}(\mathbf{G}^2)} \tag{5}$$

has under $H_0 : F_1 = \cdots = F_a$, approximately a $\chi^2_{\hat{f}}$-distribution with estimated degrees of freedom $\hat{f} = (a-1) \cdot (\text{tr}\mathbf{G})^2/\text{tr}(\mathbf{G}^2)$ (Munzel and Brunner, 2000a,b). For F-distribution approximations to the ANOVA-type test statistic, see Section 3.1 below.

Lawley-Hotelling-Type Statistic.

$$(a-1)\mathrm{tr}\big(\mathbf{H}\mathbf{G}^{-1}\big) \tag{6}$$

has under $H_0 : F_1 = \cdots = F_a$, as $n \to \infty$, a, p fixed, asymptotically a $\chi^2_{p(a-1)}$-distribution (Munzel and Brunner, 2000b).

Bartlett-Nanda-Pillai-Type Statistic.

$$(N-a)\cdot \mathrm{tr}\Big((a-1)\mathbf{H}\big((a-1)\mathbf{H}+(N-a)\mathbf{G}\big)^{-}\Big) \tag{7}$$

has under $H_0 : F_1 = \cdots = F_a$, as $n \to \infty$, a, p fixed, asymptotically a $\chi^2_{\rho(a-1)}$-distribution with $\rho = \mathrm{rank}(\Sigma_Y) \leq p$ (Harrar and Bathke, 2007).

2.3 Asymptotic Results for the Unbalanced Design

In the unbalanced design (i.e., the sample sizes $n_1, \ldots, n_a$ are different), there are different ways to define hypothesis and error sums of squares for the multivariate tests. Define the matrices

$$\begin{aligned}
\mathbf{H}_1 &= \frac{1}{a-1}\sum_{i=1}^{a} n_i(\bar{\mathbf{R}}_{i\cdot} - \bar{\mathbf{R}}_{\cdot\cdot})(\bar{\mathbf{R}}_{i\cdot} - \bar{\mathbf{R}}_{\cdot\cdot})' \,, \\
\mathbf{H}_2 &= \frac{1}{a-1}\sum_{i=1}^{a} (\bar{\mathbf{R}}_{i\cdot} - \tilde{\mathbf{R}}_{\cdot\cdot})(\bar{\mathbf{R}}_{i\cdot} - \tilde{\mathbf{R}}_{\cdot\cdot})' \,, \\
\mathbf{G}_1 &= \frac{1}{N-a}\sum_{i=1}^{a}\sum_{j=1}^{n_i} (\mathbf{R}_{ij} - \bar{\mathbf{R}}_{i\cdot})(\mathbf{R}_{ij} - \bar{\mathbf{R}}_{i\cdot})' \,, \\
\mathbf{G}_2 &= \frac{1}{a-1}\sum_{i=1}^{a}\left(1-\frac{n_i}{N}\right)\frac{1}{n_i-1}\sum_{j=1}^{n_i} (\mathbf{R}_{ij} - \bar{\mathbf{R}}_{i\cdot})(\mathbf{R}_{ij} - \bar{\mathbf{R}}_{i\cdot})' \,, \\
\mathbf{G}_3 &= \frac{1}{a}\sum_{i=1}^{a}\frac{1}{n_i(n_i-1)}\sum_{j=1}^{n_i} (\mathbf{R}_{ij} - \bar{\mathbf{R}}_{i\cdot})(\mathbf{R}_{ij} - \bar{\mathbf{R}}_{i\cdot})' \,.
\end{aligned}$$

Here, we use the following notation to indicate two different ways to define an overall mean. The elements of the $p \times 1$ vector $\bar{\mathbf{R}}_{\cdot\cdot}$ are $\bar{R}^{(k)}_{\cdot\cdot} = \frac{1}{N}\sum_{i=1}^{a}\sum_{j=1}^{n_i} R^{(k)}_{ij}$, whereas the elements of $\tilde{\mathbf{R}}_{\cdot\cdot}$ are $\tilde{R}^{(k)}_{\cdot\cdot} = \frac{1}{a}\sum_{i=1}^{a}\bar{R}^{(k)}_{i\cdot} = \frac{1}{a}\sum_{i=1}^{a}\frac{1}{n_i}\sum_{j=1}^{n_i} R^{(k)}_{ij}$. If the design is balanced, both expressions are identical.

The two matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ both calculate mean squares due to treatment. However, $\mathbf{H}_1$ uses a weighted average (weighted by sample sizes), whereas $\mathbf{H}_2$ uses an unweighted average. In a similar way, the different error sum of squares matrices $\mathbf{G}_1$, $\mathbf{G}_2$, and $\mathbf{G}_3$ correspond to different weights.

The multivariate tests are now constructed similar to the balanced design. They are based on one of the pairs $(\mathbf{H}_1, \mathbf{G}_1)$, $(\mathbf{H}_1, \mathbf{G}_2)$, or $(\mathbf{H}_2, \mathbf{G}_3)$. In each of these pairs, both matrices are under the null hypothesis consistent estimators of the same covariance matrix. Note that in a balanced design with $n_i \equiv n$, $i = 1, \ldots, a$, the matrices $\mathbf{G}_1$ and $\mathbf{G}_2$ are identical, and furthermore, $\mathbf{H}_2 = n^{-1} \cdot \mathbf{H}_1$ and $\mathbf{G}_3 = n^{-1} \cdot \mathbf{G}_1 = n^{-1} \cdot \mathbf{G}_2$. Because of these relations, in a balanced design, each of the three pairs will lead to the same test statistic which is also equivalent to the procedure based on the pair $(\mathbf{H}, \mathbf{G})$ defined for the balanced situation.

2.3.1 Results for a Large Number of Samples a

Define

$$\bar{n} = \frac{1}{a}\sum_{i=1}^{a} n_i \quad \text{and} \quad \underline{n} = \frac{1}{a}\sum_{i=1}^{a} \frac{1}{n_i}.$$

The following standardized multivariate nonparametric test statistics have under $H_0 : F_1 = \cdots = F_a$, as $a \to \infty$, n_i, p fixed, asymptotically a standard normal distribution. The corresponding asymptotic theory can be found in Harrar and Bathke (2006).

2.3.2 Tests Based on $(\mathbf{H}_1, \mathbf{G}_1)$

ANOVA-Type Statistic.

$$\sqrt{\frac{a}{\hat{\tau}_{AN}^{(1)}}}\left(\frac{\mathrm{tr}(\mathbf{H}_1)}{\mathrm{tr}(\mathbf{G}_1)} - 1\right), \qquad \text{where} \tag{8}$$

$$\hat{\tau}_{AN}^{(1)} = \frac{1}{(\mathrm{tr}\mathbf{G}_1)^2}\left(\frac{2\bar{n}}{\bar{n}-1}\mathrm{tr}(\mathbf{G}_1^2) + \frac{\bar{n}(\bar{n}\underline{n}-1)}{(\bar{n}-1)^2}\left(\hat{\mu}_4 - 2\mathrm{tr}(\mathbf{G}_1^2) - (\mathrm{tr}\mathbf{G}_1)^2\right)\right)$$

$$\text{and} \quad \hat{\mu}_4 = \frac{1}{N}\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left[(\mathbf{R}_{ij} - \frac{N+1}{2}\mathbf{1})'(\mathbf{R}_{ij} - \frac{N+1}{2}\mathbf{1})\right]^2.$$

Lawley-Hotelling-Type Statistic.

$$\sqrt{\frac{a}{\hat{\tau}_{LH}^{(1)}}}\left(\mathrm{tr}(\mathbf{H}_1\mathbf{G}_1^-) - r_1\right), \qquad \text{where } r_1 \text{ is the rank of } \mathbf{G}_1, \tag{9}$$

$$\hat{\tau}_{LH}^{(1)} = \frac{2\bar{n}r_1}{\bar{n}-1} + \frac{\bar{n}(\bar{n}\underline{n}-1)}{(\bar{n}-1)^2}(\hat{\mu}_4 - 2r_1 - r_1^2), \tag{10}$$

$$\text{and} \quad \hat{\mu}_4 = \frac{1}{N}\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left[(\mathbf{R}_{ij} - \frac{N+1}{2}\mathbf{1})'(\frac{1}{N^2}\mathbf{G}_1)^-(\mathbf{R}_{ij} - \frac{N+1}{2}\mathbf{1})\right]^2.$$

Bartlett-Nanda-Pillai-Type Statistic.

$$\sqrt{\frac{a}{\hat{\tau}_{LH}^{(1)}}}\left(\frac{N-1}{N-a}\right)\left((N-1)\mathrm{tr}\Big(\mathbf{H}_1\big((a-1)\mathbf{H}_1 + (N-a)\mathbf{G}_1\big)^-\Big) - r_2\right), \tag{11}$$

where r_2 is the rank of $(a-1)\mathbf{H}_1 + (N-a)\mathbf{G}_1$, and $\hat{\tau}_{LH}^{(1)}$ is defined as in display (10).

2.3.3 Tests Based on $(\mathbf{H}_1, \mathbf{G}_2)$

ANOVA-Type Statistic.

$$\sqrt{\frac{a}{\hat{\tau}_{AN}^{(2)}}}\left(\frac{\text{tr}(\mathbf{H}_1)}{\text{tr}(\mathbf{G}_2)} - 1\right) \ , \quad \text{where} \quad \hat{\tau}_{AN}^{(2)} = \frac{\text{tr}(\mathbf{G}_2^2)}{(\text{tr}\mathbf{G}_2)^2}\frac{2}{a}\sum_{i=1}^{a}\frac{n_i}{n_i - 1}(1 - \frac{n_i}{N})^2 \ . \tag{12}$$

Lawley-Hotelling-Type Statistic.

$$\sqrt{\frac{a}{\hat{\tau}_{LH}^{(2)}}}\left(\text{tr}(\mathbf{H}_1\mathbf{G}_2^-) - r_1^{(2)}\right) \ , \tag{13}$$

$$\text{where } r_1^{(2)} \text{ is the rank of } \mathbf{G}_2 \text{ and } \ \hat{\tau}_{LH}^{(2)} = \text{rank}(\mathbf{G}_2)\frac{2}{a}\sum_{i=1}^{a}\frac{n_i}{n_i - 1}(1 - \frac{n_i}{N})^2 \ .$$

2.3.4 Tests Based on $(\mathbf{H}_2, \mathbf{G}_3)$

ANOVA-Type Statistic.

$$\sqrt{\frac{a}{\hat{\tau}_{AN}^{(3)}}}\left(\frac{\text{tr}(\mathbf{H}_2)}{\text{tr}(\mathbf{G}_3)} - 1\right) \ , \quad \text{where} \quad \hat{\tau}_{AN}^{(3)} = \frac{\text{tr}\mathbf{G}_2^2}{(\text{tr}\mathbf{G}_3)^2}\frac{2}{a}\sum_{i=1}^{a}\frac{1}{n_i(n_i - 1)} \ . \tag{14}$$

Lawley-Hotelling-Type Statistic.

$$\sqrt{\frac{a}{\hat{\tau}_{LH}^{(3)}}}\left(\text{tr}(\mathbf{H}_2\mathbf{G}_3^-) - r_1^{(3)}\right) \ , \tag{15}$$

$$\text{where } r_1^{(3)} \text{ is the rank of } \mathbf{G}_3 \text{ and } \ \hat{\tau}_{LH}^{(3)} = \text{tr}(\mathbf{G}_2\mathbf{G}_3^-)^2\frac{2}{a}\sum_{i=1}^{a}\frac{1}{n_i(n_i - 1)} \ .$$

2.3.5 Results for Large Sample Sizes n_i

The results mentioned in Section 2.2.2 are also valid in the unbalanced case, as $\min n_i \to \infty$. The ANOVA-type and Lawley-Hotelling-type statistics proposed by Munzel and Brunner (2000a,b) are based on $(\mathbf{H}_2, \mathbf{G}_3)$, whereas the Bartlett-Nanda-Pillai-type statistic suggested in Harrar and Bathke (2007) is based on $(\mathbf{H}_1, \mathbf{G}_1)$.

3 Finite Approximations and Comparison

The above test statistics are derived under the theoretical assumption that $a \to \infty$ or $\min n_i \to \infty$. In practice, these design constants are always finite. We consider the following approximations and investigate their relative performance.

1. F distribution approximations (Brunner, Dette, and Munk 1997; Srivastava and Fujikoshi 2006) with estimated degrees of freedom (ANOVA-type statistic)

2. Chi-squared distribution expansion (Lawley-Hotelling and Bartlett-Nanda-Pillai tests)

3. Fujikoshi (1975) expansion (Lawley-Hotelling and Bartlett-Nanda-Pillai tests)

4. McKeon (1974) approximation for the Lawley-Hotelling test

5. Muller (1998) approximation for the Bartlett-Nanda-Pillai tests

3.1 F Distribution Approximations for the ANOVA-Type Statistic

The distribution of the ANOVA-type statistic $\text{tr}(\mathbf{H})/\text{tr}(\mathbf{G})$ can be approximated by a χ^2 distribution or by an F distribution with estimated degrees of freedom. Munzel and Brunner (2000a,b) suggest to approximate the distribution of $\text{tr}(\mathbf{H}_2)/\text{tr}(\mathbf{G}_3)$ by an F distribution with estimated numerator degrees of freedom $\hat{f}$ and denominator degrees of freedom ∞. This is equivalent to using a $\chi^2_{\hat{f}}$-distribution, divided by $\hat{f}$. The degrees of freedom estimator is given by

$$\hat{f} = (a-1)\frac{\text{tr}(\mathbf{G}_3)^2}{\text{tr}(\mathbf{G}_3^2)} \; . \tag{16}$$

Alternatively, following the approach by Brunner, Dette, and Munk (1997), is is possible to construct an F distribution approximation with estimated numerator and denominator degrees of freedom. The numerator d.f. estimator is given by equation (16) above, and the denominator d.f. are estimated by

$$\hat{f}_0 = \frac{a^2}{(a-1)\sum\limits_{i=1}^{a}\frac{1}{n_i-1}} \cdot \hat{f}. \tag{17}$$

This approximation is originally derived for the ANOVA-type statistic based on $(\mathbf{H}_2, \mathbf{G}_3)$, but the same d.f. estimators also provide a good approximation for the test statistic based on $(\mathbf{H}_1, \mathbf{G}_1)$.

Based on Bai and Saranadasa's (1996) test for the multivariate two-sample problem Srivastava and Fujikoshi (2006) have suggested the following degrees of freedom estimators for the parametric ANOVA-Type statistic. We have included this approximation, that is specifically designed for the large p situation, in the following form into our simulation study.

The distribution of $\text{tr}(\mathbf{H})/\text{tr}(\mathbf{G})$ is approximated by an F distribution with numerator degrees of freedom $(a-1)f_S$ and denominator degrees of freedom $(N-a)f_S$, where

$$f_S = \frac{(N-a-1)(N-a+2)}{(N-a)^2}\frac{\text{tr}(\mathbf{G})^2}{\text{tr}(\mathbf{G})^2 - \frac{1}{N-a}\text{tr}(\mathbf{G}^2)} \tag{18}$$

3.2 Chi-Squared Distribution Expansion for Lawley-Hotelling and Bartlett-Nanda-Pillai Tests Based on $(\mathbf{H}_1, \mathbf{G}_1)$

For the Lawley-Hotelling and Bartlett-Nanda-Pillai tests, an approximation is possible based on an expansion of the quantiles of the sampling distribution of the test statistics (Ito 1956, 1960; Fujikoshi 1973; Muirhead 1970; Davis 1970; see also Anderson, 1984, pp.325–328 and Munzel and Brunner, 2002a,b).

Instead of using the asymptotic normal distribution, the upper α-quantile of the distribution of $(a-1)\text{tr}(\mathbf{H}_1\mathbf{G}_1^-)$ (Lawley-Hotelling test) is approximated by

$$\chi^2_{p(a-1);\alpha} + \frac{1}{2(N-a)}\Big(\frac{p+a}{p(a-1)+2}\chi^4_{p(a-1);\alpha} + (p-a+2)\chi^2_{p(a-1);\alpha}\Big), \tag{19}$$

where $\chi^2_{p(a-1);\alpha}$ denotes the upper α-quantile of the χ^2-distribution with $p(a-1)$ degrees of freedom.

Similarly, the upper α-quantile of $(N-a)\text{tr}\Big((a-1)\mathbf{H}_1[(a-1)\mathbf{H}_1 + (N-a)\mathbf{G}_1]^-\Big)$ (Bartlett-Nanda-Pillai test) is approximated by

$$\chi^2_{p(a-1);\alpha} + \frac{1}{2(N-a)}\Big(-\frac{p+a}{p(a-1)+2}\chi^4_{p(a-1);\alpha} + (p-a+2)\chi^2_{p(a-1);\alpha}\Big). \tag{20}$$

3.3 Fujikoshi Approximation for Tests Based on $(\mathbf{H}_1, \mathbf{G}_1)$

An alternative approximation is due to Fujikoshi (1975). It uses a Cornish-Fisher expansion for the quantiles of the distribution of the test statistics. Designed originally for the $a > n_i$ case, it also performs well when $n_i > a$.

Fujikoshi (1975) derived asymptotic expansion formulae for approximating the null distributions of the Lawley-Hotelling and Barlett-Nanda-Pillai test statistics for the multivariate linear hypothesis under normality. The asymptotic framework is that both hypothesis and error degrees of freedom tend to infinity.

We have used the asymptotic expansions of Fujikoshi to calculate the Cornish-Fisher expansions for the percentiles. Denote the first six Hermite polynomials by: $h_1(x) = 1$, $h_2(x) = -x$, $h_3(x) = x^2 - 1$, $h_4(x) = -x^3 + 3x$, $h_5(x) = x^4 - 6x^2 + 3$, and $h_6(x) = -x^5 + 10x^3 - 15x$. Let $m = \mu^{-1}(N-1) - r$, and define

$$e = \frac{N-a}{N-1} \text{ and } h = \frac{a-1}{N-1}. \tag{21}$$

Then, for $0 < \alpha < 1$, the upper α-quantile for both Lawley-Hotelling and Bartlett-Nanda-Pillai test have the following general form.

$$\begin{aligned} Z_\alpha &+ \frac{1}{\sqrt{m}}\{a_1h_1(Z_\alpha) + a_3h_3(Z_\alpha)\} - \frac{1}{m}\Big\{b_2h_2(Z_\alpha) + b_4h_4(Z_\alpha) + b_6h_6(Z_\alpha) \\ &+ Z_\alpha(a_1 + a_3h_3(Z_\alpha))\left(\frac{1}{2}a_1 + a_3\left(\frac{1}{2}h_3(Z_\alpha) - 2\right)\right)\Big\} \end{aligned}$$

where Z_α denotes the upper α-quantile of a standard normal variate.

For the Lawley-Hotelling statistic defined by

$$\sqrt{\frac{m}{\tau^2}} \cdot \frac{a-1}{N-a} \cdot \left[\mathrm{tr}(\mathbf{H}_1\mathbf{G}_1^{-1}) - p\right] \tag{22}$$
$$\text{where} \quad \tau^2 = 2p\frac{(a-1)(N-1)}{(N-a)^2},\ m = (N-a)-(p+1),$$

we use $\mu^{-1} = e$, $r = p+1$, $a_1 = \frac{1}{\tau}\left(p(p+1)h(\mu e^2)^{-1}\right)$, $a_3 = \frac{1}{\tau^3}\left(4ph(2-e)(3\mu^2e^5)^{-1}\right)$, $b_2 = \frac{1}{\tau^2}\left(p(p+1)[\frac{1}{2}(p^2+p+8)h^2+3he](\mu^2e^4)^{-1} - rph(\mu e^3)^{-1}\right)$, $b_4 = \frac{1}{\tau^4}\left(2ph[\frac{2}{3}p(p+1)h(2-e)+e^2-5e+5](\mu^3e^7)^{-1}\right)$, and $b_6 = \frac{1}{2}a_3^2$.

For the Bartlett-Nanda-Pillai statistic defined by

$$\sqrt{\frac{m}{\tau^2}}\left[\mathrm{tr}\left((a-1)\mathbf{H}_1[(a-1)\mathbf{H}_1 + (N-a)\mathbf{G}_1]^{-1}\right) - ph\right] \tag{23}$$
$$\text{where} \quad \tau^2 = 2p\frac{(a-1)(N-a)}{(N-1)^2},\ m = (N-1),$$

we use $\mu^{-1} = 1$, $r = 0$, $a_1 = 0$, $a_3 = \frac{1}{\tau^3}\left(\frac{4}{3}phe\mu^{-2}(e-h)\right)$, $b_2 = \frac{1}{\tau^2}\left(-phe\mu^{-2}(p+1)\right)$, $b_4 = \frac{1}{\tau^4}\left(2phe\mu^{-3}(e^2+h^2-3he)\right)$, and $b_6 = \frac{1}{2}a_3^2$.

3.4 Fujikoshi Approximation for Tests Based on $(\mathbf{H}_1, \mathbf{G}_2)$ and $(\mathbf{H}_2, \mathbf{G}_3)$

Similar approximations to Lawley-Hotelling type tests as in the previous section can be constructed based on the hypothesis and error matrices $(\mathbf{H}_1, \mathbf{G}_2)$ and $(\mathbf{H}_2, \mathbf{G}_3)$. The detailed derivations can be found in the technical report (Bathke, Harrar, and Madden 2007). For $(\mathbf{H}_1, \mathbf{G}_2)$, define

$$k = \frac{1}{a-1}\sum_{i=1}^{a}\left(1-\frac{n_i}{N}\right)^2\frac{1}{n_i-1} \quad \text{and} \quad \nu = \frac{(a-1)^2}{\sum_{i=1}^{a}\left(1-\frac{n_i}{N}\right)^2\frac{1}{n_i-1}},$$
$$\mathbf{G}_2^* = (1/k)\mathbf{G}_2, \quad h = \frac{a-1}{N-1}, \quad \text{and} \quad e = \frac{\nu}{N-1}.$$

Then, the procedure described in the preceding section can be applied, with the terms e and h in display (21) replaced by the corresponding terms defined here, and the matrix $\mathbf{G}_2^*$ replacing $\mathbf{G}_1$ in equation (22).

For the test based on $(\mathbf{H}_2, \mathbf{G}_3)$, define

$$k_1 = \frac{1}{a-1}\frac{\sum_{i=1}^{a}\frac{1}{n_i^2}}{\sum_{i=1}^{a}\frac{1}{n_i}}, \quad \nu_1 = \frac{a-1}{a}\frac{\left(\sum_{i=1}^{a}\frac{1}{n_i}\right)^2}{\sum_{i=1}^{a}\frac{1}{n_i^2}}, \quad k_2 = \frac{1}{a}\frac{\sum_{i=1}^{a}\frac{1}{n_i^2(n_i-1)}}{\sum_{i=1}^{a}\frac{1}{n_i}}, \quad \nu_2 = \frac{\left(\sum_{i=1}^{a}\frac{1}{n_i}\right)^2}{\sum_{i=1}^{a}\frac{1}{n_i^2(n_i-1)}},$$
$$\mathbf{H}_2^* = (1/k_1)\mathbf{H}_2, \quad \mathbf{G}_3^* = (1/k_2)\mathbf{G}_3, \quad h = \frac{\nu_1}{N-1}, \quad \text{and} \quad e = \frac{\nu_2}{N-1},$$

and substitute $\mathbf{H}_2^*$ and $\mathbf{G}_3^*$ for $\mathbf{H}_1$ and $\mathbf{G}_1$ respectively in display (22).

3.5 McKeon Approximation for the Lawley-Hotelling Test

McKeon (1974) suggested an approximation for the normal theory sampling distribution of the Lawley-Hotelling test. This approximation is based on fitting the first two moments of an F distribution, while specifying the value of the numerator degrees of freedom. Let $U = \operatorname{tr}\Big((a-1)\mathbf{H}\big((N-a)\mathbf{G}\big)^{-}\Big)$. The distribution of U is approximated by $g \cdot F_{K,D}$, a "stretched" F distribution with degrees of freedom K and D. The quantities g, K, and D are defined in the following equations.

$$K = p(a-1), \quad D = 4 + \frac{p(a-1)+2}{B-1}, \text{ where } B = \frac{(N-p-2)(N-a-1)}{(N-a-p)(N-a-p-3)},$$
$$\text{and} \quad g = \frac{p(a-1)(D-2)}{(N-a-p-1)D}.$$

3.6 Muller Approximation for the Bartlett-Nanda-Pillai Test

Based on McKeon's (1974) moment estimator approach for the Lawley-Hotelling test, Muller (1998) proposed the following F approximation for the distribution of the parametric Bartlett-Nanda-Pillai test that appears to perform better than the approximation originally proposed by Pillai (1955).

Let $V = \operatorname{tr}\Big((a-1)\mathbf{H}\big((a-1)\mathbf{H} + (N-a)\mathbf{G}\big)^{-}\Big)$. The distribution of $\frac{(V/\gamma)/\nu_1}{(1-V/\gamma)/\nu_2}$ is approximated by an F-distribution with degrees of freedom ν_1 and ν_2 that are defined in the following equations.

$$\gamma = \min(a-1, p),$$
$$\nu_1 = \frac{p(a-1)}{\gamma(N-1)}\Big[\frac{\gamma(N-a+\gamma-p)(N+2)(N-1)}{(N-a)(N-p)} - 2\Big], \text{ and}$$
$$\nu_2 = \frac{N-a+\gamma-p}{N}\Big[\frac{\gamma(N-a+\gamma-p)(N+2)(N-1)}{(N-a)(N-p)} - 2\Big].$$

4 Simulation Study

In order to compare the different tests and approximations for finite samples, we have carried out extensive simulation studies using SAS IML. Note that the available tests can be compared on different levels: First comparing the three different test types ANOVA-type, Lawley-Hotelling, and Bartlett-Nanda-Pillai; then studying the effect of the different hypothesis and error sums of squares in unbalanced designs, and finally investigating the different small sample approximations. In the simulation studies, we have chosen design parameters that reasonably resemble many real life applications for plant disease response data with small a and n_i. This led to the following choices.

- The number of variables is $p = 2$, $p = 4$, or $p = 16$ (the latter for an extreme condition).
- The number of samples is $a = 6$ or $a = 12$.

- The design is either balanced with sample size $n = 4$ or $n = 6$, or it is unbalanced with $n_i = 3$ and $n_i = 4$ each for half of the samples (denoted $n = 3, 4$ in the simulation tables) or with $n_i = 5$ and $n_i = 6$ each for half of the samples (denoted $n = 5, 6$).

- The underlying distribution of the simulated observations is multivariate normal with 10% outliers, and with different correlation structures between the p variables: Strong positive correlation, that is, all pairwise correlations are 0.9; moderate positive correlation, where all pairwise correlations are 0.5; and negative correlation where all pairwise correlations between the p variables are -0.5 ($p = 2$), -0.3 ($p = 4$), or -0.05 ($p = 16$). The different chosen negative correlation values are due to the fact that, for example, a multivariate normal distribution with $p = 16$ and all pairwise correlations of -0.5 would not have a positive definite covariance matrix. One of the variables (for $p = 16$: four of the variables) is then rounded to an ordinal scale with 10 levels. This corresponds to the commonly arising situation that the variables are partially ordinal, and partially quantitative.

- The power is simulated by location shifts (adding a constant to the observations) in one of the following two ways.
  - The observations in one sixth of the factor levels are shifted by 1.0, whereas there is no location shift for the other levels.
  - The observations in one third of the factor levels are shifted by 0.5, and in another third of the levels, the observations are shifted by 1.0.

- The number of simulation runs is always 10,000, resulting in a standard error of less than 0.002 for the α-level simulations, and not more than 0.005 for the power simulations.

In Tables 2-5, we present simulated α-level and power for the most preferable nonparametric test statistics. These are:

- ANOVA Type test: asymptotic version (A_{as}, see equation 8), and two different F approximations (A_F, based on equations 16 and 17; A_{FS}, see equation 18)

- Lawley-Hotelling test: approximations by McKeon (1974) (LH_{McK}, see Section 3.5) and Fujikoshi (1975) (LH_{Fu}, see equation 22 and Section 3.4)

- Bartlett-Nanda-Pillai test: approximations by Fujikoshi (1975) (BNP_{Fu}, see equation 23) and Muller (1998) (BNP_{Mu}, see Section 3.6)

In general, those tests for which no simulation results are reported below performed less favorably in the simulations.

In the following, we summarize the key findings from our simulation study which can serve as recommendations on how to choose the best nonparametric multivariate tests for the one-way layout.

Correlation	Test Statistic	Balanced Design, Number of Variables $p = 2$			
		$a = 6$		$a = 12$	
		$n = 4$	$n = 6$	$n = 4$	$n = 6$
$q = 0.9$	A_{as}	11.2 (25.6, 29.3)	9.7 (33.7, 39.5)	8.6 (28.2, 34.6)	8.2 (41.3, 50.5)
	A_F	5.5 (14.6, 18.4)	5.2 (22.6, 29.0)	5.4 (19.3, 25.0)	5.2 (31.7, 40.9)
	A_{FS}	6.0 (16.8, 20.0)	6.3 (24.0, 30.4)	6.0 (20.6, 26.5)	5.7 (32.8, 42.6)
	LH_{McK}	4.9 (10.8, 13.7)	4.8 (16.3, 21.7)	5.2 (14.3, 18.5)	4.4 (22.9, 31.8)
	LH_{Fu}	4.9 (11.0, 14.2)	4.9 (16.6, 21.9)	5.2 (14.5, 18.9)	4.5 (23.1, 32.0)
	BNP_{Mu}	5.0 (10.8, 13.1)	5.1 (15.8, 21.1)	4.7 (14.3, 17.2)	4.9 (21.6, 29.0)
	BNP_{Fu}	4.9 (10.7, 12.9)	5.0 (15.7, 21.0)	4.6 (14.2, 17.2)	4.9 (21.6, 28.9)
$q = 0.5$	A_{as}	9.8 (25.4, 30.9)	9.1 (35.4, 41.7)	9.0 (29.1, 37.2)	7.9 (44.6, 54.4)
	A_F	4.9 (15.1, 18.7)	4.4 (23.9, 30.4)	4.7 (21.4, 27.1)	5.1 (35.5, 45.6)
	A_{FS}	5.2 (16.6, 20.1)	5.7 (25.8, 31.5)	5.5 (21.2, 27.3)	5.1 (35.9, 46.4)
	LH_{McK}	5.2 (11.9, 16.1)	5.4 (19.2, 25.5)	5.2 (16.8, 22.3)	5.2 (27.7, 36.6)
	LH_{Fu}	5.3 (12.9, 16.8)	5.5 (19.4, 25.8)	5.3 (17.1, 22.6)	5.2 (28.0, 36.6)
	BNP_{Mu}	5.4 (12.6, 15.5)	5.3 (19.0, 23.3)	5.3 (15.5, 20.3)	5.2 (26.6, 34.2)
	BNP_{Fu}	5.3 (12.4, 15.3)	5.2 (18.8, 23.1)	5.3 (15.5, 20.2)	5.2 (26.6, 34.2)
$q = -0.5$	A_{as}	10.0 (24.9, 29.5)	9.0 (33.1, 42.1)	8.5 (28.4, 36.2)	7.9 (43.8, 56.1)
	A_F	4.8 (12.9, 16.9)	4.9 (22.5, 30.4)	5.0 (18.3, 25.6)	4.9 (34.3, 45.1)
	A_{FS}	5.3 (14.1, 18.4)	5.6 (23.2, 30.5)	5.4 (20.0, 25.8)	5.1 (33.3, 45.9)
	LH_{McK}	5.0 (20.6, 26.9)	5.2 (35.3, 44.7)	5.2 (31.5, 40.7)	5.4 (53.0, 64.0)
	LH_{Fu}	5.2 (21.2, 27.4)	5.3 (35.6, 45.0)	5.3 (31.9, 41.0)	5.5 (53.2, 64.3)
	BNP_{Mu}	5.1 (20.4, 24.8)	5.2 (32.2, 41.2)	5.2 (28.5, 35.6)	4.6 (49.4, 61.4)
	BNP_{Fu}	5.0 (20.3, 24.6)	5.1 (32.0, 41.0)	5.1 (28.5, 35.5)	4.6 (49.3, 61.4)

Table 2: Simulated α-levels [in percent] for the proposed nonparametric multivariate tests (in parentheses: their respective power for location shifts as described in the text). Nominal α is 5%. Number of simulations is 10,000.

Correlation	Test Statistic	Balanced Design, Number of Variables $p = 4$			
		$a = 6$		$a = 12$	
		$n = 4$	$n = 6$	$n = 4$	$n = 6$
$q = 0.9$	A_{as}	10.5 (27.4, 34.3)	9.5 (37.7, 46.0)	9.0 (31.9, 39.6)	8.0 (48.6, 57.9)
	A_F	6.1 (17.8, 22.2)	5.5 (27.6, 35.0)	6.1 (24.0, 31.7)	6.1 (40.0, 49.5)
	A_{FS}	8.1 (22.3, 27.5)	6.8 (31.8, 38.8)	7.4 (28.1, 36.4)	6.6 (42.9, 53.1)
	LH_{McK}	4.3 (8.1, 11.2)	4.5 (12.9, 17.8)	4.8 (11.5, 15.6)	5.0 (19.2, 24.7)
	LH_{Fu}	5.0 (9.7, 12.1)	4.8 (12.9, 18.6)	5.0 (12.0, 16.4)	4.9 (19.7, 26.1)
	BNP_{Mu}	4.6 (8.6, 10.7)	4.5 (12.7, 16.2)	4.7 (11.1, 13.5)	4.9 (18.2, 22.8)
	BNP_{Fu}	4.3 (8.2, 10.2)	4.4 (12.4, 16.0)	4.7 (11.0, 13.4)	4.9 (18.1, 22.6)
$q = 0.5$	A_{as}	8.9 (29.7, 36.7)	8.4 (44.5, 52.6)	7.7 (38.3, 47.1)	7.8 (58.4, 68.6)
	A_F	4.4 (18.9, 24.7)	5.1 (33.8, 43.8)	4.9 (29.6, 38.4)	5.0 (49.4, 62.0)
	A_{FS}	6.1 (23.7, 29.2)	5.9 (36.4, 46.7)	6.0 (33.1, 41.2)	5.8 (53.1, 63.3)
	LH_{McK}	5.7 (11.3, 15.0)	5.1 (18.5, 25.1)	4.9 (16.0, 21.3)	4.9 (27.2, 37.5)
	LH_{Fu}	5.9 (13.0, 16.6)	5.4 (19.5, 26.2)	5.2 (16.8, 22.4)	5.1 (27.7, 38.1)
	BNP_{Mu}	5.3 (12.4, 13.7)	5.1 (17.3, 22.0)	5.0 (15.1, 19.1)	4.9 (24.8, 31.6)
	BNP_{Fu}	5.1 (12.0, 13.4)	5.0 (17.0, 21.5)	4.9 (15.0, 19.0)	4.9 (24.7, 31.4)
$q = -0.3$	A_{as}	7.8 (27.5, 36.2)	8.0 (45.6, 58.0)	7.2 (37.7, 50.2)	6.9 (64.8, 78.8)
	A_F	4.1 (16.1, 22.1)	4.3 (33.5, 46.2)	4.2 (27.4, 40.3)	4.9 (55.6, 72.1)
	A_{FS}	4.9 (20.3, 26.9)	5.4 (36.4, 50.2)	5.0 (30.2, 42.4)	4.9 (57.3, 74.8)
	LH_{McK}	5.4 (42.9, 53.3)	5.6 (73.5, 81.4)	5.3 (68.8, 80.5)	5.2 (93.7, 96.9)
	LH_{Fu}	5.7 (45.5, 56.3)	5.9 (74.7, 82.2)	5.6 (69.7, 81.4)	5.3 (94.2, 97.0)
	BNP_{Mu}	5.1 (35.1, 38.2)	5.4 (61.5, 68.5)	5.2 (53.1, 60.5)	5.1 (84.8, 90.7)
	BNP_{Fu}	4.9 (34.4, 37.3)	5.3 (60.9, 68.0)	5.1 (52.9, 60.2)	5.0 (84.6, 90.5)

Table 3: Simulated α-levels [in percent] for the proposed nonparametric multivariate tests (in parentheses: their respective power for location shifts as described in the text). Nominal α is 5%. Number of simulations is 10,000.

Correlation	Test Statistic	Balanced Design, Number of Variables $p = 16$			
		$a = 6$		$a = 12$	
		$n = 4$	$n = 6$	$n = 4$	$n = 6$
$q = 0.9$	A_{as}	10.7 (30.7, 37.2)	9.4 (43.0, 50.8)	9.1 (36.8, 45.6)	8.8 (54.4, 65.7)
	A_F	6.6 (21.2, 26.9)	5.9 (32.3, 41.6)	6.2 (30.0, 38.1)	6.0 (47.7, 58.7)
	A_{FS}	10.4 (30.3, 35.5)	8.4 (39.8, 48.1)	9.2 (38.2, 46.2)	7.9 (53.7, 64.4)
	LH_{McK}	6.3 (8.0, 7.8)	4.0 (8.1, 10.1)	5.0 (9.0, 10.5)	4.7 (10.9, 14.3)
	LH_{Fu}	5.5 (7.2, 6.8)	5.6 (10.8, 13.0)	6.4 (11.1, 12.9)	4.9 (12.4, 15.9)
	BNP_{Mu}	4.2 (7.0, 6.1)	4.0 (7.7, 8.3)	4.1 (7.8, 7.5)	4.1 (10.0, 10.5)
	BNP_{Fu}	1.4 (2.3, 2.1)	2.9 (5.8, 5.9)	3.7 (7.0, 6.8)	4.0 (9.5, 10.0)
$q = 0.5$	A_{as}	7.6 (38.9, 46.7)	7.7 (58.9, 68.6)	7.0 (51.7, 62.4)	7.4 (76.0, 85.3)
	A_F	4.9 (28.2, 36.8)	5.1 (50.4, 60.8)	5.0 (45.3, 57.2)	5.5 (71.5, 82.3)
	A_{FS}	8.7 (39.9, 49.6)	7.2 (57.5, 67.7)	8.0 (53.5, 63.9)	7.3 (75.9, 86.2)
	LH_{McK}	6.6 (8.2, 7.9)	5.1 (10.9, 12.9)	5.0 (11.1, 12.2)	4.5 (17.4, 22.8)
	LH_{Fu}	5.8 (7.3, 7.0)	6.4 (14.1, 16.4)	6.3 (13.6, 15.3)	5.0 (18.6, 24.4)
	BNP_{Mu}	5.3 (9.2, 8.4)	5.3 (12.2, 12.0)	4.6 (10.9, 11.1)	4.3 (15.4, 17.8)
	BNP_{Fu}	1.8 (3.4, 2.9)	3.9 (9.4, 9.2)	4.0 (10.0, 10.1)	4.0 (14.9, 16.9)
$q = -0.05$	A_{as}	2.8 (47.2, 64.5)	4.5 (87.0, 95.4)	4.2 (76.8, 91.2)	4.7 (98.9, 99.9)
	A_F	1.3 (31.6, 49.0)	2.0 (78.8, 91.6)	2.3 (69.7, 86.4)	3.4 (98.3, 99.8)
	A_{FS}	4.8 (55.5, 70.9)	5.1 (88.4, 96.3)	4.6 (80.2, 92.0)	5.0 (98.9, 99.9)
	LH_{McK}	6.5 (14.9, 17.5)	5.2 (80.2, 89.1)	5.6 (82.5, 91.4)	5.5 (99.8, 100)
	LH_{Fu}	5.4 (13.7, 15.7)	6.7 (86.0, 92.0)	7.1 (85.9, 93.5)	5.7 (99.9, 100)
	BNP_{Mu}	5.2 (22.8, 18.6)	5.2 (49.0, 43.0)	5.2 (40.2, 38.5)	5.2 (75.5, 75.4)
	BNP_{Fu}	1.7 (10.6, 8.7)	3.7 (43.0, 36.8)	4.6 (37.9, 36.4)	4.9 (74.5, 74.3)

Table 4: Simulated α-levels [in percent] for the proposed nonparametric multivariate tests (in parentheses: their respective power for location shifts as described in the text). Nominal α is 5%. Number of simulations is 10,000.

		Unbalanced Design			
		$p=2$		$p=4$	
Correlation	Test	$a=12$			
	Statistic	$n=3,4$	$n=5,6$	$n=3,4$	$n=5,6$
	A_{as}	9.3 (25.3, 32.4)	8.8 (37.1, 45.6)	9.7 (28.6, 35.9)	8.3 (44.0, 53.5)
	$A_F(H_1,G_1)$	5.0 (15.9, 21.1)	5.6 (28.7, 36.8)	5.6 (20.2, 25.5)	5.7 (34.8, 45.2)
	$A_F(H_2,G_3)$	5.7 (16.7, 21.2)	5.6 (28.4, 36.3)	6.4 (20.7, 25.9)	5.9 (34.9, 44.5)
$q=0.9$	A_{FS}	6.1 (18.7, 22.9)	5.5 (29.3, 38.6)	7.8 (24.8, 31.8)	7.1 (39.0, 49.2)
	LH_{McK}	4.9 (12.6, 16.5)	4.8 (20.3, 27.1)	4.8 (10.5, 13.8)	4.5 (17.0, 22.5)
	$LH_{Fu}(H_1,G_1)$	5.0 (12.9, 16.7)	4.9 (20.5, 27.3)	5.2 (11.1, 14.5)	4.7 (17.3, 23.1)
	$LH_{Fu}(H_1,G_2)$	4.8 (12.6, 16.0)	4.8 (20.1, 26.9)	4.9 (10.5, 13.8)	4.5 (17.2, 22.7)
	$LH_{Fu}(H_2,G_3)$	3.6 (9.4, 12.3)	3.5 (16.3, 22.4)	3.3 (7.2, 9.9)	3.2 (13.2, 17.7)
	BNP_{Mu}	4.8 (12.7, 16.0)	4.8 (19.3, 25.9)	5.0 (10.1, 12.4)	4.5 (16.4, 19.8)
	BNP_{Fu}	4.8 (12.7, 15.9)	4.8 (19.3, 25.8)	5.0 (10.0, 12.3)	4.4 (16.2, 19.7)
	A_{as}	8.7 (26.7, 32.0)	8.2 (41.0, 50.1)	7.6 (32.3, 41.4)	8.1 (52.9, 63.8)
	$A_F(H_1,G_1)$	4.9 (16.3, 22.1)	5.2 (32.0, 41.5)	5.1 (23.4, 30.2)	5.1 (44.5, 56.7)
	$A_F(H_2,G_3)$	5.4 (17.3, 22.2)	5.2 (32.1, 41.4)	5.4 (23.7, 30.2)	5.3 (44.2, 56.3)
$q=0.5$	A_{FS}	5.2 (17.5, 23.4)	5.5 (31.3, 40.8)	6.2 (27.2, 34.5)	6.0 (46.3, 59.0)
	LH_{McK}	5.0 (14.3, 18.2)	4.9 (25.2, 32.9)	5.2 (13.5, 18.3)	5.2 (23.7, 32.9)
	$LH_{Fu}(H_1,G_1)$	5.1 (14.5, 18.6)	5.0 (25.3, 33.1)	5.7 (14.2, 19.3)	5.4 (24.0, 33.5)
	$LH_{Fu}(H_1,G_2)$	4.9 (14.2, 17.7)	4.9 (25.1, 32.8)	5.2 (13.8, 18.7)	5.4 (23.9, 33.1)
	$LH_{Fu}(H_2,G_3)$	3.6 (10.8, 14.0)	3.6 (20.8, 27.5)	3.4 (10.0, 13.9)	3.7 (19.1, 27.5)
	BNP_{Mu}	4.9 (14.2, 17.1)	4.9 (24.3, 30.4)	5.0 (12.7, 16.5)	5.3 (21.7, 28.2)
	BNP_{Fu}	4.9 (14.1, 17.1)	4.9 (24.2, 30.3)	4.9 (12.6, 16.3)	5.2 (21.6, 28.0)
	A_{as}	9.0 (25.0, 31.0)	8.1 (40.1, 50.9)	7.0 (32.6, 41.5)	6.6 (58.6, 73.0)
	$A_F(H_1,G_1)$	4.7 (16.5, 20.5)	5.1 (29.4, 40.0)	4.0 (22.1, 30.2)	4.0 (47.9, 64.5)
	$A_F(H_2,G_3)$	5.3 (16.7, 20.7)	5.3 (29.5, 39.9)	4.6 (22.2, 29.9)	4.2 (47.2, 63.6)
$q=-0.5$	A_{FS}	5.0 (16.8, 21.6)	4.8 (30.2, 41.0)	4.9 (24.7, 35.3)	5.3 (50.5, 66.8)
(p=2)	LH_{McK}	4.9 (25.9, 32.9)	4.9 (47.8, 59.6)	4.9 (58.4, 69.9)	4.7 (90.3, 95.1)
or	$LH_{Fu}(H_1,G_1)$	5.1 (26.3, 33.5)	5.0 (48.0, 59.8)	5.4 (59.9, 71.0)	4.9 (90.5, 95.3)
$q=-0.3$	$LH_{Fu}(H_1,G_2)$	4.8 (25.1, 32.6)	5.0 (47.6, 59.1)	5.1 (58.7, 69.8)	4.9 (90.3, 95.1)
(p=4)	$LH_{Fu}(H_2,G_3)$	3.4 (19.7, 26.0)	3.9 (41.5, 53.4)	3.7 (49.3, 61.0)	3.5 (87.1, 93.0)
	BNP_{Mu}	5.1 (23.6, 29.5)	4.9 (44.4, 55.0)	4.9 (43.7, 49.8)	4.5 (79.1, 86.3)
	BNP_{Fu}	5.1 (23.6, 29.5)	4.9 (44.3, 54.9)	4.8 (43.4, 49.6)	4.4 (78.9, 86.2)

Table 5: Simulated α-levels [in percent] for the proposed nonparametric multivariate tests (in parentheses: their respective power for location shifts as described in the text). Nominal α is 5%. Number of simulations is 10,000.

Note that, even though we have carried out an extensive simulation study and compared the performance of several different available tests in numerous settings that are of practical relevance, it is possible that a design setting could be constructed in which a different test has the highest power. We think that the suggestions below give a useful general view that should serve as an approximate guide, but when choosing the appropriate test, we also recommend carrying out a simulation study that best imitates the specific experimental design settings encountered in the application of interest. Summarizing the simulations, we found:

1. In general, the Bartlett-Nanda-Pillai type tests (BNP_{Mu} and BNP_{Fu}) were always a reliable, although usually a slightly conservative choice, never substantially exceeding the fixed α-level. However, they had low power, especially when the different variables were strongly positively correlated. In fact, they generally had the lowest power of all the tests considered. Regarding the different approximations, there was no practically significant difference between the performances of the Muller (1998) and the Fujikoshi (1975; Harrar and Bathke, 2007) approximations, unless the number p of variables was large and a was small, in which case the Muller approximation was preferable (see Table 4). Another practical advantage of the Muller approximation is that it is implemented in the SAS procedure Proc GLM.

2. Comparing the different approximations for the Lawley-Hotelling type test, the differences between the McKeon (1974) (LH_{McK}) and the Fujikoshi (1975) (LH_{Fu}) approximations were minimal in terms of α-levels and power, as were the differences between the hypothesis and error matrix pairs $(\mathbf{H}_1, \mathbf{G}_1)$ and $(\mathbf{H}_1, \mathbf{G}_2)$ in the unbalanced case. The pair $(\mathbf{H}_2, \mathbf{G}_3)$ seemed to lead to a more conservative test and subsequently lower power in an unbalanced design (Table 5). For balanced data sets, the Lawley-Hotelling tests were slightly anti-conservative for moderate positive or negative correlations, but at large positive correlation, the tests were slightly conservative, except for large p (Table 4), in which type I errors fluctuated around 5%.

3. The asymptotic version of the ANOVA Type statistics (A_{as}) performed poorly, as expected, with high type I error rates. Although these tests also had the highest power, their use could not be recommended based on the α-levels found (Tables 2-5). For balanced or unbalanced data, the α-levels somewhat exceeded 5% for the small-sample approximation of the ANOVA Type statistic when there was a high correlation of the response variables. For moderate positive or negative correlation, however, the type I error rates were lower, and the approximation (A_F) became conservative, especially with $p = 16$. The test based on use of estimated numerator and denominator d.f. (A_F, see equations 16 and 17) performed best overall, for balanced or unbalanced designs. However, with $p = 16$ and average negative correlation, only A_{FS} (see equation 18) gave α-levels close to 5%. For unbalanced designs, the performance of A_F using different pairs of hypothesis and error matrices was in general very similar. Combination $(\mathbf{H}_2, \mathbf{G}_3)$ led to somewhat anti-conservative tests when the response variables had a strong positive correlation.

4. In choosing the most powerful test, it is important to take the empirical correlation matrix between the different variables into account. For data with positively correlated variables, the ANOVA-type test approximation A_F with numerator and denominator degrees of freedom estimators (16) and (17), respectively, performs in general best. In contrast, the Lawley-Hotelling type tests perform better in the case of negatively correlated variables. An exception is the situation with large p and uncorrelated or negatively correlated variables in which the ANOVA-type test should be used with the degrees of freedom estimators based on (18) (Srivastava and Fujikoshi, 2006) (A_{FS}).

5 Example

A study was conducted in a commercial farm to evaluate the effects of three different fungicides (pesticides) on the control of fruit and foliar diseases of strawberry. A section of a 4-year-old strawberry planting was divided into 16 3-meter long single-row plots, and four treatments were randomly assigned to four plots each: sprayed with Kocide 2000 WG five times; sprayed with Elevate 50 WG plus Switch 62.5 WG four times; sprayed with V-10135 20 WP (experimental fungicide from Valent Corp.) three times; or not sprayed (control). All fruit were harvested and visually evaluated for symptoms of the fungus-caused disease grey mold (also known as Botrytis fruit rot), and symptoms of other fruit rots (caused by various fungal species). Total weight of all harvested fruit was determined ($X^{(1)}$). The percent of fruit with symptoms of Botrytis ($X^{(2)}$) and other species ($X^{(3)}$) was determined for each plot. Finally, the severity of symptoms on the foliage (leaflets) of Phomopsis leaf blight (another fungal-caused disease) was assessed with a 0-3 ordinal scale, where 0 represents disease-free and 3 represents 40% or more of the foliage covered by lesions. Thirty leaflets were measured in each plot, and the median value of these measurements was determined ($X^{(4)}$). The data are given in Table 6.

Thus, the design is a one-way layout with $a = 4$, $n = 4$, and $p = 4$. The test results based on using all the small-sample approximations are given in Table 7. All tests were significant (at the 5% level).

A simulation was done to determine the α-level and power for this design, using generally the same approach as described in Section 4. We also performed simulations for a parametric analysis of the non-ranked data. Some of the results are given here for the tests with the most desirable properties in the larger simulation study. For the approximations A_F, LH_{McK}, LH_{Fu}, BNP_{Mu}, and BNP_{Fu}, the simulated type I error rates were 3.4-5.8%, 4.7-5.6%, 5.2-6.2%, 4.5-5.3%, and 4.1-4.7%, respectively. The tests A_{as} and A_{FS} could not be recommended because of high type I error rates. Power curves were generated using two different models for the alternative. In model (A), all four variables change their means under alternative by a location shift, and this shift is different for the four groups. This is modeled by adding multiples of 0, 1, 2, or 3 to the means of control and the three treatment groups, respectively. In model (B), the four variables change differentially in the alternative, with one variable not changing at all, one variable changing more strongly, and the remaining two variables changing less strongly and into opposing directions. In such a scenario, the strong effect of one variable could be masked by the

Treatment	Replicate	Weight	Botrytis [%]	Other [%]	Phomopsis
	1	6.90	4.10	17.24	1.0
Kocide	2	8.30	5.13	5.65	1.0
(5 times)	3	8.40	6.07	8.80	1.5
	4	7.95	2.72	9.51	1.5
Elevate	1	8.60	1.19	17.06	1.0
plus	2	8.50	0.55	12.86	1.0
Switch	3	8.20	0.74	6.76	0.5
(4 times)	4	9.50	0.99	1.84	1.0
	1	6.20	4.29	4.64	1.0
V-10135	2	9.00	1.56	3.03	3.0
(3 times)	3	6.80	0.88	5.60	0.0
	4	8.50	2.42	8.66	2.0
not	1	7.50	15.60	13.08	1.0
sprayed	2	6.70	10.28	14.43	1.0
(control)	3	8.70	13.29	10.92	2.5
	4	7.40	18.38	16.03	3.0

Table 6: Original data of the study to evaluate the effects of three different fungicides (pesticides) on the control of fruit and foliar diseases of strawberry. The four treatments are: sprayed with Kocide 2000 WG five times; sprayed with Elevate 50 WG plus Switch 62.5 WG four times; sprayed with V-10135 20 WP three times; or not sprayed (control). The four response variables are: total fruit weight; percent of fruit with symptoms of Botrytis; percentage of fruit with symptoms of other species; severity of symptoms on the foliage of Phomopsis.

Test	Statistic	Num df	Den df	p-value
A_F	2.984	6.836	27.343	0.0191
A_{FS}	2.984	9.024	36.095	0.0092
LH_{McK}	8.241	12	12	0.0025
LH_{Fu}	12.116	N/A	N/A	0.0001
BNP_{Mu}	1.477	15.333	42.167	0.0060
BNP_{Fu}	2.319	N/A	N/A	0.0093

Table 7: Results from different multivariate nonparametric tests. The data is displayed in Table 6.

(A)	Fungicide			
	Control	1	2	3
$X^{(1)}$	0	+1	+2	+3
$X^{(2)}$	0	+1	+2	+3
$X^{(3)}$	0	+1	+2	+3
$X^{(4)}$	0	+1	+2	+3

(B)	Fungicide			
	Control	1	2	3
$X^{(1)}$	0	+2	+4	+6
$X^{(2)}$	0	1	2	3
$X^{(3)}$	0	0	0	0
$X^{(4)}$	0	-1	-2	-3

Table 8: Schematic for the simulated location shifts under alternative in power simulations, model (A) (left) and (B) (right).

presence of other variables with no or weak effect, so that the multivariate approach may suffer from lower power. The least favorable situation for a multivariate test is the extreme case in which only *one* variable displays an effect, and *all* other variables are unaffected. This situation will be addressed in a separate manuscript. See Table 8 for a schematic of the simulated location changes under the two alternative models. The simulated power curves for the tests A_F, LH_{McK}, and BNP_{Fu} are displayed in Figure 1. The tests LH_{Fu} and BNP_{Mu} are not displayed because, in general, they performed very similar to LH_{McK} and BNP_{Fu}, respectively.

The power simulations indicated that the proposed nonparametric tests A_F and LH_{McK} had in general the highest power, whereas BNP_{Fu} had the lowest power. This is consistent with the simulation results presented in the previous section. In model (B), the two tests A_F and LH_{McK} performed similarly. In model (A), the nonparametric ANOVA-type test had the highest power when the data was positively correlated, while the nonparametric Lawley-Hotelling test was in general most the most powerful test under a negative correlation structure. Also, the nonparametric tests performed generally better than their parametric counterparts in model (A). In model (B), the nonparametric tests had higher power in detecting alternatives that were closer to the null hypothesis, but for alternatives further away from the null hypothesis, the power curves of the parametric tests approached the ones of the corresponding nonparametric tests (LH_{McK} and BNP_{Fu}) or even crossed it (A_F). This could be due to the fact that in the nonparametric approach, the influence of any subset of the variables on the test statistic is bounded, whereas it is unbounded in the parametric multivariate tests. Note that in model (B), an effect exists only for three of the four variables, and a strong effect only for one of the variables.

6 Discussion

There are several small-sample finite approximations for the three nonparametric tests of marginal distributions based on ranks, ANOVA-type, Lawley-Hotelling type, and Bartlett-Nanda-Pillai type tests. The simulations here do not indicate a single best choice for all circumstances. This is not surprising, given that there is no single "best" test for parametric hypothesis testing in the multivariate one-way layout (*cf.* Anderson 1984, pp.330–333; Rencher 2002, pp.176–178). With an overall average negative correlation

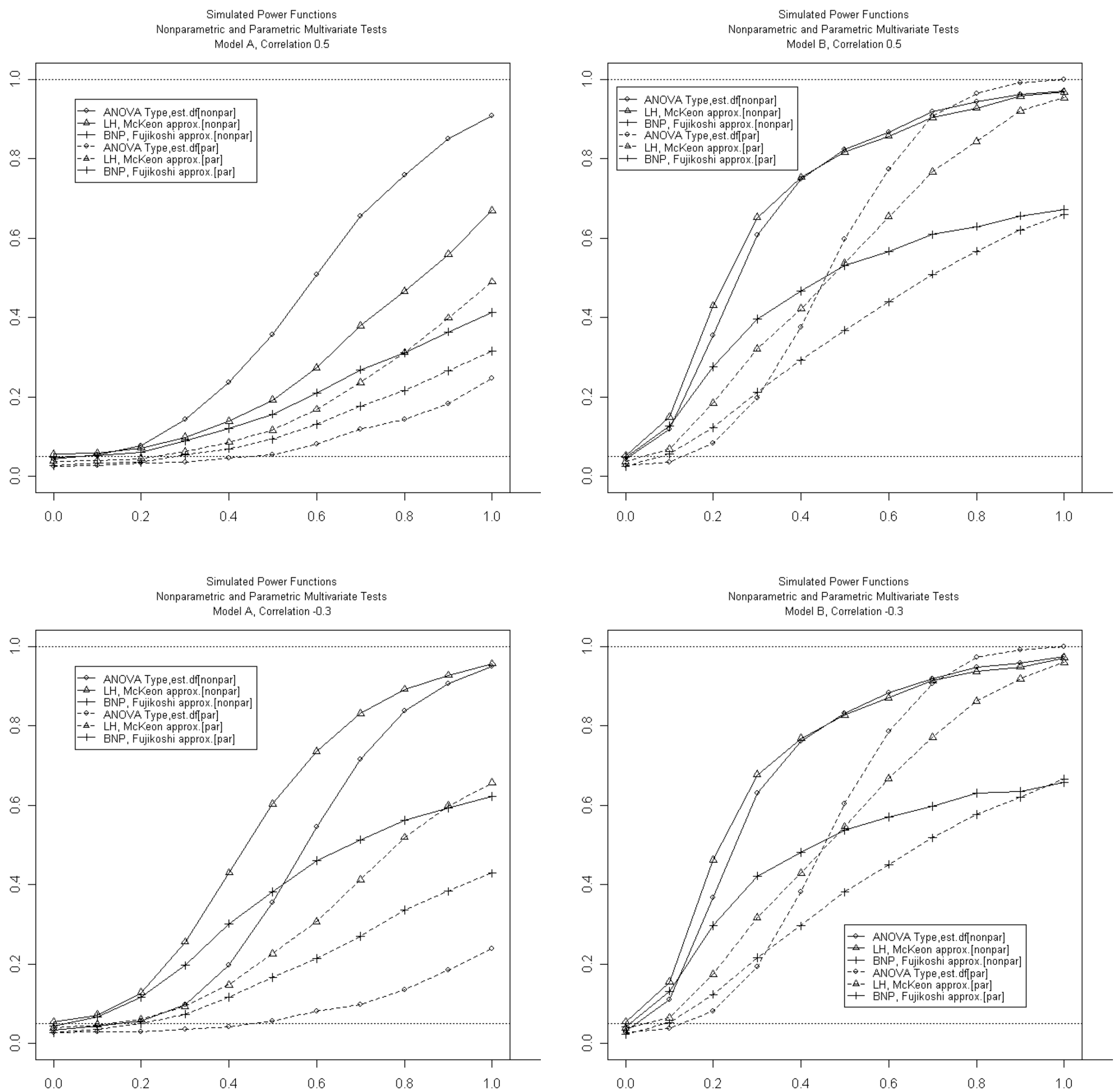

Figure 1: Simulated power of proposed nonparametric and of parametric multivariate tests in comparison: A_F, LH_{McK}, and BNP_{Fu}. Sample size $n = 4$ per level, $a = 4$ levels, $p = 4$ variables. Underlying distribution is multivariate normal with correlation=0.5 (above) or correlation=-0.3 (below). Alternative is modeled through location shift according to model (A) (left) or (B) (right).

of response variables, the finite approximations LH_{McK} and LH_{Fu} performed well in terms of α levels and power. Except when p was very high (16 here), the A_F approximation for the ANOVA-Type test performed fairly well with negative correlations. With a positive correlation, the A_F test performed best overall. Thus, it would be advantageous for data analysts to calculate correlations of the response variables before deciding on a test statistic to use. The performance of the Bartlett-Nanda-Pillai type tests, in terms of type I error rates, was generally little affected by the correlations of the responses. The small-sample approximations were just slightly conservative, although the BNP_{Fu} test became problematic at $p = 16$. The problem with this test was with power. With positive correlation, the Bartlett-Nanda-Pillai type tests had lower power than the A_F test, and with negative correlation, these tests had lower power than the Lawley-Hotelling type tests. We have written a program using PROC IML of SAS to calculate all test statistics and p-values, which is available from the senior author.

When there is a significant effect of group (treatment) on the marginal distributions, the next step often is to use contrasts to test individual hypotheses of interest, such as pairwise comparisons of treatments. Munzel and Brunner (2000a) show how to accomplish this using the ANOVA-type statistic in equation (5). All the global test statistics in this paper can be utilized for tests of contrasts.

References

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

Bai, Z., and Saranadasa, H. (1996), "Effect of High Dimension: By an Example of a Two-Sample Problem," *Statistica Sinica*, **6**, 311–329.

Bartlett, M.S. (1939), "A Note on Tests of Significance in Multivariate Analysis," *Proceedings of the Cambridge Philosophical Society*, **35**, 180–185.

Bathke, A.C., and Harrar, S.W. (2007), "Nonparametric Methods in Multivariate Factorial Designs for Large Number of Factor Levels," *Journal of Statistical Planning and Inference*, accepted for publication.

Bathke, A.C., Harrar, S.W., and Madden, L.V. (2007), "How to Compare Small Multivariate Samples Using Nonparametric Tests," Technical Report 399, Department of Statistics, University of Kentucky.

Box, G.E.P. (1954), "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification," *Annals of Mathematical Statistics*, **25**, 290–302.

Brunner, E., Dette, H., and Munk, A. (1997), "Box-Type Approximations in Nonparametric Factorial Designs," *Journal of the American Statistical Association*, **92**, 1494–1502.

Brunner, E., Munzel, U., and Puri, M.L. (1999), "Rank-Score Tests in Factorial Designs with Repeated Measures," *Journal of Multivariate Analysis*, **70**, 286–317.

Brunner, E., and Puri, M.L. (2001), "Nonparametric Methods in Factorial Designs," *Statistical Papers*, **42**, 1–52.

Chatfield, J.A., Draper, E.A., Cochran, K.D., and Herms, D.A. (2000), "Evaluation of Crabapples for Apple Scab at the Secrest Arboretum in Wooster, Ohio," *The Ohio State University, Ohio Agricultural Research and Development Center Special Circular 177. Ornamental Plants: Annual Reports and Research Reviews.*

Davis, A.W. (1970), “A System of Linear Differential Equations for the Distribution of Hotelling’s Generalized T_0^2,” *Annals of Statistics*, **39, 3**, 815–832.

Dempster, A. P. (1958), “A High Dimensional Two Sample Significance Test,” *Annals of Mathematical Statistics*, **29**, 995–1010.

Dempster, A. P. (1960), “A Significance Test for the Separation of Two Highly Multivariate Small Samples,” *Biometrics*, **16**, 41–50.

Fujikoshi, Y. (1973), “Monotonicity of the Power Functions of Some Tests in General MANOVA Models,” *Annals of Statistics*, **1**, 388–391.

Fujikoshi, Y. (1975), “Asymptotic Formulas for Non-Null Distributions of Three Statistics for Multivariate Linear Hypothesis,” *Annals of the Institute of Statistical Mathematics*, **27**, 99–108.

Harrar, S.W., and Bathke, A.C. (2007), “A Nonparametric Version of the Bartlett-Nanda-Pillai Multivariate Test: Asymptotics, Approximations, and Applications,” *American Journal of Mathematical and Management Sciences*, accepted for publication.

Harrar, S.W., and Bathke, A.C. (2006), “Nonparametric Methods for Unbalanced Multivariate Data and Many Factor Levels,” submitted.

Horst, L.E., Locke, J., Krause, C.R., McMahaon, R.W., Madden, L.V., and Hoitink, H.A.J. (2005), “Suppression of Botrytis blight of Begonia by Trichoderma hamatum 382 in peat and compost-amended potting mixes,” *Plant Disease*, **89**, 1195–1200.

Hotelling, H. (1947), “Multivariate Quality Control, Illustrated by the Air Testing of Sample Bombsights,” *In: Techniques of Statistical Analysis, C. Eisenhart, M. Hastay, and W.A. Wallis (Eds.)*, McGraw-Hill, New York, 111–184.

Hotelling, H. (1951), “A Generalized T Test and Measure of Multivariate Dispersion,” *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, J. Neyman (Ed.)*, University of California, Los Angeles and Berkeley, 23–41.

Ito, K. (1956), “Asymptotic Formulae for the Distribution of Hotelling’s Generalized T_0^2 Statistic,” *Annals of Mathematical Statistics*, **27**, 1091–1105.

Ito, K. (1960), “Asymptotic Formulae for the Distribution of Hotelling’s Generalized T_0^2 Statistic, II,” *Annals of Mathematical Statistics*, **31**, 1148–1153.

Kruskal, W.H. (1952), “A Nonparametric Test for the Several Sample Problem,” *Annals of Mathematical Statistics*, **23**, 525–540.

Lawley, D.N. (1938), “A Generalization of Fisher’s z Test,” *Biometrika*, **30**, 180–187.

Lévy, P. (1925), *Calcul des Probabilités*, Paris: Gauthiers-Villars.

Magnus, J.R., and Neudecker, H. (1979). “The Commutation Matrix: Some Properties and Applications,” *Annals of Statistics*, **7**, 381–394.

McKeon, J.J. (1974), “F Approximations to the Distribution of Hotelling’s T_0^2,” *Biometrika*, **61, 2**, 381–383.

Muirhead, R.J. (1970), *Aspects of Multivariate Statistical Theory*, New York: Wiley.

Muller, K.E. (1998), “A New F Approximation for the Pillai-Bartlett Trace Under H_0,” *Journal of Computational and Graphical Statistics*, **7, 1**, 131–137.

Munzel, U. (1999), “Linear Rank Score Statistics When Ties Are Present,” *Statistics and Probability Letters*, 41, 389–395.

Munzel, U., and Brunner, E. (2000a), "Nonparametric Methods in Multivariate Factorial Designs," *Journal of Statistical Planning and Inference*, **88, 1**, 117–132.

Munzel, U., and Brunner, E. (2000b), "Nonparametric Tests in the Unbalanced Multivariate One-Way Design," *Biometrical Journal*, **42, 7**, 837–854; Erratum (2001) **43, 6**, 791–792.

Nanda, D.N. (1950), "Distribution of the Sum of Roots of a Determinantal Equation Under a Certain Condition," *Annals of Mathematical Statistics*, **21**, 432–439.

Omer, M.A., Johnson D.A., and Rowe, R.C. (2000), "Recovery of Verticillium dahliae from North American Certified Seed Potatoes and Characterization of Strains by Vegetative Compatibility and Aggressiveness," *American Journal of Potato Research*, **77**, 325–331.

Pillai, K.C.S. (1955), "Some New Test Criteria in Multivariate Analysis," *Annals of Mathematical Statistics*, **26**, 117–121.

Puri, M.L., and Sen, P.K. (1966), "On A Class of Multivariate Multisample Rank-Order Tests," *Sankhya*, **A 28**, 353–376.

Rencher, A.C. (2002), *Methods of Multivariate Analysis* (Second Edition), New York: Wiley.

Rowe, R.C., Riedel, R.M., and Martin, M.J. (1985), "Synergistic interactions between Verticillium dahliae and Pratylenchus penetrans in potato early dying disease," *Phytopathology*, **75**, 412–418.

Ruymgaart, F.H. (1980), "A Unified Approach to the Asymptotic Distribution Theory of Certain Midrank Statistics," *In: Statistique non Parametrique Asymptotique, J.P. Raoult (Ed.)*, Lecture Notes on Mathematics, No. 821, Springer, Berlin, 1–18.

Satterthwaite, F.E. (1946), "An Approximate Distribution of Estimates of Variance Components," *Biometrics*, **20**, 110–114.

Sen, P.K. (1969), "Nonparametric Tests for Multivariate Interchangeability. Part Two: The Problem of MANOVA in Two-Way Layouts," *Sankhya*, **A 31**, 145–156.

Sen, P.K. (1970), "Asymptotic Distribution of a Class of Multivariate Rank Order Statistics," *Calcutta Statistical Association Bulletin*, **19, 73**, 23–31.

Srivastava, M.S., and Fujikoshi, Y. (2006), "Multivariate Analysis of Variance with Fewer Observations Than the Dimension," *Journal of Multivariate Analysis*, **97**, 1927–1940.

Thompson, G.L. (1990), "Asymptotic Distribution of Rank Statistics under Dependencies with Multivariate Application," *Journal of Multivariate Analysis*, **33**, 183–211.

Thompson, G.L. (1991), "A Unified Approach to Rank Tests for Multivariate and Repeated Measures Designs," *Journal of the American Statistical Association*, **86**, 410–419.

Welch, B.L. (1951), "On the Comparison of Several Mean Values," *Biometrika*, **38**, 330–336.