**Bootstrapping Analogs of the One Way MANOVA Test**

Hasthika S. Rupasinghe Arachchige Don and David J. Olive

Department of Mathematical Sciences, Appalachian State University, Boone, NC 28608-2092, hasthika@appstate.edu

Department of Mathematics, Southern Illinois University, Carbondale, Illinois 62901-4408, dolive@.siu.edu

**Keywords** Behrens–Fisher problem; bootstrap; confidence region; coordinatewise median.

**Mathematics Subject Classification** Primary 62H15; Secondary 62G09.

Abstract

Analogs of the classical one way MANOVA model have recently been suggested that do not assume that population covariance matrices are equal or that the error vector distribution is known. These tests are based on the sample mean and sample covariance matrix corresponding to each of the $p$ populations. We show how to extend these tests using other measures of location such as the trimmed mean or coordinatewise median. These new bootstrap tests can have some outlier resistance, and can perform better than the tests based on the sample mean if the error vector distribution is heavy tailed.

## 1. Introduction

Suppose there are $p$ independent random samples from $p$ groups or populations. Multivariate tests are often used to test whether the mean measurements are the same or differ across $p$ groups. The one way MANOVA test has null hypothesis $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_p$, and the test assumes that each group has the same population covariance matrix. Zhang and Liu (2013) and Konietschke, Bathke, Harrar, and Pauly (2015) developed analogs of the one way MANOVA test that do not assume that the population covariance matrices are equal or that the error vector distribution is known. These tests are based on the sample mean and sample covariance matrix $(\overline{\boldsymbol{y}}_i, \boldsymbol{S}_i)$ corresponding to the random sample from the

$i$th population. We show how to extend these tests using other measures of location such as the trimmed mean or coordinatewise median. Bootstrap confidence regions are used since estimating the asymptotic covariance matrix can be difficult. These new bootstrap tests can have some outlier resistance, and can perform better than the tests based on the sample mean if the error vector distribution is heavy tailed or skewed.

The analogs of the one way MANOVA model use independent random samples of size $n_i$ from $p$ different populations (treatments), or $n_i$ cases are randomly assigned to $p$ treatment groups where $n = \sum_{i=1}^{p} n_i$. Assume that $m$ response variables $\boldsymbol{y}_{ij} = (Y_{ij1}, ..., Y_{ijm})^T$ are measured for the $i$th treatment group and the $j$th case (often an individual or thing) in the group. Hence $i = 1, ..., p$ and $j = 1, ..., n_i$. Then

$$\boldsymbol{y}_{ij} = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_{ij}$$

where $\boldsymbol{\epsilon}_{i1}, ..., \boldsymbol{\epsilon}_{in_i}$ are independent and identically distributed (iid), and it is often assumed that $E(\boldsymbol{y}_{ij}) = \boldsymbol{\mu}_i$ and $\text{Cov}(\boldsymbol{y}_{ij}) = \text{Cov}(\boldsymbol{\epsilon}_{ij}) = \boldsymbol{\Sigma}_i$ for $i = 1, ..., p$.

The classical one way MANOVA model assumes that $\text{Cov}(\boldsymbol{\epsilon}_{ij}) = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ is the same for each of the $p$ populations or groups. This homogeneity assumption is very strong. Fujikoshi (2002) and Kakizawa (2009) derived the large sample theory, assuming that the error vectors are iid with unknown distribution, for the one way MANOVA tests based on the Pillai's trace statistic, Hotelling Lawley trace statistic, and Wilks' lambda. This theory is reviewed in Olive (2017a, ch. 10).

Large sample theory can be used to derive a test that does not need the equal population covariance matrix assumption $\boldsymbol{\Sigma}_i \equiv \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$. Let the statistic $T_i$ be a location estimator such as the sample mean or coordinatewise median. This test was used by Zhang and Liu (2013) and Konietschke, Bathke, Harrar, and Pauly (2015) with $T_i = \overline{\boldsymbol{y}}_i$ and $\hat{\boldsymbol{\Sigma}}_i = \boldsymbol{S}_i$.

The large sample theory for the test that uses $T_i$ for each group is much simpler than the large sample theory for the one way MANOVA test. To simplify the large sample theory, assume $n_i = \pi_i n$ where $0 < \pi_i < 1$ and $\sum_{i=1}^{p} \pi_i = 1$. Assume $H_0$ is true, and let $\boldsymbol{\mu}_i = \boldsymbol{\mu}$ for

$i = 1, ..., p$. Suppose $\sqrt{n_i}(T_i - \boldsymbol{\mu}) \xrightarrow{D} N_m(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$, and $\sqrt{n}(T_i - \boldsymbol{\mu}) \xrightarrow{D} N_m(\boldsymbol{0}, \boldsymbol{\Sigma}_i/\pi_i)$. Let

$$\boldsymbol{w} = \begin{bmatrix} T_1 - T_p \\ T_2 - T_p \\ \vdots \\ T_{p-2} - T_p \\ T_{p-1} - T_p \end{bmatrix}. \tag{1}$$

Then $\sqrt{n}\boldsymbol{w} \xrightarrow{D} N_{m(p-1)}(\boldsymbol{0}, \boldsymbol{\Sigma_w})$ with $\boldsymbol{\Sigma_w} = (\boldsymbol{\Sigma}_{ij})$ where $\boldsymbol{\Sigma}_{ij} =$
$\dfrac{\boldsymbol{\Sigma}_p}{\pi_p}$ for $i \neq j$, and $\boldsymbol{\Sigma}_{ii} = \dfrac{\boldsymbol{\Sigma}_i}{\pi_i} + \dfrac{\boldsymbol{\Sigma}_p}{\pi_p}$ for $i = j$. Hence the Wald-type statistic

$$t_0 = n\boldsymbol{w}^T \hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}}^{-1} \boldsymbol{w} = \boldsymbol{w}^T \left( \frac{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}}}{n} \right)^{-1} \boldsymbol{w} \xrightarrow{D} \chi^2_{m(p-1)}$$

as the $n_i \to \infty$ if $H_0$ is true. Here

$$\frac{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}}}{n} = \begin{bmatrix} \frac{\hat{\boldsymbol{\Sigma}}_1}{n_1} + \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} & \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} & \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} & \cdots & \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} \\ \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} & \frac{\hat{\boldsymbol{\Sigma}}_2}{n_2} + \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} & \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} & \cdots & \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} & \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} & \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} & \cdots & \frac{\hat{\boldsymbol{\Sigma}}_{p-1}}{n_{p-1}} + \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} \end{bmatrix}$$

is a block matrix where the off diagonal block entries equal $\hat{\boldsymbol{\Sigma}}_p/n_p$ and the $i$th diagonal block entry is $\dfrac{\hat{\boldsymbol{\Sigma}}_i}{n_i} + \dfrac{\hat{\boldsymbol{\Sigma}}_p}{n_p}$ for $i = 1, ..., (p-1)$.

Reject $H_0$ if

$$t_0 > m(p-1) F_{m(p-1), d_n}(1 - \delta) \tag{2}$$

where $d_n = \min(n_1, ..., n_p)$. It may make sense to relabel the groups so that $n_p$ is the largest $n_i$ or $\hat{\boldsymbol{\Sigma}}_p/n_p$ has the smallest generalized variance of the $\hat{\boldsymbol{\Sigma}}_i/n_i$. This test may start to outperform the one way MANOVA test if $n \geq (m+p)^2$ and $n_i \geq 40m$ for $i = 1, ..., p$.

If the sequence of positive integers $d_n \to \infty$ and $W_n \sim F_{r,d_n}$, then $rW_n \xrightarrow{D} \chi^2_r$. Using an $F_{r,d_n}$ cutoff instead of a $\chi^2_r$ cutoff is similar to using a $t_{d_n}$ cutoff instead of a standard normal $N(0,1)$ cutoff for inference. Instead of rejecting $H_0$ when $t_0 > \chi^2_{r,1-\delta}$, reject $H_0$ when

$$t_0 > rF_{r,d_n,1-\delta} = \frac{rF_{r,d_n,1-\delta}}{\chi^2_{r,1-\delta}} \chi^2_{r,1-\delta}.$$

The term $\dfrac{rF_{r,d_n,1-\delta}}{\chi^2_{r,1-\delta}}$ can be regarded as a small sample correction factor that improves the test's performance for small samples. Here $P(W_n \leq \chi^2_{r,\delta}) = \delta$ if $W_n$ has a $\chi^2_r$ distribution, and $P(W_n \leq F_{r,d_n,\delta}) = P(W_n \leq F_{r,d_n}(\delta)) = \delta$ if $W_n$ has an $F_{r,d_n}$ distribution.

If $\boldsymbol{T} = (T_1^T, T_2^T, ..., T_P^T)^T$, $\boldsymbol{\nu} = (\boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T, ..., \boldsymbol{\mu}_p^T)^T$, $\boldsymbol{c}$ is a constant vector, and $\boldsymbol{A}$ is a full rank $r \times mp$ matrix with rank $r$, then a large sample test of the form $H_0 : \boldsymbol{A\nu} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{A\nu} \neq \boldsymbol{\theta}_0$ uses

$$\boldsymbol{A}\sqrt{n}(\boldsymbol{T} - \boldsymbol{\nu}) \xrightarrow{D} \boldsymbol{u} \sim N_r\left(\boldsymbol{0}, \boldsymbol{A} \ diag\left(\frac{\boldsymbol{\Sigma}_1}{\pi_1}, \frac{\boldsymbol{\Sigma}_2}{\pi_2}, ..., \frac{\boldsymbol{\Sigma}_p}{\pi_p}\right) \ \boldsymbol{A}^T\right). \tag{3}$$

When $H_0$ is true, the Wald-type statistic

$$t_0 = [\boldsymbol{AT} - \boldsymbol{\theta}_0]^T \left[\boldsymbol{A} \ diag\left(\frac{\hat{\boldsymbol{\Sigma}}_1}{n_1}, \frac{\hat{\boldsymbol{\Sigma}}_2}{n_2}, ..., \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p}\right) \ \boldsymbol{A}^T\right]^{-1} [\boldsymbol{AT} - \boldsymbol{\theta}_0] \xrightarrow{D} \chi^2_r.$$

Section 2 shows how to get a bootstrap confidence region that can be used to test $H_0$ when $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}}$ or the $\hat{\boldsymbol{\Sigma}}_i$ are unknown or difficult to estimate. Section 3 gives some simulations and an example.

## 2. Bootstrapping Hypothesis Tests and the Prediction Region Method

Consider testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ where $\boldsymbol{\theta}_0$ is a known $r \times 1$ vector. Given training data $\boldsymbol{z}_1, ..., \boldsymbol{z}_n$, a large sample $100(1-\delta)\%$ confidence region for $\boldsymbol{\theta}$ is a set $\mathcal{A}_n$ such that $P(\boldsymbol{\theta} \in \mathcal{A}_n) \to 1 - \delta$ as $n \to \infty$. Then reject $H_0$ if $\boldsymbol{\theta}_0$ is not in the confidence region $\mathcal{A}_n$.

We will use the Olive(2017a, 2018) prediction region method to make a confidence region, and some notation is needed. Let $Z_n = \hat{\boldsymbol{\theta}}$ be an $r \times 1$ vector. Let $Z_1^*, ..., Z_B^*$ be the bootstrap sample. Let

$$\overline{Z}^* = \frac{1}{B}\sum_{i=1}^{B} Z_i^* \quad \text{and} \quad \boldsymbol{S}^*_{\mathrm{Z}} = \frac{1}{\mathrm{B}-1}\sum_{\mathrm{i=1}}^{\mathrm{B}}(\mathrm{Z}^*_{\mathrm{i}} - \overline{\mathrm{Z}}^*)(\mathrm{Z}^*_{\mathrm{i}} - \overline{\mathrm{Z}}^*)^{\mathrm{T}}$$

be the sample mean and sample covariance matrix of the bootstrap sample. Let the $i$th *squared sample Mahalanobis distance* be the scalar

$$D_i^2 = D_i^2(\overline{Z}^*, \boldsymbol{S}^*_Z) = D^2_{Z_i^*}(\overline{Z}^*, \boldsymbol{S}^*_Z) = (Z_i^* - \overline{Z}^*)^T[\boldsymbol{S}^*_Z]^{-1}(Z_i^* - \overline{Z}^*) \tag{4}$$

for each observation $Z_i^*$. Similarly,

$$D_i^2 = D^2_{Z_i}(W, \boldsymbol{C}) = (Z_i - W)^T\boldsymbol{C}^{-1}(Z_i - W).$$

Let $q_B = \min(1-\delta+0.05, 1-\delta+r/B)$ for $\delta > 0.1$ and

$$q_B = \min(1-\delta/2, 1-\delta+10\delta r/B), \quad \text{otherwise.} \tag{5}$$

If $1-\delta < 0.999$ and $q_B < 1-\delta+0.001$, set $q_B = 1-\delta$. Let $D_{(U_B)}$ be the $100q_B$th sample quantile of the $D_i$. This correction factor helps correct undercoverage when $B \geq 50r$ is finite.

Following Bickel and Ren (2001), let the vector of parameters $\boldsymbol{\theta} = Z(F)$, the statistic $Z_n = Z(F_n)$, and $Z^* = Z(F_n^*)$ where $F$ is the cdf of iid $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$, $F_n$ is the empirical cdf, and $F_n^*$ is the empirical cdf of $\boldsymbol{x}_1^*, ..., \boldsymbol{x}_n^*$, a sample from $F_n$ using the nonparametric bootstrap. If $\sqrt{n}(F_n - F) \xrightarrow{D} \boldsymbol{z}_F$, a Gaussian random process, and if $Z$ is sufficiently smooth (has a Hadamard derivative $\dot{Z}(F)$), then $\sqrt{n}(Z_n - \boldsymbol{\theta}) \xrightarrow{D} \boldsymbol{u}$ and $\sqrt{n}(Z_i^* - Z_n) \xrightarrow{D} \boldsymbol{u}$ with $\boldsymbol{u} = \dot{Z}(F)\boldsymbol{z}_F$. Olive (2017ab) used these results to show that if $\boldsymbol{u} \sim N_r(\boldsymbol{0}, \boldsymbol{\Sigma}_A)$, then $\sqrt{n}(\overline{Z}^* - Z_n) \xrightarrow{D} \boldsymbol{0}$, $\sqrt{n}(Z_i^* - \overline{Z}^*) \xrightarrow{D} \boldsymbol{u}$, $\sqrt{n}(\overline{Z}^* - \boldsymbol{\theta}) \xrightarrow{D} \boldsymbol{u}$, and that the prediction region method large sample $100(1-\delta)\%$ confidence region for $\boldsymbol{\theta}$ is $\{\boldsymbol{w} : (\boldsymbol{w} - \overline{Z}^*)^T [\boldsymbol{S}_Z^*]^{-1} (\boldsymbol{w} - \overline{Z}^*) \leq D^2_{(U_B)}\} =$

$$\{\boldsymbol{w} : D^2_{\boldsymbol{w}}(\overline{Z}^*, \boldsymbol{S}_Z^*) \leq D^2_{(U_B)}\} \tag{6}$$

where $D^2_{(U_B)}$ is computed from $D_i^2 = (Z_i^* - \overline{Z}^*)^T [\boldsymbol{S}_Z^*]^{-1} (Z_i^* - \overline{Z}^*)$ for $i = 1, ..., B$. Note that the corresponding test for $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ rejects $H_0$ if $(\overline{Z}^* - \boldsymbol{\theta}_0)^T [\boldsymbol{S}_Z^*]^{-1} (\overline{Z}^* - \boldsymbol{\theta}_0) > D^2_{(U_B)}$. Simpler proofs are in Pelawa Watagoda and Olive (2018). This procedure is basically the one sample Hotelling's $T^2$ test applied to the $Z_i^*$ using $\boldsymbol{S}_Z^*$ as the estimated covariance matrix and replacing the $\chi^2_{r,1-\delta}$ cutoff by $D^2_{(U_B)}$.

The modified Bickel and Ren (2001) large sample $100(1-\delta)\%$ confidence region for $\boldsymbol{\theta}$ is $\{\boldsymbol{w} : (\boldsymbol{w} - Z_n)^T [\boldsymbol{S}_Z^*]^{-1} (\boldsymbol{w} - Z_n) \leq D^2_{(U_B, Z)}\} =$

$$\{\boldsymbol{w} : D^2_{\boldsymbol{w}}(Z_n, \boldsymbol{S}_Z^*) \leq D^2_{(U_B, Z)}\} \tag{7}$$

where $D^2_{(U_B,Z)}$ is computed from $D_i^2 = (Z_i^* - Z_n)^T [\boldsymbol{S}_Z^*]^{-1} (Z_i^* - Z_n)$.

These two confidence regions are asymptotically equivalent if $\boldsymbol{u} \sim N_r(\boldsymbol{0}, \boldsymbol{\Sigma}_A)$, so that $\sqrt{n}(\overline{Z}^* - Z_n) \xrightarrow{D} \boldsymbol{0}$. Bickel and Ren (2001) showed that their method can work when Hadamard differentiability fails. The location model with means, medians, and trimmed means is one example where the Bickel and Ren (2001, p. 96) method works.

Next we will check some of the sufficient conditions for (6) to be a confidence region if $\sqrt{n}(T_i - \boldsymbol{\mu}_i) \xrightarrow{D} N_m(\mathbf{0}, \boldsymbol{\Sigma}_i/\pi_i)$ and if each coordinate $T_{ij}$ is Hadamard differentiable. Since the univariate sample mean, sample median, and sample trimmed mean are Hadamard differentiable and asymptotically normal, each coordinate satisfies $\sqrt{n}(T_{ij} - \overline{T}^*_{ij}) \xrightarrow{D} 0$ for $j = 1, ..., m$. Hence $\sqrt{n}(T_i - \overline{T}^*_i) \xrightarrow{D} \mathbf{0}$ for $i = 1, ..., p$, and $\sqrt{n}(\boldsymbol{T} - \overline{\boldsymbol{T}}^*) \xrightarrow{D} \mathbf{0}$ where $\overline{\boldsymbol{T}}^* = (\overline{T}^{*T}_1, ..., \overline{T}^{*T}_p)^T$. Thus $\sqrt{n}(Z_n - \overline{Z}^*) \xrightarrow{D} \mathbf{0}$ where $Z_n = \boldsymbol{AT}$. If $\boldsymbol{\theta} = \boldsymbol{A\nu}$, then $\sqrt{n}(Z_n - \boldsymbol{\theta}) \xrightarrow{D} \boldsymbol{u}$ and $\sqrt{n}(\overline{Z}^* - \boldsymbol{\theta}) \xrightarrow{D} \boldsymbol{u}$ where $\boldsymbol{u}$ is given by Equation (3). Hence

$$D^2_{\boldsymbol{\theta}} = (\overline{Z}^* - \boldsymbol{\theta})^T [\boldsymbol{S}^*_Z]^{-1} (\overline{Z}^* - \boldsymbol{\theta})^T = \sqrt{n}(\overline{Z}^* - \boldsymbol{\theta})^T [n\boldsymbol{S}^*_Z]^{-1} \sqrt{n}(\overline{Z}^* - \boldsymbol{\theta})^T \approx \boldsymbol{u}^T [n\boldsymbol{S}^*_Z]^{-1} \boldsymbol{u},$$

for probability calculations, and there exists a cutoff $\hat{D}^2_{1-\delta}$ such that (6) (using cutoff $\hat{D}^2_{1-\delta}$) is a large sample confidence region for $\boldsymbol{\theta}$ provided $n\boldsymbol{S}^*_Z$ does not get too ill conditioned. This is a much weaker condition than

$$n\boldsymbol{S}^*_Z \xrightarrow{P} \boldsymbol{\Sigma}_A = \boldsymbol{A} \ diag\left(\frac{\boldsymbol{\Sigma}_1}{\pi_1}, \frac{\boldsymbol{\Sigma}_2}{\pi_2}, ..., \frac{\boldsymbol{\Sigma}_p}{\pi_p}\right) \ \boldsymbol{A}^T.$$

See Machado and Parente (2005) for regularity conditions for $n\boldsymbol{S}^*_Z \xrightarrow{P} \boldsymbol{\Sigma}_A$. If $\sqrt{n}(T_i - \boldsymbol{\mu}_i) \xrightarrow{D} \boldsymbol{u}_i \sim N_m(\mathbf{0}, \boldsymbol{\Sigma}_i/\pi_i)$ and $\sqrt{n}(T^*_i - T_i) \xrightarrow{D} \boldsymbol{u}_i$ then (6) is a confidence region for $\boldsymbol{\theta}$. If also $n\boldsymbol{S}^*_Z \xrightarrow{P} \boldsymbol{\Sigma}_A$, then $D^2_{(U_B)} \xrightarrow{P} \chi^2_{r,1-\delta}$ under $H_0$. In the simulations with $H_0$ true and $n_i$ large, the confidence region coverage was near the nominal and the average value of $D^2_{(U_B)}$ tended to be near $\chi^2_{r,1-\delta}$. Hence (6) with $\hat{D}^2_{1-\delta} = D^2_{(U_B)}$ worked reasonably well.

Fréchet differentiability implies Hadamard differentiability, and many statistics are shown to be Hadamard differentiable in Bickel and Ren (2001), Clarke (1986, 2000), Fernholtz (1983), Gill (1989), Ren (1991), and Ren and Sen (1995).

To bootstrap the test $H_0 : \boldsymbol{A\nu} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{A\nu} \neq \boldsymbol{\theta}_0$, use $Z_n = \boldsymbol{AT}$. Take a sample of size $n_j$ with replacement from the $n_j$ cases for each group for $j = 1, 2, ..., p$ to obtain $T^*_j$ and $\boldsymbol{T}^*_1$. Repeat $B$ times to obtain $\boldsymbol{T}^*_1, ..., \boldsymbol{T}^*_B$. Then $Z^*_i = \boldsymbol{AT}^*_i$ for $i = 1, ..., B$. We will illustrate this method with the analog for the one way MANOVA test for $H_0 : \boldsymbol{A\theta} = \mathbf{0}$ which is equivalent to $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_p$, where $\mathbf{0}$ is an $r \times 1$ vector of zeroes with $r = m(p-1)$. Then $Z_n = \boldsymbol{AT} = \boldsymbol{w}$ given by Equation (1). Hence the $m(p-1) \times 1$ vector

$Z_i^* = \boldsymbol{A}\boldsymbol{T}_i^* = ((T_1^* - T_p^*)^T, ..., (T_{p-1}^* - T_p^*)^T)^T$ where $T_j$ is a robust location estimator, such as the coordinatewise median or trimmed mean, applied to the cases in the $j$th treatment group. The prediction region method fails to reject $H_0$ if $\mathbf{0}$ is in the resulting confidence region. Alternative methods for bootstrapping confidence regions are given in Ghosh and Polanski (2014).

## 3. Example and Simulations

In our simulations we used large sample sizes. We may need $B \geq 50m(p-1)$, $n \geq (m+p)^2$, and $n_i \geq 40m$. If the $n_i$ are not large, the one way MANOVA test can be regarded as a regularized estimator, and can perform better than the tests that do not assume equal population covariance matrices. Konietschke, Bathke, Harrar, and Pauly (2015) give some bootstrap methods that can simulate well for small $n_i$ if the sample means and covariance matrices $(\overline{\boldsymbol{y}}_i, \boldsymbol{S}_i)$ are used.

**Example.** The Cornwell and Trumbull (1994) North Carolina Crime data consists of 630 observations on 24 variables. This data set is available online from (https://vincentarelbundock.github.io/Rdatasets/datasets.html). Region is a categorical variable with three categories: Central, West, and Other with the number of observations 238, 147, and 245 respectively, and forms the three groups. The $m = 5$ variables are $Y_1 = wsta$ = weekly wage of state employees, $Y_2 = avgsen$ = average sentence days, $Y_3 = prbarr$ = 'probability' of arrest, $Y_5 = prbconv$ = 'probability' of conviction, and $Y_5 = taxpc$ = tax revenue per capita. There were a few outliers and boxplots of the variables, not shown, showed that the sample medians of the three groups were nearly the same for all 5 variables. The variables were highly skewed with different amounts of skew for the three groups. Hence the location measures other than the population coordinatewise median likely do differ. The test with the coordinatewise median had $D_0 = 4.086$ with the cutoff of $D_{(U_B)} = 4.32$ and failed to reject $H_0$. Note that $\sqrt{\chi^2_{10,0.95}} = 4.28$ and $\sqrt{10F_{10,147,0.95}} = 4.36$. The classical one way MANOVA test had a p-value of 0.001 and rejected the null hypothesis.

The simulation used 5000 runs with $B$ bootstrap samples and $p = 3$ groups. Olive (2017ab) suggests that the prediction region method can give good results when the number of bootstrap samples $B \geq 50r = 50m(p-1)$, and the simulation used various values of $B$. The sample mean, coordinatewise median, and coordinatewise 25% trimmed mean were the statistics $T$ used. The classical one way MANOVA Hotelling Lawley test statistic was also used.

Four types of data distributions $\boldsymbol{w}_i$ were considered that were identical for $i = 1, 2$, and 3. Then $\boldsymbol{y}_1 = \sigma_1 \boldsymbol{C}\boldsymbol{w}_1 + \delta_1 \boldsymbol{1}$, $\boldsymbol{y}_2 = \sigma_2 \boldsymbol{C}\boldsymbol{w}_2 + \delta_2 \boldsymbol{1}$, and $\boldsymbol{y}_3 = \sigma_3 \boldsymbol{C}\boldsymbol{w}_3 + \delta_3 \boldsymbol{1}$ or $\boldsymbol{y}_3 = \boldsymbol{w}_3$ where $\boldsymbol{1} = (1, .., 1)^T$ is a vector of ones and $\boldsymbol{C} = diag(1, \sqrt{2}, ..., \sqrt{m})$. The $\boldsymbol{w}_i$ distributions were the multivariate normal distribution $N_m(\boldsymbol{0}, \boldsymbol{I})$, the mixture distribution $0.6N_m(\boldsymbol{0}, \boldsymbol{I}) + 0.4N_m(\boldsymbol{0}, 25\boldsymbol{I})$, the multivariate $t$ distribution with 4 degrees of freedom, and the multivariate lognormal distribution shifted to have nonzero mean $\boldsymbol{\mu} = 0.649\ \boldsymbol{1}$, but a population coordiatewise median of $\boldsymbol{0}$. If $\sigma_1 = 1$ and $\delta_i = 0$ for $i = 1, 2, 3$, note that $\text{Cov}(\boldsymbol{y}_2) = \sigma_2^2\ \text{Cov}(\boldsymbol{y}_1)$, and for the first three distributions, $E(\boldsymbol{y}_i) = E(\boldsymbol{w}_i) = \boldsymbol{0}$. If $\boldsymbol{y}_3 = \boldsymbol{w}_3$ then $\text{Cov}(\boldsymbol{y}_3) = c\boldsymbol{I}_m$ for some constant $c > 0$. If $\sigma_1 = 1$ and $\boldsymbol{y}_3 = \sigma_3 \boldsymbol{C}\boldsymbol{w}_3 + \delta_3 \boldsymbol{1}$, then $\text{Cov}(\boldsymbol{y}_3) = \sigma_3^2\ \text{Cov}(\boldsymbol{y}_1)$.

Adding the same type and proportion of outliers to all three groups often resulted in three distributions that were still similar. Hence outliers were added to the first group but not the second or third, making the covariance structures of the three groups quite different. The outlier proportion was $100\gamma\%$. Let $\boldsymbol{y}_1 = (y_{11}, ..., y_{1m})^T$. The five outlier types for group 1 were type 1: a tight cluster at the major axis $(0, ..., 0, z)^T$, type 2: a tight cluster at the minor axis $(z, 0, ..., 0)^T$, type 3: $N_m(z\boldsymbol{1}, diag(1, ..., m))$, type 4: $y_{1m}$ replaced by $z$, and type 5: $y_{11}$ replaced by $z$. The quantity $z$ determines how far the outliers are from the clean data.

Let the *coverage* be the proportion of times that $H_0$ is rejected. We want the *coverage* near 0.05 when $H_0$ is true and the coverage close to 1.0 for good power when $H_0$ is false. With 5000 runs, an observed *coverage* inside of (0.04, 0.06) suggests that the true *coverage* is close to the nominal 0.05 coverage when $H_0$ is true.

The new tests worked well with all the distributions and with the different covariance settings. Tables 1 through 4 show simulation results for two distributions with various covariance settings. We took $\delta_1 = \delta_3 = 0$ and $B =$ the size of the bootstrap sample. Balanced and unbalanced designs have also been considered. For Tables 1 and 2, $\boldsymbol{\Sigma}_i \propto diag(1, 2, .., m)$

for $i = 1, 2, 3$. For Tables 3 and 4, $\sigma_2 = \sigma_3 = 1$, and $\boldsymbol{\Sigma}_3 = c\boldsymbol{I}$ does not have the same shape as $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$. Tables 1 and 3 are for the multivariate normal (MVN) distribution. The classical test works well with multivariate normal data when the covariance matrices are the same, but the type I error tends to be higher than the nominal level when the covariance matrices differ. The classical test can be too conservative when the design is unbalanced. Having an unbalanced design and different covariance matrices was the worst case scenario for the classical test regardless of the data distribution. The bootstrap tests using the mean and coordinatewise trimmed mean usually performed well but occasionally had coverage near 0.07. Tables 2 and 4 are for the lognormal distribution, where the location measures other than the coordinatewise median differ if $\sigma_2 \neq \sigma_3$ (then coverage near 1 is desired).

Figures 1, 2, and 3 generated power curves for the bootstrap tests and for the Zhang and Liu (2013) MANOVA type test (2) based on the sample means $\overline{\boldsymbol{y}}_i$ and $\boldsymbol{S}_i$ for the 3 groups. The bootstrap test based on the sample means bootstraps the test (2). For these power curves, group $i$ has mean $\boldsymbol{\mu}_i = \delta_i \mathbf{1}$ where $\delta_2 = 2\ \delta_1$ and $\delta_3 = 3\ \delta_1$. When $\delta_1$ increases, the distance between the mean vectors increases. The power curves for the bootstrap test based on the sample means and for test (2) were always similar. Figure 1 shows the power curve for clean MVN data with a balanced design where the groups have the same covariance matrices. Here the three mean based tests had similar power. The power curve for the classical test was poor for the next two figures. Figure 2 shows clean MVN data with $m = 5, \sigma_1 = 1, \sigma_2 = 2, \sigma_3 = 5, n_1 = 200, n_2 = 400$, and $n_3 = 600$. Figure 3 used settings similar to Figure 2 with the multivariate $t_4$ distribution, and the coordinatewise trimmed mean had the best power.

Simulations were also done for type I error with contamination using the five types of outliers, and $(\gamma, z) = (0.1, 10)$ or $(0.05,\ 20)$. In Table 5 with $m = 5$, the test with the coordinatewise median works reasonably well (close to the nominal coverage) for 10% outliers with all the distributions and for all the outlier types with the exception of outlier type 3. All the other tests, including the classical test, failed. Results were similar with $m = 10$, $n_i = 800$, $B = 1000$, and $\gamma = 0.05$. Increasing $z$ as $m$ increases can help, but if $m$ and $\gamma$ are

Table 1: Type I error for clean MVN data with $\boldsymbol{\Sigma}_3 \neq c\boldsymbol{I}$

| $m$ | $n_1$ | $n_2$ | $n_3$ | $B$ | $\sigma_2$ | $\sigma_3$ | Median | Mean | Tr.Mn | Class |
|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 200 | 200 | 200 | 400 | 1 | 1 | 0.0422 | 0.0562 | 0.0552 | 0.0460 |
| | | | | 1000 | 1 | 1 | 0.0486 | 0.0602 | 0.0598 | 0.0510 |
| | | | | 400 | 2 | 3 | 0.0506 | 0.0670 | 0.0606 | 0.0680 |
| | | | | 1000 | 2 | 3 | 0.0482 | 0.0580 | 0.0590 | 0.0680 |
| 5 | 200 | 400 | 600 | 400 | 1 | 1 | 0.0506 | 0.0542 | 0.0598 | 0.0474 |
| | | | | 1000 | 1 | 1 | 0.0492 | 0.0542 | 0.0554 | 0.0472 |
| | | | | 400 | 2 | 3 | 0.0474 | 0.0580 | 0.0576 | 0.0066 |
| | | | | 1000 | 2 | 3 | 0.0532 | 0.0626 | 0.0618 | 0.0074 |
| 10 | 400 | 400 | 400 | 800 | 1 | 1 | 0.0508 | 0.0724 | 0.0712 | 0.0558 |
| | | | | 2000 | 1 | 1 | 0.0516 | 0.0652 | 0.0644 | 0.0526 |
| | | | | 800 | 2 | 3 | 0.0562 | 0.0640 | 0.0686 | 0.0656 |
| | | | | 2000 | 2 | 3 | 0.0554 | 0.0624 | 0.0630 | 0.0704 |
| 10 | 400 | 800 | 1200 | 800 | 1 | 1 | 0.0510 | 0.0594 | 0.0626 | 0.0456 |
| | | | | 2000 | 1 | 1 | 0.0470 | 0.0578 | 0.0576 | 0.0494 |
| | | | | 800 | 2 | 3 | 0.0468 | 0.0576 | 0.0572 | 0.0008 |
| | | | | 2000 | 2 | 3 | 0.0440 | 0.0574 | 0.0534 | 0.0034 |
| 20 | 800 | 800 | 800 | 1600 | 1 | 1 | 0.0474 | 0.0724 | 0.0652 | 0.0496 |
| | | | | 4000 | 1 | 1 | 0.0504 | 0.0662 | 0.0668 | 0.0494 |
| | | | | 1600 | 2 | 3 | 0.0566 | 0.0728 | 0.0618 | 0.0772 |
| | | | | 4000 | 2 | 3 | 0.0592 | 0.0644 | 0.0672 | 0.0638 |
| 20 | 800 | 1600 | 2400 | 1600 | 1 | 1 | 0.0562 | 0.0644 | 0.0648 | 0.0492 |
| | | | | 4000 | 1 | 1 | 0.0504 | 0.0564 | 0.0618 | 0.0462 |
| | | | | 1600 | 2 | 3 | 0.0530 | 0.0654 | 0.0650 | 0.0000 |
| | | | | 4000 | 2 | 3 | 0.0472 | 0.0632 | 0.0620 | 0.0008 |

Table 2: Type I error for clean lognormal data with $\boldsymbol{\Sigma}_3 \neq c\boldsymbol{I}$

| $m$ | $n_1$ | $n_2$ | $n_3$ | $B$ | $\sigma_2$ | $\sigma_3$ | Median | Mean | Tr.Mn | Class |
|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 200 | 200 | 200 | 400 | 1 | 1 | 0.0368 | 0.0628 | 0.0478 | 0.0436 |
| | | | | 1000 | 1 | 1 | 0.0402 | 0.0596 | 0.0486 | 0.0452 |
| | | | | 400 | 2 | 3 | 0.0432 | 0.9996 | 0.1004 | 0.9994 |
| | | | | 1000 | 2 | 3 | 0.0448 | 1.0000 | 0.0980 | 0.9996 |
| 5 | 200 | 400 | 600 | 400 | 1 | 1 | 0.0446 | 0.0768 | 0.0568 | 0.0476 |
| | | | | 1000 | 1 | 1 | 0.0426 | 0.0724 | 0.0530 | 0.0530 |
| | | | | 400 | 2 | 3 | 0.0428 | 1.0000 | 0.2068 | 1.0000 |
| | | | | 1000 | 2 | 3 | 0.0428 | 1.0000 | 0.2002 | 1.0000 |
| 10 | 400 | 400 | 400 | 800 | 1 | 1 | 0.0450 | 0.0658 | 0.0622 | 0.0450 |
| | | | | 2000 | 1 | 1 | 0.0472 | 0.0716 | 0.0542 | 0.0498 |
| | | | | 800 | 2 | 3 | 0.0532 | 1.0000 | 0.2858 | 1.0000 |
| | | | | 2000 | 2 | 3 | 0.0458 | 1.0000 | 0.2706 | 1.0000 |
| 10 | 400 | 800 | 1200 | 800 | 1 | 1 | 0.0434 | 0.0754 | 0.0542 | 0.0546 |
| | | | | 2000 | 1 | 1 | 0.0502 | 0.0708 | 0.0526 | 0.0462 |
| | | | | 800 | 2 | 3 | 0.0438 | 1.0000 | 0.6448 | 1.0000 |
| | | | | 2000 | 2 | 3 | 0.0372 | 1.0000 | 0.6394 | 1.0000 |
| 20 | 800 | 800 | 800 | 1600 | 1 | 1 | 0.0482 | 0.0680 | 0.0580 | 0.0470 |
| | | | | 4000 | 1 | 1 | 0.0412 | 0.0678 | 0.0582 | 0.0486 |
| | | | | 1600 | 2 | 3 | 0.0530 | 1.0000 | 0.8714 | 1.0000 |
| | | | | 4000 | 2 | 3 | 0.0516 | 1.0000 | 0.8622 | 1.0000 |
| 20 | 800 | 1600 | 2400 | 1600 | 1 | 1 | 0.0470 | 0.0756 | 0.0648 | 0.0532 |
| | | | | 4000 | 1 | 1 | 0.0520 | 0.0684 | 0.0652 | 0.0464 |
| | | | | 1600 | 2 | 3 | 0.0480 | 1.0000 | 0.9980 | 1.0000 |
| | | | | 4000 | 2 | 3 | 0.0442 | 1.0000 | 0.9988 | 1.0000 |

Table 3: Type I error for clean MVN data with $\boldsymbol{\Sigma}_3 = c\boldsymbol{I}$

| $m$ | $n_1$ | $n_2$ | $n_3$ | $B$ | Median | Mean | Tr.Mn | Class |
|---|---|---|---|---|---|---|---|---|
| 5 | 200 | 200 | 200 | 400 | 0.0482 | 0.0682 | 0.0638 | 0.0650 |
| | | | | 1000 | 0.0500 | 0.0684 | 0.0610 | 0.0592 |
| 5 | 200 | 400 | 600 | 400 | 0.0566 | 0.0604 | 0.0648 | 0.1354 |
| | | | | 1000 | 0.0472 | 0.0526 | 0.0534 | 0.1278 |
| 10 | 400 | 400 | 400 | 800 | 0.0512 | 0.0636 | 0.0610 | 0.0604 |
| | | | | 2000 | 0.0506 | 0.0608 | 0.0632 | 0.0584 |
| 10 | 400 | 800 | 1200 | 800 | 0.0570 | 0.0658 | 0.0642 | 0.2422 |
| | | | | 2000 | 0.0536 | 0.0536 | 0.0536 | 0.2224 |
| 20 | 800 | 800 | 800 | 1600 | 0.0662 | 0.0740 | 0.0734 | 0.0638 |
| | | | | 4000 | 0.0562 | 0.0668 | 0.0600 | 0.0604 |
| 20 | 800 | 1600 | 2400 | 1600 | 0.0566 | 0.0638 | 0.0628 | 0.4308 |
| | | | | 4000 | 0.0560 | 0.0702 | 0.0658 | 0.4308 |

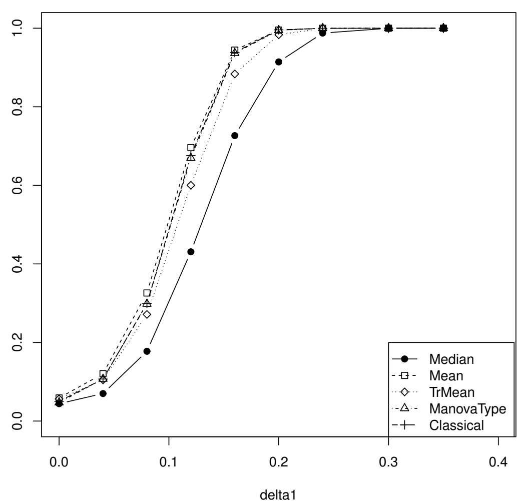


Figure 1: Power curve for clean MVN data with $m = 5, \sigma_1 = 1, \sigma_2 = 1, \sigma_3 = 1, n_1 = 200, n_2 = 200$, and $n_3 = 200$.

Table 4: Type I error for clean lognormal data with $\boldsymbol{\Sigma}_3 = c\boldsymbol{I}$

| $m$ | $n_1$ | $n_2$ | $n_3$ | $B$ | Median | Mean | Tr.Mn | Class |
|---|---|---|---|---|---|---|---|---|
| 5 | 200 | 200 | 200 | 400 | 0.0424 | 0.8744 | 0.0652 | 0.7208 |
| | | | | 1000 | 0.0446 | 0.8790 | 0.0686 | 0.7220 |
| 5 | 200 | 400 | 600 | 400 | 0.0470 | 0.9950 | 0.0864 | 0.9980 |
| | | | | 1000 | 0.0460 | 0.9976 | 0.0884 | 0.9990 |
| 10 | 400 | 400 | 400 | 800 | 0.0440 | 1.0000 | 0.2404 | 1.0000 |
| | | | | 2000 | 0.0438 | 1.0000 | 0.2424 | 1.0000 |
| 10 | 400 | 800 | 1200 | 800 | 0.0524 | 1.0000 | 0.4256 | 1.0000 |
| | | | | 2000 | 0.0520 | 1.0000 | 0.4384 | 1.0000 |
| 20 | 800 | 800 | 800 | 1600 | 0.0576 | 1.0000 | 0.9674 | 1.0000 |
| | | | | 4000 | 0.0602 | 1.0000 | 0.9668 | 1.0000 |
| 20 | 800 | 1600 | 2400 | 1600 | 0.0588 | 1.0000 | 0.9994 | 1.0000 |
| | | | | 4000 | 0.0504 | 1.0000 | 0.9996 | 1.0000 |

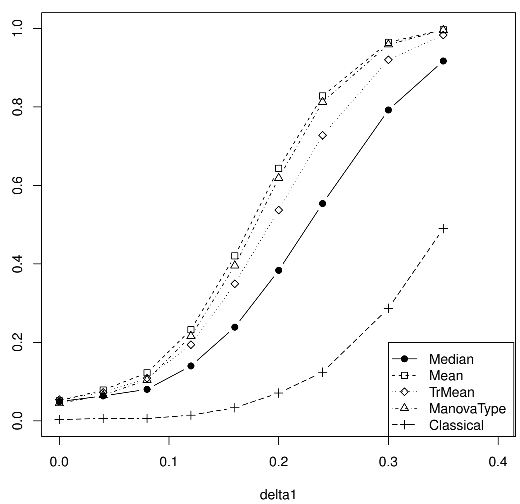


Figure 2: Power curve for clean MVN data with $m = 5, \sigma_1 = 1, \sigma_2 = 2, \sigma_3 = 5, n_1 = 200, n_2 = 400$, and $n_3 = 600$.

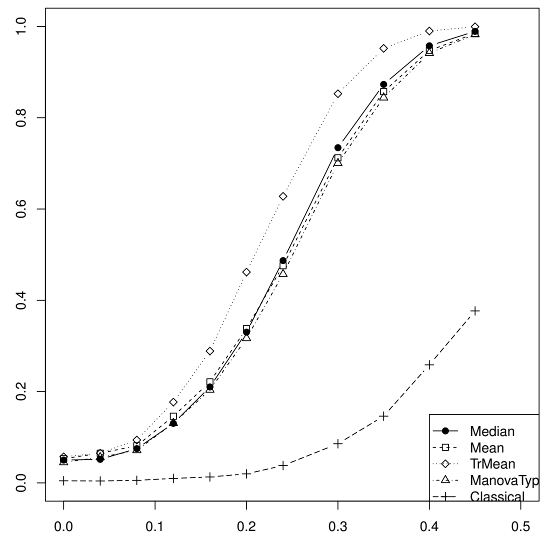


Figure 3: Power curve for clean multivariate $t_4$ data with $m = 5, \sigma_1 = 1, \sigma_2 = 2, \sigma_3 = 5, n_1 = 200, n_2 = 400,$ and $n_3 = 600$.

large enough, then the outliers move the coordinatewise median of the first group enough so that the test tends to reject $H_0$.

**4. Conclusions**

Bootstrapping different estimators of multivariate location provides an alternative to the one way MANOVA test that assumes the population covariance matrices of the $p$ groups are the same. The bootstrap test and test (2) were similar when the sample means $\overline{\boldsymbol{y}}_i$ were used. A larger simulation is in Rupasinghe Arachchige Don (2017). Rupasinghe Arachchige Don and Pelawa Watagoda (2017) consider bootstrapping analogs of the two sample Hotelling's $T^2$ test, and Konietschke, Bathke, Harrar, and Pauly (2015) suggest a method for bootstrapping analogs of the one way MANOVA model that may be useful even if the $n_i$ are not all large. References for robust one way MANOVA tests are in Finch and French (2013), Todorov and Filzmoser (2010), Van Aelst and Willems (2011), Wilcox (1995), and Zhang and Liu (2013).

The *R* software was used in the simulation. See R Core Team (2016). Programs are in the Olive (2017a) collection of *R* functions *mpack.txt* available from (http://lagrange.math.siu.edu /Olive/mpack.txt). The function `manbtsim2` was used to simulate the tests of hypotheses,

Table 5: Type I error with contaminated data: $m = 5$, $\gamma = 0.1$

| Dist. | $n_1 = n_2 = n_3$ | $B$ | outlier | Median | Mean | Tr.Mn | Class |
|---|---|---|---|---|---|---|---|
| 1 | 200 | 1000 | 1 | 0.0638 | 0.8034 | 0.1572 | 0.9302 |
| | | | 2 | 0.0504 | 0.9826 | 0.1488 | 1.0000 |
| | | | 3 | 0.2720 | 0.9994 | 0.4024 | 1.0000 |
| | | | 4 | 0.0966 | 0.7862 | 0.1546 | 0.9236 |
| | | | 5 | 0.0840 | 0.9854 | 0.1268 | 1.0000 |
| 2 | 200 | 1000 | 1 | 0.0488 | 0.1832 | 0.1068 | 0.1812 |
| | | | 2 | 0.0376 | 0.4880 | 0.1042 | 0.5428 |
| | | | 3 | 0.1994 | 0.7502 | 0.2206 | 0.8978 |
| | | | 4 | 0.0858 | 0.1848 | 0.1080 | 0.1830 |
| | | | 5 | 0.0780 | 0.4688 | 0.0974 | 0.5400 |
| 3 | 200 | 1000 | 1 | 0.0598 | 0.6046 | 0.1554 | 0.7094 |
| | | | 2 | 0.0464 | 0.9356 | 0.1366 | 0.9946 |
| | | | 3 | 0.2590 | 0.9944 | 0.3882 | 1.0000 |
| | | | 4 | 0.0946 | 0.5824 | 0.1486 | 0.6926 |
| | | | 5 | 0.0828 | 0.9216 | 0.1270 | 0.9928 |
| 4 | 200 | 1000 | 1 | 0.0426 | 0.9880 | 0.1998 | 0.9624 |
| | | | 2 | 0.0416 | 0.9924 | 0.1396 | 0.9884 |
| | | | 3 | 0.1762 | 1.0000 | 0.3980 | 1.0000 |
| | | | 4 | 0.0708 | 0.9902 | 0.1892 | 0.9674 |
| | | | 5 | 0.0766 | 0.9950 | 0.1508 | 0.9932 |

and `predreg` computes the confidence region given the bootstrap values. The function `manbtsim4` adds the test given by Equation (2) using the $(\overline{\boldsymbol{y}}_i, \boldsymbol{S}_i)$, which is very similar to the bootstrap test with the sample means.

**Acknowledgments**

**References**

Bickel, P. J., Ren, J. –J. (2001). The bootstrap in hypothesis testing. In: de Gunst, M., Klaassen, C., van der Vaart, A. Eds. *State of the Art in Probability and Statistics: Festschrift for William R. van Zwet*. The Institute of Mathematical Statistics, CA: Hayward, pp. 91-112.

Clarke, B. R. (1986). Nonsmooth analysis and Fréchet differentiability of $M$ functionals. *Probability Theory and Related Fields* 73:137-209.

Clarke, B. R. (2000). A review of differentiability in relation to robustness with an application to seismic data analysis. *Proceedings of the Indian National Science Academy, A* 66:467-482.

Cornwell, C., Trumbull, W. N. (1994). Estimating the economic model of crime with panel data. *Review of Economics and Statistics* 76:360-366.

Fernholtz, L. T. (1983). *von Mises Calculus for Statistical Functionals*. New York: Springer.

Finch, H., French, B. (2013). A "Monte Carlo" comparison of robust MANOVA test statistics. *Journal of Modern Applied Statistical Methods* 12:35-81.

Fujikoshi, Y. (2002). Asymptotic expansions for the distributions of multivariate basic statistics and one-way MANOVA tests under nonnormality. *Journal of Statistical Planning and Inference* 108:263-282.

Ghosh, S., Polansky, A. M. (2014). Smoothed and iterated bootstrap confidence regions for parameter vectors. *Journal of Multivariate Analysis* 132:171-182.

Gill, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von

Mises method, part 1. *Scandinavian Journal of Statistics* 16:97-128.

Kakizawa, Y. (2009). Third-order power comparisons for a class of tests for multivariate linear hypothesis under general distributions. *Journal of Multivariate Analysis* 100:473-496.

Konietschke, F., Bathke, A. C., Harrar, S. W., Pauly, M. (2015). Parametric and nonparametric bootstrap methods for general MANOVA. *Journal of Multivariate Analysis* 140:291-301.

Machado, J. A. F., Parente, P. (2005). Bootstrap estimation of covariance matrices via the percentile method. *Econometrics Journal* 8:70-78.

Olive, D. J. (2017a). *Robust Multivariate Analysis.* New York: Springer.

Olive, D. J. (2017b). Bootstrapping hypothesis tests and confidence regions. Unpublished Manuscript with the bootstrap material from Olive (2017a) at (http://lagrange.math.siu.edu /Olive/ppvselboot.pdf).

Olive, D. J. (2018). Applications of hyperellipsoidal prediction regions. *Statistical Papers.* To appear.

Pelawa Watagoda, L. C. R., Olive, D. J. (2018). Bootstrapping multiple linear regression after variable selection. Unpublished manuscript at (http://lagrange.math.siu.edu/Olive/ppboottest.pdf).

R Core Team (2016). R: *A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing. (www.R-project.org).

Ren, J.-J. (1991). On Hadamard differentiability of extended statistical functional. *Journal of Multivariate Analysis* 39:30-43.

Ren, J.-J., Sen, P. K. (1995). Hadamard differentiability on D[0,1]$^p$. *Journal of Multivariate Analysis* 55:14-28.

Rupasinghe Arachchige Don, H. S. (2017). Bootstrapping Analogs of the One Way MANOVA

Test. (Ph.D. Thesis), Southern Illinois University, USA, at (http://lagrange.math.siu.edu/Olive/shasthikaphd.pdf).

Rupasinghe Arachchige Don, H. S., Pelawa Watagoda, L. C. R. (2018). Bootstrapping analogs of the two sample Hotelling's $T^2$ test. *Communications and Statistics: Theory and Methods* to appear.

Todorov, V., Filzmoser, P. (2010). Robust statistics for the one-way MAVOVA. *Computational Statistics & Data Analysis* 54:37-48.

Van Aelst, S., Willems, G. (2011). Robust and efficient one-way MANOVA tests. *Journal of the American Statistical Association* 106:706-718.

Wilcox, R. R. (1995). Simulation results on solutions to the multivariate "Behrens-Fisher" problem via trimmed means. *The Statistician* 44:213-225.

Zhang, J.–T., Liu, X. (2013). A modified Bartlett test for heteroscedastic one-way MANOVA. *Metrika* 76:135-152.