BOOTSTRAPPING ANALOGS OF THE ONE WAY MANOVA TEST

by

Hasthika S. Rupasinghe Arachchige Don

M.S., Southern Illinois University, 2013

A Dissertation
Submitted in Partial Fulfillment of the Requirements for the
Doctoral Degree

Department of Mathematics
in the Graduate School
Southern Illinois University Carbondale
July, 2017

AN ABSTRACT OF THE DISSERTATION OF

HASTHIKA S. RUPASINGHE ARACHCHIGE DON, for the Doctor of Philosophy degree in MATHEMATICS, presented on DATE OF DEFENSE, at Southern Illinois University Carbondale.

TITLE: BOOTSTRAPPING ANALOGS OF THE ONE WAY MANOVA TEST

MAJOR PROFESSOR: Dr. D. J. Olive

The classical one way MANOVA model is used to test whether the mean measurements are the same or differ across p groups, and assumes that the covariance matrix of each group is the same. This work suggests using the Olive (2017abc) bootstrap technique to develop analogs of one way MANOVA test. A large sample theory test has also been developed. The bootstrap tests can have considerable outlier resistance, and the tests do not need the population covariance matrices to be equal. The two sample Hotelling's T^2 test is the special case of the one way MANOVA model when $p = 2$.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 MANOVA

Multivariate analysis of variance (MANOVA) is analogous to an ANOVA with more than one dependent variable. ANOVA tests for the difference in means between two or more groups, while MANOVA tests for the difference in two or more vectors of means.

Definition. The multivariate analysis of variance (MANOVA) model $\boldsymbol{y}_i = \boldsymbol{B}^T\boldsymbol{x}_i + \boldsymbol{\epsilon}_i$ for $i = 1, \ldots, n$ has $m \geq 2$ response variables $Y_1, \ldots Y_m$ and p predictor variables $x_1, x_2, \ldots, x_p$. The ith case is $(\boldsymbol{x}_i^T, \boldsymbol{y}_i^T) = (x_{i1}, ..., x_{ip}, Y_{i1}, ..., Y_{im})$. If a constant $x_{i1} = 1$ is in the model, then x_{i1} could be omitted from the case.

For the MANOVA model predictors are indicator variables. Sometimes the trivial predictor $\mathbf{1}$ is also in the model.

The MANOVA model in matrix form is $\boldsymbol{Z} = \boldsymbol{XB} + \boldsymbol{E}$ and has $E(\boldsymbol{\epsilon}_k) = \mathbf{0}$ and $Cov(\boldsymbol{\epsilon}_k) = \boldsymbol{\Sigma_\epsilon} = (\sigma_{ij})$ for $k = 1, \ldots, n$. Also $E(\mathbf{e_i}) = \mathbf{0}$ while $Cov(\mathbf{e_i}, \mathbf{e_j}) = \sigma_{\mathbf{ij}}\boldsymbol{I_n}$ for $i, j = 1, \ldots, m$. Then $\boldsymbol{B}$ and $\boldsymbol{\Sigma}_\epsilon$ are unknown matrices of parameters to be estimated.

$$\boldsymbol{Z} = \begin{pmatrix} Y_{1,1} & Y_{1,2} & \cdots & Y_{1,m} \\ Y_{2,1} & Y_{2,2} & \cdots & Y_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n,1} & Y_{n,2} & \cdots & Y_{n,m} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Y}_1 & \boldsymbol{Y}_2 & \cdots & \boldsymbol{Y}_m \end{pmatrix} = \begin{pmatrix} \boldsymbol{y}_1^T \\ \vdots \\ \boldsymbol{y}_n^T \end{pmatrix}$$

The $n \times p$ matrix $\boldsymbol{X}$ is not necessarily of full rank p, and

$$\boldsymbol{X} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,p} \end{pmatrix} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_p \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_n^T \end{pmatrix}$$

where often $\mathbf{v}_1 = \mathbf{1}$

The $p \times m$ matrix

$$\boldsymbol{B} = \begin{pmatrix} \beta_{1,1} & \beta_{1,2} & \cdots & \beta_{1,m} \\ \beta_{2,1} & \beta_{2,2} & \cdots & \beta_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{p,1} & \beta_{p,2} & \cdots & \beta_{p,m} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \cdots & \boldsymbol{\beta}_m \end{pmatrix}.$$

The $n \times m$ matrix

$$\boldsymbol{E} = \begin{pmatrix} \epsilon_{1,1} & \epsilon_{1,2} & \cdots & \epsilon_{1,m} \\ \epsilon_{2,1} & \epsilon_{2,2} & \cdots & \epsilon_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_{n,1} & \epsilon_{n,2} & \cdots & \epsilon_{n,m} \end{pmatrix} = \begin{pmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_m \end{pmatrix} = \begin{pmatrix} \boldsymbol{\epsilon}_1^T \\ \vdots \\ \boldsymbol{\epsilon}_n^T \end{pmatrix}.$$

Each response variable in a MANOVA model follows an ANOVA model $\boldsymbol{Y}_j = \boldsymbol{X}\boldsymbol{\beta}_j + \boldsymbol{e}_j$ for $j = 1, \ldots, m$ where it is assumed that $E(\boldsymbol{e}_i) = \boldsymbol{0}$ and $Cov(\boldsymbol{e}_j) = \sigma_{jj}\boldsymbol{I}_n$.

MANOVA models are often fit by least squares. The least squares estimators $\hat{\boldsymbol{B}}$ of $\boldsymbol{B}$ are

$$\hat{\boldsymbol{B}} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^- \boldsymbol{X}^T\boldsymbol{Z} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 & \hat{\boldsymbol{\beta}}_2 & \ldots & \hat{\boldsymbol{\beta}}_m \end{pmatrix}$$

where $\left(\boldsymbol{X}^T\boldsymbol{X}\right)^-$ is a generalized inverse of $\boldsymbol{X}^T\boldsymbol{X}$. If $\boldsymbol{X}$ has a full rank then $\left(\boldsymbol{X}^T\boldsymbol{X}\right)^- = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}$ and $\hat{\boldsymbol{B}}$ is unique.

Definition. The predicted values or fitted values

$$\hat{\boldsymbol{Z}} = \boldsymbol{X}\hat{\boldsymbol{B}} = \begin{pmatrix} \hat{\boldsymbol{Y}}_1 & \hat{\boldsymbol{Y}}_2 & \cdots & \hat{\boldsymbol{Y}}_m \end{pmatrix} = \begin{pmatrix} \hat{Y}_{1,1} & \hat{Y}_{1,2} & \cdots & \hat{Y}_{1,m} \\ \hat{Y}_{2,1} & \hat{Y}_{2,2} & \cdots & \hat{Y}_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{Y}_{n,1} & \hat{Y}_{n,2} & \cdots & \hat{Y}_{n,m} \end{pmatrix}.$$

The residuals $\hat{\boldsymbol{E}} = \boldsymbol{Z} - \hat{\boldsymbol{Z}} = \boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{B}}$.

Finally,

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}} = \frac{(\boldsymbol{Z} - \hat{\boldsymbol{Z}})^T(\boldsymbol{Z} - \hat{\boldsymbol{Z}})}{n - p} = \frac{\hat{\boldsymbol{E}}^T\hat{\boldsymbol{E}}}{n - p}.$$

1.2 ONE WAY MANOVA

Assume that there are independent random samples of size n_i from p different populations, or n_i cases are randomly assigned to p treatment groups. Let $n = \sum_{i=1}^{p} n_i$ be the total sample size. Also assume that m response variables $\boldsymbol{y}_{ij} = (Y_{ij1}, \ldots, Y_{ijm})^T$ are measured for ith treatment group and jth case. Assume $E(\boldsymbol{y}_{ij}) = \boldsymbol{\mu}_i$ and $Cov(\boldsymbol{y}_{ij}) = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$.

The one way MANOVA is used to test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_p$. Note that if $m = 1$ the one way MANOVA model becomes the one way ANOVA model. One might think that performing m ANOVA tests is sufficient to test the above hypotheses. But the separate ANOVA tests would not take the correlation between the m variables into account. On the other hand the MANOVA test will take the correlation into account.

Let $\bar{\boldsymbol{y}} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} \boldsymbol{y}_{ij}/n$ be the overall mean. Let $\bar{\boldsymbol{y}}_i = \sum_{j=1}^{n_i} \boldsymbol{y}_{ij}/n_i$. Several $m \times m$ matrices will be useful. Let $\boldsymbol{S}_i$ be the sample covariance matrix corresponding to the ith treatment group. Then the within sum of squares and cross products matrix is $\boldsymbol{W} = (n_1 - 1)\boldsymbol{S}_1 + \cdots + (n_p - 1)\boldsymbol{S}_p = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_i)(\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_i)^T$. Then $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}} = \boldsymbol{W}/(n - p)$. The treatment or between sum of squares and cross products matrix is

$$\boldsymbol{B}_T = \sum_{i=1}^{p} n_i (\overline{\boldsymbol{y}}_i - \overline{\boldsymbol{y}})(\overline{\boldsymbol{y}}_i - \overline{\boldsymbol{y}})^T.$$

The total corrected (for the mean) sum of squares and cross products matrix is $\boldsymbol{T} = \boldsymbol{B}_T + \boldsymbol{W} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}})(\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}})^T$. Note that $\boldsymbol{S} = \boldsymbol{T}/(n - 1)$ is the usual sample covariance matrix of the $\boldsymbol{y}_{ij}$ if it is assumed that all n of the $\boldsymbol{y}_{ij}$ are iid so that the $\boldsymbol{\mu}_i \equiv \boldsymbol{\mu}$ for $i = 1, ..., p$.

The one way MANOVA model is $\boldsymbol{y}_{ij} = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_{ij}$ where the $\boldsymbol{\epsilon}_{ij}$ are iid with $E(\boldsymbol{\epsilon}_{ij}) = \boldsymbol{0}$ and $Cov(\boldsymbol{\epsilon}_{ij}) = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$. The summary one Way MANOVA table is shown bellow.

Source	matrix	df
Treatment or Between	$\boldsymbol{B}_T$	$p - 1$
Residual or Error or Within	$\boldsymbol{W}$	$n - p$
Total (Corrected)	T	$n - 1$

There are three commonly used test statistics to test the above hypotheses. Namely,

1. Hotelling Lawley trace statistic: $U = tr(\boldsymbol{B}_T \boldsymbol{W}^{-1}) = tr(\boldsymbol{W}^{-1} \boldsymbol{B}_T)$.

2. Wilks' lambda: $\Lambda = \frac{|\boldsymbol{W}|}{|\boldsymbol{B}_T + \boldsymbol{W}|}$.

3. Pillai's trace statistic: $\boldsymbol{V} = tr(\boldsymbol{B}_T \boldsymbol{T}^{-1}) = tr(\boldsymbol{T}^{-1}\boldsymbol{B}_T)$.

If the $\boldsymbol{y}_{ij} - \boldsymbol{\mu}_j$ are iid with common covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$, and if H_0 is true, then under regularity conditions Fujikoshi (2002) showed

1. $(n-m-p-1)U \xrightarrow{D} \chi^2_{m(p-1)}$,

2. $-[n - 0.5(m+p-2)]log(\Lambda) \xrightarrow{D} \chi^2_{m(p-1)}$, and

3. $(n-1)V \xrightarrow{D} \chi^2_{m(p-1)}$.

Note that the common covariance matrix assumption implies that each of the p treatment groups or populations has the same covariance matrix $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ for $i = 1, ..., p$, an extremely strong assumption. Kakizawa (2009) and Olive, Pelawa Watagoda, and Rupasinghe Arachchige Don (2015) show that similar results hold for the multivariate linear model. The common covariance matrix assumption, $\text{Cov}(\boldsymbol{\epsilon}_k) = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ for $k = 1, ..., n$, is often reasonable for the multivariate linear regression model.

1.3 TWO GROUP CASE

Suppose there are two independent random samples from two populations or groups. A common multivariate two sample test of hypotheses is $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ versus $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ where $\boldsymbol{\mu}_i$ is a population location measure of the ith population for $i = 1, 2$. The two sample Hotelling's T^2 test is the classical method, and is a special case of the one way MANOVA model if the two populations are assumed to have the same population covariance matrix.

1.3.1 Two Sample Hotelling's T^2 Test

Suppose there are two independent random samples $\boldsymbol{x}_{1,1}, ..., \boldsymbol{x}_{n_1,1}$ and $\boldsymbol{x}_{1,2}, ..., \boldsymbol{x}_{n_2,2}$ from two populations or groups, and that it is desired to test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ versus $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ where the $\boldsymbol{\mu}_i$ are $p \times 1$ vectors. Assume that T_i satisfy a central limit type theorem $\sqrt{n}(T_i - \boldsymbol{\mu}_i) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$ for $i = 1, 2$ where the $\boldsymbol{\Sigma}_i$ are positive definite.

To simplify large sample theory, assume $n_1 = kn_2$ for some positive real number k. Let $\hat{\boldsymbol{\Sigma}}_i$ be a consistent nonsingular estimator of $\boldsymbol{\Sigma}_i$. Then

$$\begin{pmatrix} \sqrt{n_1}\,(T_1 - \boldsymbol{\mu}_1) \\ \sqrt{n_2}\,(T_2 - \boldsymbol{\mu}_2) \end{pmatrix} \xrightarrow{D} N_{2p}\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 \end{pmatrix} \right],$$

or

$$\begin{pmatrix} \sqrt{n_2}\,(T_1 - \boldsymbol{\mu}_1) \\ \sqrt{n_2}\,(T_2 - \boldsymbol{\mu}_2) \end{pmatrix} \xrightarrow{D} N_{2p}\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \frac{\boldsymbol{\Sigma}_1}{k} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 \end{pmatrix} \right].$$

Hence

$$\sqrt{n_2}\,[(T_1 - T_2) - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)] \xrightarrow{D} N_p\left(\mathbf{0}, \frac{\boldsymbol{\Sigma}_1}{k} + \boldsymbol{\Sigma}_2\right).$$

Using $n\boldsymbol{B}^{-1} = \left(\frac{\boldsymbol{B}}{n}\right)^{-1}$ and $n_2 k = n_1$, if $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, then

$$n_2(T_1 - T_2)^T \left(\frac{\boldsymbol{\Sigma}_1}{k} + \boldsymbol{\Sigma}_2\right)^{-1} (T_1 - T_2) =$$

$$(T_1 - T_2)^T \left(\frac{\boldsymbol{\Sigma}_1}{n_1} + \frac{\boldsymbol{\Sigma}_2}{n_2}\right)^{-1} (T_1 - T_2) \xrightarrow{D} \chi^2_p.$$

Hence

$$T_0^2 = (T_1 - T_2)^T \left(\frac{\hat{\boldsymbol{\Sigma}}_1}{n_1} + \frac{\hat{\boldsymbol{\Sigma}}_2}{n_2}\right)^{-1} (T_1 - T_2) \xrightarrow{D} \chi^2_p. \tag{1.1}$$

Note that k drops out of the above result.

If the sequence of positive integers $d_n \to \infty$ and $Y_n \sim F_{p,d_n}$, then $Y_n \xrightarrow{D} \chi^2_p/p$. Using an F_{p,d_n} distribution instead of a χ^2_p distribution is similar to using a t_{d_n} distribution instead of a standard normal $N(0,1)$ distribution for inference. Instead of rejecting H_0 when $T_0^2 > \chi^2_{p,1-\delta}$, reject H_0 when

$$T_0^2 > pF_{p,d_n,1-\delta} = \frac{pF_{p,d_n,1-\delta}}{\chi^2_{p,1-\delta}} \chi^2_{p,1-\delta}.$$

The term $\frac{pF_{p,d_n,1-\delta}}{\chi^2_{p,1-\delta}}$ can be regarded as a small sample correction factor that improves the test's performance for small samples. For example, use $d_n = \min(n_1 - p, n_2 - p)$. Here $P(Y_n \leq \chi^2_{p,\delta}) = \delta$ if Y_n has a χ^2_p distribution, and $P(Y_n \leq F_{p,d_n,\delta}) = \delta$ if Y_n has an F_{p,d_n} distribution.

The two sample Hotelling's T^2 test is the classical method. If it is not assumed that the population covariance matrices are equal, then this test uses the sample mean and sample covariance matrix $T_i = \overline{\boldsymbol{x}}_i$ and $\hat{\boldsymbol{\Sigma}}_i = \boldsymbol{S}_i$ applied to each sample. This test has considerable robustness to the assumption that both populations have a multivariate normal distribution and to the assumption that the populations have a common population covariance matrix $\boldsymbol{\Sigma}$, but the test can be very poor if outliers are present.

Alternative statistics to the sample mean can be useful, but large sample tests of the form of (1.1) need practical consistent estimators $\hat{\boldsymbol{\Sigma}}_i$ of the two asymptotic covariance matrices $\boldsymbol{\Sigma}_i$.

Chapter 2 gives theory and methods for bootstrapping hypotheses tests and shows how to apply the bootstrap to test the hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{c}$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{c}$. Chapter 3 suggests using the Olive (2017abc) bootstrap technique to develop analogs of the Hotelling's T^2 test that use a statistic T_i, such as the coordinatewise median, applied to the ith sample for $i = 1, 2$. These tests are useful if the asymptotic covariance matrix is unknown or difficult to estimate. The new tests can have considerable outlier resistance, and the tests do not need the population covariance matrices to be equal. Chapter 4 suggests using the Olive (2017abc) bootstrap technique to develop analogs of the one way MANOVA test. The new tests can have considerable outlier resistance, and the tests do not need the population covariance matrices to be equal. Chapters 5 and 6 give some simulations and examples.

CHAPTER 2

THEORY AND METHODS

2.1 NOTATION

2.1.1 Mahalanobis Distance

Let the $p \times 1$ column vector T be a multivariate location estimator, and let the $p \times p$ symmetric positive definite matrix $\boldsymbol{C}$ be a dispersion estimator. Then the ith *squared sample Mahalanobis distance* is the scalar

$$D_i^2 = D_i^2(T, \boldsymbol{C}) = D_{\boldsymbol{x}_i}^2(T, \boldsymbol{C}) = (\boldsymbol{x}_i - T)^T \boldsymbol{C}^{-1} (\boldsymbol{x}_i - T) \tag{2.1}$$

for each observation $\boldsymbol{x}_i$. Notice that the Euclidean distance of $\boldsymbol{x}_i$ from the estimate of center T is $D_i(T, \boldsymbol{I}_p)$ where $\boldsymbol{I}_p$ is the $p \times p$ identity matrix. The classical Mahalanobis distance uses $(T, \boldsymbol{C}) = (\overline{\boldsymbol{x}}, \boldsymbol{S})$, the sample mean and sample covariance matrix where

$$\overline{\boldsymbol{x}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \text{ and } \boldsymbol{S} = \frac{1}{\mathrm{n}-1} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\boldsymbol{x}_{\mathrm{i}} - \overline{\boldsymbol{x}})(\boldsymbol{x}_{\mathrm{i}} - \overline{\boldsymbol{x}})^{\mathrm{T}}. \tag{2.2}$$

2.2 PREDICTION REGION

A large sample $100(1-\delta)\%$ prediction region is the hyperellipsoid

$$\{\boldsymbol{w} : D_{\boldsymbol{w}}^2(\overline{\boldsymbol{x}}, \boldsymbol{S}) \leq D_{(c)}^2\} = \{\boldsymbol{w} : D_{\boldsymbol{w}}(\overline{\boldsymbol{x}}, \boldsymbol{S}) \leq D_{(c)}\} \tag{2.3}$$

for appropriate c. Using $c = \lceil n(1-\delta) \rceil$ covers about $100(1-\delta)\%$ of the training data cases $\boldsymbol{x}_i$, but the prediction region will have coverage lower than the nominal coverage of $1-\delta$ for moderate n. This result is not surprising since empirically statistical methods perform worse on test data. Increasing c will improve the coverage for moderate samples. Let $q_n = \min(1-\delta+0.05, 1-\delta+p/n)$ for $\delta > 0.1$ and

$$q_n = \min(1-\delta/2, 1-\delta+10\delta p/n), \quad \text{otherwise.} \tag{2.4}$$

If $1-\delta < 0.999$ and $q_n < 1-\delta+0.001$, set $q_n = 1-\delta$.

Let $D_{(U_n)}$ be the $100q_n$th percentile of the D_i. Then the Olive (2013) large sample $100(1-\delta)\%$ nonparametric prediction region for a future value $\boldsymbol{x}_f$ given iid data $\boldsymbol{x}_1, ..., , \boldsymbol{x}_n$

is

$$\{\boldsymbol{w} : D^2_{\boldsymbol{w}}(\overline{\boldsymbol{x}}, \boldsymbol{S}) \leq D^2_{(U_n)}\}, \tag{2.5}$$

while the classical large sample $100(1-\delta)\%$ prediction region is

$$\{\boldsymbol{w} : D^2_{\boldsymbol{w}}(\overline{\boldsymbol{x}}, \boldsymbol{S}) \leq \chi^2_{p,1-\delta}\}. \tag{2.6}$$

2.3 PREDICTION REGION METHOD

Olive (2017bc) shows that there is a useful relationship between prediction regions and confidence regions. Consider predicting a future $p \times 1$ test vector $\boldsymbol{x}_f$, given past training data $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$. A *large sample* $100(1-\delta)\%$ *prediction region* is a set $\mathcal{A}_n$ such that $P(\boldsymbol{x}_f \in \mathcal{A}_n) \rightarrow 1-\delta$ while a large sample $100(1-\delta)\%$ confidence region for a parameter $\boldsymbol{\mu}$ is a set $\mathcal{A}_n$ such that $P(\boldsymbol{\mu} \in \mathcal{A}_n) \rightarrow 1-\delta$ as $n \rightarrow \infty$. Consider testing $H_0 : \boldsymbol{\mu} = \boldsymbol{c}$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{c}$ where $\boldsymbol{c}$ is a known $p \times 1$ vector.

The Olive (2017abc) prediction region method obtains a confidence region for $\boldsymbol{\mu}$ by applying the nonparametric prediction region (2.5) to the bootstrap sample $T^*_1, ..., T^*_B$, and the theory for the method is sketched below. Let $\overline{T}^*$ and $\boldsymbol{S}^*_T$ be the sample mean and sample covariance matrix of the bootstrap sample. Following Bickel and Ren (2001), let the vector of parameters $\boldsymbol{\mu} = T(F)$, the statistic $T_n = T(F_n)$, and $T^* = T(F^*_n)$ 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 T is sufficiently smooth (Hadamard differentiable with a Hadamard derivative $\dot{T}(F)$), then $\sqrt{n}(T_n - \boldsymbol{\mu}) \xrightarrow{D} \boldsymbol{X}$ and $\sqrt{n}(T^*_i - \overline{T}^*) \xrightarrow{D} \boldsymbol{X}$ with $\boldsymbol{X} = \dot{T}(F)\boldsymbol{z}_F$. Olive (2017bc) uses these results to show that if $\boldsymbol{X} \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_T)$, then $\sqrt{n}(\overline{T}^* - T_n) \xrightarrow{D} \boldsymbol{0}$, $\sqrt{n}(\overline{T}^* - \boldsymbol{\mu}) \xrightarrow{D} \boldsymbol{X}$, and that the prediction region method large sample $100(1-\delta)\%$ confidence region for $\boldsymbol{\mu}$ is

$$\{\boldsymbol{w} : (\boldsymbol{w} - \overline{T}^*)^T [\boldsymbol{S}^*_T]^{-1} (\boldsymbol{w} - \overline{T}^*) \leq D^2_{(U_B)}\} = \{\boldsymbol{w} : D^2_{\boldsymbol{w}}(\overline{T}^*, \boldsymbol{S}^*_T) \leq D^2_{(U_B)}\} \tag{2.7}$$

where $D^2_{(U_B)}$ is computed from $D^2_i = (T^*_i - \overline{T}^*)^T [\boldsymbol{S}^*_T]^{-1} (T^*_i - \overline{T}^*)$ for $i = 1, ..., B$. Note that the corresponding test for $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ rejects H_0 if $(\overline{T}^* - \boldsymbol{\mu}_0)^T [\boldsymbol{S}^*_T]^{-1} (\overline{T}^* - \boldsymbol{\mu}_0) > D^2_{(U_B)}$. This procedure is basically the one sample Hotelling's T^2 test applied to the T^*_i using $\boldsymbol{S}^*_T$ as the estimated covariance matrix and replacing the $\chi^2_{p,1-\delta}$ cutoff by $D^2_{(U_B)}$.

2.3.1 Testing $H_0 : \boldsymbol{\mu} = \boldsymbol{c}$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{c}$ Using the Prediction Region Method

The prediction region method for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{c}$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{c}$ is simple. Let $\hat{\boldsymbol{\mu}}$ be a consistent estimator of $\boldsymbol{\mu}$ and make a bootstrap sample $\boldsymbol{w}_i = \hat{\boldsymbol{\mu}}_i^* - \boldsymbol{c}$ for $i = 1, ..., B$. Make the nonparametric prediction region (2.7) for the $\boldsymbol{w}_i$ and fail to reject H_0 if $\mathbf{0}$ is in the prediction region, reject H_0 otherwise.

The Bickel and Ren (2001) hypothesis testing method is equivalent to using confidence region (2.7) with $\overline{T}^*$ replaced by T_n and U_B replaced by $\lceil B(1-\delta) \rceil$. If region (2.7) or the Bickel and Ren (2001) region is a large sample $100(1-\delta)\%$ confidence region, then so is the other region if $\sqrt{n}(\overline{T}^* - T_n) \xrightarrow{D} \mathbf{0}$. Hadamard differentiability and asymptotic normality are sufficient conditions for both regions to be large sample confidence regions if $\boldsymbol{S}_T^* \xrightarrow{D} \boldsymbol{\Sigma}_T$, but 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. Since the univariate sample mean, sample median, and sample trimmed mean are Hadamard differentiable and asymptotically normal, each coordinate satisfies $\sqrt{n}(T_{in} - \overline{T}_i^*) \xrightarrow{D} 0$ for $i = 1, ..., p$. Hence $\sqrt{n}(T_n - \overline{T}^*) \xrightarrow{D} \mathbf{0}$, and (2.7) is a large sample $100(1-\delta)\%$ confidence region if T_n is the coordinatewise sample mean, median, or trimmed mean.

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

2.4 A RELATIONSHIP BETWEEN THE ONE-WAY MANOVA TEST AND THE HOTELLING LAWLEY TRACE TEST

An alternative method for one way MANOVA is to use the model $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}$ with $\boldsymbol{X}$, $\boldsymbol{Z}$ and $\boldsymbol{B}$ as follows.

Let

$$\boldsymbol{Y}_{ij} = \begin{pmatrix} Y_{ij1} \\ \vdots \\ Y_{ijm} \end{pmatrix} = \boldsymbol{\mu}_i + \boldsymbol{e}_{ij}, \qquad E\boldsymbol{Y}_{ij} = \boldsymbol{\mu}_i = \begin{pmatrix} \mu_{ij1} \\ \vdots \\ \mu_{ijm} \end{pmatrix}$$

for $i = 1, \ldots, p$ and $j = 1, \ldots, n_i$

Then $\boldsymbol{X}$ is a full rank where the ith column of $\boldsymbol{X}$ is an indicator for group $i-1$ for $i = 2, \ldots, p$, and $\boldsymbol{Z}$ are as follows.

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{Y}_{11}^T \\ \vdots \\ \boldsymbol{Y}_{1n_1}^T \\ \boldsymbol{Y}_{21}^T \\ \vdots \\ \boldsymbol{Y}_{2n_2}^T \\ \vdots \\ \boldsymbol{Y}_{p-1,1}^T \\ \vdots \\ \boldsymbol{Y}_{p-1,n_{p-1}}^T \\ \vdots \\ \boldsymbol{Y}_{p,1}^T \\ \vdots \\ \boldsymbol{Y}_{p,n_p}^T \end{pmatrix},$$

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix} \tag{2.8}$$

$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{\mu}_p^T \\ (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_p)^T \\ \vdots \\ (\boldsymbol{\mu}_{p-1} - \boldsymbol{\mu}_p)^T \end{pmatrix}$ *and* $\boldsymbol{L} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{I}_{p-1} \end{pmatrix}$. Note that $\boldsymbol{Y}_{ij}^T = \boldsymbol{\mu}_i^T + \boldsymbol{e}_{ij}^T$.

Then

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{pmatrix} n & n_1 & n_2 & \cdots & n_{p-1} \\ n_1 & n_1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ n_{p-2} & 0 & \cdots & n_{p-2} & 0 \\ n_{p-1} & 0 & \cdots & 0 & n_{p-1} \end{pmatrix} \tag{2.9}$$

and

$$\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1} = \frac{1}{n_p}\begin{pmatrix} 1 & -1 & -1 & \cdots & -1 \\ -1 & 1+\frac{n_p}{n_1} & 1 & \cdots & 1 \\ \vdots & & \ddots & & \vdots \\ -1 & 1 & \cdots & 1+\frac{n_p}{n_{p-2}} & 1 \\ -1 & 1 & \cdots & 1 & 1+\frac{n_p}{n_{p-1}} \end{pmatrix}. \tag{2.10}$$

Then the least square estimators $\hat{\boldsymbol{B}}$ of $\boldsymbol{B}$,

$$\hat{\boldsymbol{B}} = \begin{pmatrix} \bar{\mathbf{y}}_p^T \\ (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_p)^T \\ \vdots \\ (\bar{\mathbf{y}}_{p-1} - \bar{\mathbf{y}}_p)^T \end{pmatrix}, \qquad \textit{and} \qquad \boldsymbol{L}\hat{\boldsymbol{B}} = \begin{pmatrix} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_p)^T \\ (\bar{\mathbf{y}}_2 - \bar{\mathbf{y}}_p)^T \\ \vdots \\ (\bar{\mathbf{y}}_{p-1} - \bar{\mathbf{y}}_p)^T \end{pmatrix}.$$

Then $\boldsymbol{L}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{L}^T$ becomes

$$\boldsymbol{L}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{L}^T = \frac{1}{n_p}\begin{pmatrix} 1+\frac{n_p}{n_1} & 1 & 1 & \cdots & 1 \\ 1 & 1+\frac{n_p}{n_2} & 1 & \cdots & 1 \\ \vdots & & \ddots & & \vdots \\ 1 & 1 & \cdots & 1 & 1+\frac{n_p}{n_{p-1}} \end{pmatrix}. \tag{2.11}$$

It can be shown that the inverse of the above matrix is

$$\left[\boldsymbol{L}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{L}^T\right]^{-1} = \frac{1}{n}\begin{pmatrix} n_1(n-n_1) & -n_1n_2 & -n_1n_3 & \cdots & -n_1n_{p-1} \\ -n_1n_2 & n_2(n-n_2) & -n_2n_3 & \cdots & -n_2n_{p-1} \\ \vdots & & \ddots & & \vdots \\ -n_1n_{p-1} & -n_2n_{p-1} & \cdots & & n_{p-1}(n-n_{p-1}) \end{pmatrix}.$$

For convenience, write $\left[\boldsymbol{L}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{L}^T\right]^{-1}$ as follows

$$\left[\boldsymbol{L}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{L}^T\right]^{-1} =$$

$$\frac{1}{n}\begin{pmatrix} -n_1^2 & -n_1n_2 & -n_1n_3 & \cdots & -n_1n_{p-1} \\ -n_1n_2 & -n_2^2 & -n_2n_3 & \cdots & -n_2n_{p-1} \\ \vdots & & \ddots & & \vdots \\ -n_1n_{p-1} & -n_2n_{p-1} & \cdots & & -n_{p-1}^2 \end{pmatrix} + \begin{pmatrix} n_1 & 0 & 0 & \cdots & 0 \\ 0 & n_2 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 0 & n_{p-1} \end{pmatrix}.$$

Then,

$$\left(\boldsymbol{L}\hat{\boldsymbol{B}}\right)^T \left[\boldsymbol{L}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{L}^T\right]^{-1}\left(\boldsymbol{L}\hat{\boldsymbol{B}}\right) =$$

$$-\frac{1}{n}\sum_{i=1}^{p-1}\sum_{j=1}^{p-1} n_i n_j(\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_p)(\bar{\mathbf{y}}_j - \bar{\mathbf{y}}_p)^T + \sum_{i=1}^{p-1} n_i(\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_p)(\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_p)^T = \boldsymbol{H}.$$

Let $\boldsymbol{X}$ be as in (2.8). Then the multivariate linear regression (MREG) Hotelling Lawley test statistic for testing $H_0 : \boldsymbol{LB} = \boldsymbol{0}$ versus $H_0 : \boldsymbol{LB} \neq \boldsymbol{0}$ has

$$U = tr(\boldsymbol{W}^{-1}\boldsymbol{H}).$$

One way MANOVA is used to test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_p$. The Hotelling Lawley test statistic for testing for above hypotheses is

$$U = tr(\boldsymbol{W}^{-1}\boldsymbol{B}_T)$$

where

$$\boldsymbol{W} = (n-p)\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}} \quad and \quad \boldsymbol{B}_T = \sum_{i=1}^{p} n_i(\bar{\mathbf{y}}_i - \bar{\mathbf{y}})(\bar{\mathbf{y}}_i - \bar{\mathbf{y}})^T.$$

Theorem 2.1. *The one-way MANOVA test statistic and the Hotelling Lawley trace test statistic are the same for the design matrix as in (2.8).*

To show that the above two test statistics are equal it is sufficient to prove that $\boldsymbol{H} = \boldsymbol{B}_T$.

Proof. **Special case I:** $p = 2$ (Two group case)

Consider $\boldsymbol{H}$.

$\boldsymbol{H} = -\frac{1}{n}n_1 n_1(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T + n_1(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T$. Since $n = n_1 + n_2$,

$\boldsymbol{H} = -\frac{1}{n}(nn_1 - n_1 n_2)(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T + n_1(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T$

$\boldsymbol{H} = -n_1(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T + \frac{n_1 n_2}{n}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T + n_1(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T$

$\boldsymbol{H} = \frac{n_1 n_2}{n}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T$.

Now consider $\boldsymbol{B}_T$ with $p = 2$.

Note that $\bar{\mathbf{y}} = (n_1\bar{\mathbf{y}}_1 + n_2\bar{\mathbf{y}}_2)/n$ and

$\boldsymbol{B}_T = n_1(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}})(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}})^T + n_2(\bar{\mathbf{y}}_2 - \bar{\mathbf{y}})(\bar{\mathbf{y}}_2 - \bar{\mathbf{y}})^T$

$\boldsymbol{B}_T = \frac{n_1}{n^2}(n\bar{\mathbf{y}}_1 - n_1\bar{\mathbf{y}}_1 - n_2\bar{\mathbf{y}}_2)(n\bar{\mathbf{y}}_1 - n_1\bar{\mathbf{y}}_1 - n_2\bar{\mathbf{y}}_1)^T + \frac{n_2}{n^2}(n\bar{\mathbf{y}}_2 - n_1\bar{\mathbf{y}}_1 - n_2\bar{\mathbf{y}}_2)(n\bar{\mathbf{y}}_2 - n_1\bar{\mathbf{y}}_1 - n_2\bar{\mathbf{y}}_2)^T$

$\boldsymbol{B}_T = \frac{n_1 n_2^2}{n^2}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T + \frac{n_1^2 n_2}{n^2}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T$

$\boldsymbol{B}_T = \frac{n_1 n_2}{n}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)^T.$

Therefore $\boldsymbol{B}_T = \boldsymbol{H}$ when $p = 2$.

□

Proof. **Special case II:** $n_i = n_1 \ \forall i = 1, \ldots, p$

$$\boldsymbol{H} = -\frac{1}{n}\sum_{i=1}^{p-1}\sum_{j=1}^{p-1} n_i n_j(\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_p)(\bar{\mathbf{y}}_j - \bar{\mathbf{y}}_p)^T + \sum_{i=1}^{p-1} n_i(\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_p)(\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_p)^T.$$

Note that the i, j running from 1 through $p-1$ and i, j running from 1 through p would yield the same $\boldsymbol{H}$. Therefore $\boldsymbol{H}$ can be written as

$$\boldsymbol{H} = -\frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{p} n_i n_j(\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_p)(\bar{\mathbf{y}}_j - \bar{\mathbf{y}}_p)^T + \sum_{i=1}^{p} n_i(\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_p)(\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_p)^T.$$

Now consider the double sum in $\boldsymbol{H}$. Note that $n = n_1 p$ and

$$-\frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{p} n_i n_j(\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_p)(\bar{\mathbf{y}}_j - \bar{\mathbf{y}}_p)^T = \frac{-n_1^2}{n_1 p}\sum_{i=1}^{p}\sum_{j=1}^{p}\left(\bar{\mathbf{y}}_i\bar{\mathbf{y}}_j^T - \bar{\mathbf{y}}_i\bar{\mathbf{y}}_p^T - \bar{\mathbf{y}}_p\bar{\mathbf{y}}_j^T + \bar{\mathbf{y}}_p\bar{\mathbf{y}}_p^T\right)$$

$$= \frac{n_1}{p}\left[-\sum_{i=1}^{p}\sum_{j=1}^{p}\left(\bar{\mathbf{y}}_i\bar{\mathbf{y}}_j^T\right) + p\left(\sum_{i=1}^{p}\bar{\mathbf{y}}_i\right)\bar{\mathbf{y}}_p^T + p\bar{\mathbf{y}}_p\left(\sum_{j=1}^{p}\bar{\mathbf{y}}_j^T\right) - p^2\bar{\mathbf{y}}_p\bar{\mathbf{y}}_p^T\right]. \tag{2.12}$$

Now consider the rest of $\boldsymbol{H}$,

$$n_1\sum_{i=1}^{p}(\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_p)(\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_p)^T = n_1\sum_{i=1}^{p}\bar{\mathbf{y}}_i\bar{\mathbf{y}}_i^T - n_1\left(\sum_{i=1}^{p}\bar{\mathbf{y}}_i\right)\bar{\mathbf{y}}_p^T - n_1\bar{\mathbf{y}}_p\left(\sum_{i=1}^{p}\bar{\mathbf{y}}_i^T\right) + n_1 p\bar{\mathbf{y}}_p\bar{\mathbf{y}}_p^T. \tag{2.13}$$

Therefore by (2.12) and (2.13), it is clear that

$$\boldsymbol{H} = n_1\sum_{i=1}^{p}\bar{\mathbf{y}}_i\bar{\mathbf{y}}_i^T - \frac{n_1}{p}\sum_{i=1}^{p}\sum_{j=1}^{p}\bar{\mathbf{y}}_i\bar{\mathbf{y}}_j^T. \tag{2.14}$$

Now consider

$$\boldsymbol{B}_T = n_1\sum_{i=1}^{p}(\bar{\mathbf{y}}_i - \bar{\mathbf{y}})(\bar{\mathbf{y}}_i - \bar{\mathbf{y}})^T. \tag{2.15}$$

Let

$$\bar{\boldsymbol{Y}} = \begin{pmatrix} \bar{\mathbf{y}}_1^T \\ \bar{\mathbf{y}}_2^T \\ \vdots \\ \bar{\mathbf{y}}_p^T \end{pmatrix}. \quad Then \quad \boldsymbol{B}_T = n_1 \left[\bar{\boldsymbol{Y}}^T \bar{\boldsymbol{Y}} - \frac{1}{p} \bar{\boldsymbol{Y}}^T \mathbf{1}\mathbf{1}^{\mathbf{T}} \bar{\boldsymbol{Y}} \right].$$

Therefore, $\boldsymbol{B}_T$ becomes

$$\boldsymbol{B}_T = n_1 \sum_{i=1}^{p} \bar{\mathbf{y}}_i \bar{\mathbf{y}}_i^T - \frac{n_1}{p} \sum_{i=1}^{p} \sum_{j=1}^{p} \bar{\mathbf{y}}_i \bar{\mathbf{y}}_j^T. \tag{2.16}$$

From (2.15) and (2.16) $\boldsymbol{B}_T = \boldsymbol{H}$.

□

Proof. **General case:**

$$\boldsymbol{H} = -\frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{p} n_i n_j (\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_p)(\bar{\mathbf{y}}_j - \bar{\mathbf{y}}_p)^T + \sum_{i=1}^{p} n_i (\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_p)(\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_p)^T.$$

Now consider the double sum in $\boldsymbol{H}$.

$$-\frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{p} n_i n_j (\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_p)(\bar{\mathbf{y}}_j - \bar{\mathbf{y}}_p)^T =$$

$$-\frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{p} n_i n_j \bar{\mathbf{y}}_i \bar{\mathbf{y}}_j^T + \frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{p} n_i n_j \bar{\mathbf{y}}_i \bar{\mathbf{y}}_p^T + \frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{p} n_i n_j \bar{\mathbf{y}}_p \bar{\mathbf{y}}_j^T - \frac{1}{n} \bar{\mathbf{y}}_p \bar{\mathbf{y}}_p^T \sum_{i=1}^{p} \sum_{j=1}^{p} n_i n_j \tag{2.17}$$

$$-\frac{1}{n} \sum_{i=1}^{p} n_i \bar{\mathbf{y}}_i \sum_{j=1}^{p} n_j \bar{\mathbf{y}}_j^T + \frac{1}{n} \sum_{i=1}^{p} n_i \bar{\mathbf{y}}_i \sum_{j=1}^{p} n_j \bar{\mathbf{y}}_p^T + \frac{1}{n} \bar{\mathbf{y}}_p \sum_{i=1}^{p} n_i \sum_{j=1}^{p} n_j \bar{\mathbf{y}}_j^T - \frac{1}{n} \bar{\mathbf{y}}_p \bar{\mathbf{y}}_p^T n^2$$

$$-\frac{1}{n} n \bar{\mathbf{y}} n \bar{\mathbf{y}}^T + \frac{1}{n} \sum_{i=1}^{p} n_i \bar{\mathbf{y}}_i n \bar{\mathbf{y}}_p^T + \frac{1}{n} \bar{\mathbf{y}}_p n \sum_{j=1}^{p} n_j \bar{\mathbf{y}}_j^T - n \bar{\mathbf{y}}_p \bar{\mathbf{y}}_p^T$$

$$-n \bar{\mathbf{y}} \bar{\mathbf{y}}^T + \sum_{i=1}^{p} n_i \bar{\mathbf{y}}_i \bar{\mathbf{y}}_p^T + \bar{\mathbf{y}}_p \sum_{j=1}^{p} n_j \bar{\mathbf{y}}_j^T - n \bar{\mathbf{y}}_p \bar{\mathbf{y}}_p^T. \tag{2.18}$$

Now consider the rest of $\boldsymbol{H}$,

$$\sum_{i=1}^{p} n_i(\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_p)(\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_p)^T = \sum_{i=1}^{p} n_i\bar{\mathbf{y}}_i\bar{\mathbf{y}}_i^T - \sum_{i=1}^{p} n_i\bar{\mathbf{y}}_i\bar{\mathbf{y}}_p^T - \bar{\mathbf{y}}_p \sum_{i=1}^{p} n_i\bar{\mathbf{y}}_i^T + n\bar{\mathbf{y}}_p\bar{\mathbf{y}}_p^T. \tag{2.19}$$

Therefore by (2.18) and (2.19)

$$\boldsymbol{H} = \sum_{i=1}^{p} n_i\bar{\mathbf{y}}_i\bar{\mathbf{y}}_i^T - n\bar{\mathbf{y}}\bar{\mathbf{y}}^T. \tag{2.20}$$

Now consider

$$\boldsymbol{B}_T = \sum_{i=1}^{p} n_i(\bar{\mathbf{y}}_i - \bar{\mathbf{y}})(\bar{\mathbf{y}}_i - \bar{\mathbf{y}})^T$$

$$\boldsymbol{B}_T = \sum_{i=1}^{p} n_i\bar{\mathbf{y}}_i\bar{\mathbf{y}}_i^T - \sum_{i=1}^{p} n_i\bar{\mathbf{y}}_i\bar{\mathbf{y}}^T - \bar{\mathbf{y}} \sum_{i=1}^{p} n_i\bar{\mathbf{y}}_i^T + \bar{\mathbf{y}}\bar{\mathbf{y}}^T \sum_{i=1}^{p} n_i$$

$$\boldsymbol{B}_T = \sum_{i=1}^{p} n_i\bar{\mathbf{y}}_i\bar{\mathbf{y}}_i^T - n\bar{\mathbf{y}}\bar{\mathbf{y}}^T - \bar{\mathbf{y}}n\bar{\mathbf{y}}^T + n\bar{\mathbf{y}}\bar{\mathbf{y}}^T$$

$$\boldsymbol{B}_T = \sum_{i=1}^{p} n_i\bar{\mathbf{y}}_i\bar{\mathbf{y}}_i^T - n\bar{\mathbf{y}}_i\bar{\mathbf{y}}^T. \tag{2.21}$$

(2.20) and (2.21) proves that $\boldsymbol{H} = \boldsymbol{B}_T$.

□

2.5 CELL MEANS MODEL

$$\boldsymbol{X} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}, \qquad \boldsymbol{B} = \begin{pmatrix} \boldsymbol{\mu}_1^T \\ \vdots \\ \boldsymbol{\mu}_p^T \end{pmatrix} \quad \textit{and} \qquad \boldsymbol{L} = \begin{pmatrix} \boldsymbol{I}_{p-1} & -\mathbf{1} \end{pmatrix}$$

$$\hat{\boldsymbol{B}} = \begin{pmatrix} \bar{\mathbf{y}}_1^T \\ \vdots \\ \bar{\mathbf{y}}_p^T \end{pmatrix}, \qquad \boldsymbol{L}\hat{\boldsymbol{B}} = \begin{pmatrix} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_p)^T \\ (\bar{\mathbf{y}}_2 - \bar{\mathbf{y}}_p)^T \\ \vdots \\ (\bar{\mathbf{y}}_{p-1} - \bar{\mathbf{y}}_p)^T \end{pmatrix}.$$

Then $\boldsymbol{X}^T\boldsymbol{X} = diag\,(n_1, \ldots, n_{p-1})$ and $(\boldsymbol{X}^T\boldsymbol{X})^{-1} = diag\left(\frac{1}{n_1}, \ldots, \frac{1}{n_{p-1}}\right)$.
Then $\boldsymbol{L}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{L}^T$ becomes

$$\boldsymbol{L}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{L}^T = \frac{1}{n_p}\begin{pmatrix} 1+\frac{n_p}{n_1} & 1 & 1 & \cdots & 1 \\ 1 & 1+\frac{n_p}{n_2} & 1 & \cdots & 1 \\ \vdots & & \ddots & & \vdots \\ 1 & 1 & \cdots & 1 & 1+\frac{n_p}{n_{p-1}} \end{pmatrix}. \tag{2.22}$$

Corollary 2.2. *Theorem 2.1 does not depends on the full rank design matrix.*

Notice that the matrix equation (2.22) is the exactly same as (2.11). This is an indication that Theorem 2.1 does not depend on the design matrix.

CHAPTER 3

BOOTSTRAPPING ANALOGS OF THE TWO SAMPLE HOTELLING'S T^2 TEST

Suppose there are two independent random samples from two populations or groups. A common multivariate two sample test of hypotheses is $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ versus $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ where $\boldsymbol{\mu}_i$ is a population location measure of the ith population for $i = 1, 2$. The two sample Hotelling's T^2 test is the classical method, and is a special case of the one way MANOVA model if the two populations are assumed to have the same population covariance matrix. This chapter suggests using the Olive (2017abc) bootstrap technique to develop analogs of Hotelling's T^2 test. The new tests can have considerable outlier resistance, and the tests do not need the population covariance matrices to be equal.

3.1 APPLYING THE PREDICTION REGION METHOD TO THE TWO SAMPLE TEST

The two sample test of $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ versus $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ uses $\boldsymbol{\mu} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{c} = \boldsymbol{0}$ with $\boldsymbol{w}_i = T^*_{i1} - T^*_{i2}$ for $i = 1, ..., B$. Make the prediction region (2.7) where $T^*_i = \boldsymbol{w}_i$. Fail to reject H_0 if $\boldsymbol{0}$ is in the prediction region, reject H_0 otherwise. A sample of size n_i is drawn with replacement from $\boldsymbol{x}_{1,i}, ..., \boldsymbol{x}_{n_i,i}$ for $i = 1, 2$ to obtain the bootstrap sample.

For illustrative purposes, the simulation study will take T_i to be the coordinatewise median, the (Olive (2017b, ch. 4), Olive and Hawkins (2010), and Zhang, Olive, and Ye (2012)) RMVN estimator T_{RMVN}, the sample mean, and the 25% trimmed mean. The asymptotic covariance matrix of the coordinatewise median is difficult to estimate, while that of the RMVN estimator is unknown. The RMVN estimator has been shown to be $\sqrt{n}$ consistent on a large class of elliptically contoured distributions, but has not yet been shown to be asymptotically normal. Hence the bootstrap "test" for the RMVN estimator should be used for exploratory purposes.

The RMVN estimator $(T_{RMVN}, \boldsymbol{C}_{RMVN})$ uses a concentration algorithm. Let $(T_{-1,j}, \boldsymbol{C}_{-1,j})$ be the jth start (initial estimator) and compute all n Mahalanobis distances $D_i(T_{-1,j}, \boldsymbol{C}_{-1,j})$. At the next iteration, the classical estimator $(T_{0,j}, \boldsymbol{C}_{0,j}) = (\overline{\boldsymbol{x}}_{0,j}, \boldsymbol{S}_{0,j})$

is computed from the $c_n \approx n/2$ cases corresponding to the smallest distances. This iteration can be continued for k *concentration steps* resulting in the sequence of estimators $(T_{-1,j}, \boldsymbol{C}_{-1,j}), (T_{0,j}, \boldsymbol{C}_{0,j}), ..., (T_{k,j}, \boldsymbol{C}_{k,j})$. The result of the iteration $(T_{k,j}, \boldsymbol{C}_{k,j})$ is called the jth *attractor*. The algorithm estimator uses one of the attractors. The RMVN estimator uses the same two starts as the Olive (2004) MBA estimator: $(\overline{\boldsymbol{x}}, \boldsymbol{S})$ and $(MED(n), \boldsymbol{I}_p)$ where $MED(n)$ is the coordinatewise median. Then the location estimator T_{RMVN} can be used to test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$.

3.2 REAL DATA EXAMPLE

The Johnson (1996) STATLIB bodyfat data consists of 252 observations on 15 variables including the density determined from underwater weighing and the percent body fat measurement. Consider these two variables with two age groups: age ≤ 50 and age > 50. The test with the RMVN estimator had $D_0 = 1.78$ while the test with the coordinatewise median had $D_0 = 1.35$. Both tests had cutoffs near 2.37 and fail to reject H_0. The classical two sample Hotelling's T^2 test rejects H_0 with a test statistic of 4.74 and a p-value of 0.001.

The DD plots, shown in Figures 3.1 and 3.2, reveal five outliers. After deleting the outliers, the three tests all fail to reject H_0. The RMVN test had $D_0 = 1.63$ with cutoff 2.25, the coordinatewise median test had $D_0 = 1.22$ with cutoff 2.38, and the classical test had test statistic 2.39 with a p-value of 0.09.

See the simulation set up and the simulation results in Chapter 5.

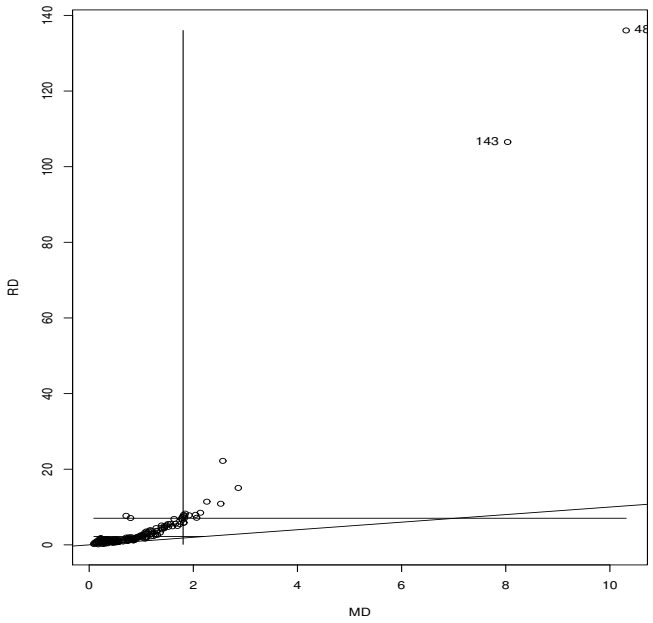

Figure 3.1. DD plot for the age ≤ 50 group.

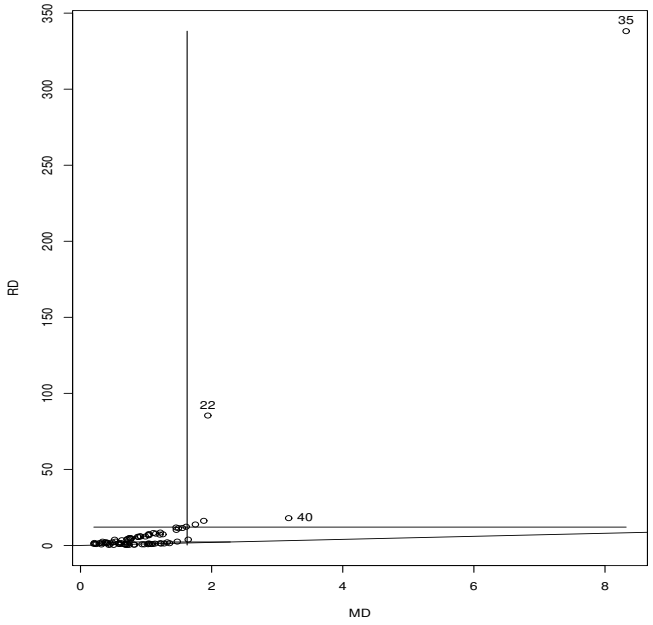

Figure 3.2. DD plot for the age > 50 group.

CHAPTER 4

BOOTSTRAPPING ANALOGS OF THE ONE WAY MANOVA TEST

The classical one way MANOVA model is used to test whether the mean measurements are the same or differ across p groups, and assumes that the covariance matrix of each group is the same. This chapter suggests using the Olive (2017abc) bootstrap technique to develop analogs of the one way MANOVA test. The new tests can have considerable outlier resistance, and the tests do not need the population covariance matrices to be equal.

The multivariate linear model

$$\boldsymbol{y}_i = \boldsymbol{B}^T \boldsymbol{x}_i + \boldsymbol{\epsilon}_i$$

for $i = 1, ..., n$ has $m \geq 2$ response variables $Y_1, ..., Y_m$ and p predictor variables $x_1, x_2, ..., x_p$. The ith case is $(\boldsymbol{x}_i^T, \boldsymbol{y}_i^T) = (x_{i1}, x_{i2}, ..., x_{ip}, Y_{i1}, ..., Y_{im})$. The model is written in matrix form as $\boldsymbol{Z} = \boldsymbol{XB} + \boldsymbol{E}$ where the matrices are defined below. The model has $E(\boldsymbol{\epsilon}_k) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{\epsilon}_k) = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = (\sigma_{ij})$ for $k = 1, ..., n$. Then the $p \times m$ coefficient matrix $\boldsymbol{B} = \begin{bmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \dots & \boldsymbol{\beta}_m \end{bmatrix}$ and the $m \times m$ covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ are to be estimated, and $E(\boldsymbol{Z}) = \boldsymbol{XB}$ while $E(Y_{ij}) = \boldsymbol{x}_i^T \boldsymbol{\beta}_j$. The $\boldsymbol{\epsilon}_i$ are assumed to be iid. The univariate linear model corresponds to $m = 1$ response variable, and is written in matrix form as $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$. Subscripts are needed for the m univariate linear models $\boldsymbol{Y}_j = \boldsymbol{X\beta}_j + \boldsymbol{e}_j$ for $j = 1, ..., m$ where $E(\boldsymbol{e}_j) = \boldsymbol{0}$. For the multivariate linear model, $\text{Cov}(\boldsymbol{e}_i, \boldsymbol{e}_j) = \sigma_{ij} \ \boldsymbol{I}_n$ for $i, j = 1, ..., m$ where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix.

The $n \times m$ matrix

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Y}_1 & \boldsymbol{Y}_2 & \dots & \boldsymbol{Y}_m \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_1^T \\ \vdots \\ \boldsymbol{y}_n^T \end{bmatrix}.$$

The $n \times p$ design matrix $\boldsymbol{X}$ of predictor variables is not necessarily of full rank p, and

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_n^T \end{bmatrix}$$

where often $\boldsymbol{v}_1 = \mathbf{1}$.

The $p \times m$ matrix

$$\boldsymbol{B} = \left[\begin{array}{cccc} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \ldots & \boldsymbol{\beta}_m \end{array} \right].$$

The $n \times m$ matrix

$$\boldsymbol{E} = \left[\begin{array}{cccc} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \ldots & \boldsymbol{e}_m \end{array} \right] = \left[\begin{array}{c} \boldsymbol{\epsilon}_1^T \\ \vdots \\ \boldsymbol{\epsilon}_n^T \end{array} \right].$$

Considering the ith row of $\boldsymbol{Z}, \boldsymbol{X}$, and $\boldsymbol{E}$ shows that $\boldsymbol{y}_i^T = \boldsymbol{x}_i^T \boldsymbol{B} + \boldsymbol{\epsilon}_i^T$.

The multivariate linear regression model and one way MANOVA model are special cases of the multivariate linear model, but using double subscripts will be useful for describing the one way MANOVA model. Suppose there are 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 ith treatment group and the jth case (often an individual or thing) in the group. Hence $i = 1, ..., p$ and $j = 1, ..., n_i$. The Y_{ijk} follow different one way ANOVA models for $k = 1, ..., m$. Assume $E(\boldsymbol{y}_{ij}) = \boldsymbol{\mu}_i$ and $\mathrm{Cov}(\boldsymbol{y}_{ij}) = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$. Hence the p treatments have different mean vectors $\boldsymbol{\mu}_i$, but common covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$.

The one way MANOVA is used to test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_p$. Often $\boldsymbol{\mu}_i = \boldsymbol{\mu} + \boldsymbol{\tau}_i$, so H_0 becomes $H_0 : \boldsymbol{\tau}_1 = \cdots = \boldsymbol{\tau}_p$. If $m = 1$, the one way MANOVA model is the one way ANOVA model. MANOVA is useful since it takes into account the correlations between the m response variables. The Hotelling's T^2 test that uses a common covariance matrix is a special case of the one way MANOVA model with $p = 2$.

4.1 AN ALTERNATIVE TO THE USUAL ONE WAY MANOVA

A useful one way MANOVA model is $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}$ where $\boldsymbol{X}$ is the full rank matrix where the first column of $\boldsymbol{X}$ is $\boldsymbol{v}_1 = \mathbf{1}$ and the ith column $\boldsymbol{v}_i$ of $\boldsymbol{X}$ is an indicator for group $i-1$ for $i = 2, ..., p$. For example, $\boldsymbol{v}_3 = (\mathbf{0}^T, \mathbf{1}^T, \mathbf{0}^T, ..., \mathbf{0}^T)^T$ where the p vectors in $\boldsymbol{v}_3$ have lengths $n_1, n_2, ..., n_p$, respectively. Then $\hat{\beta}_{1k} = \overline{Y}_{p0k} = \hat{\mu}_{pk}$ for $k = 1, ..., m$, and

$$\hat{\beta}_{ik} = \overline{Y}_{i-1,0k} - \overline{Y}_{p0k} = \hat{\mu}_{i-1,k} - \hat{\mu}_{pk}$$

for $k = 1, ..., m$ and $i = 2, ..., p$. Thus testing $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_p$ is equivalent to testing $H_0 : \boldsymbol{LB} = \boldsymbol{0}$ where $\boldsymbol{L} = [\boldsymbol{0}\ \boldsymbol{I}_{p-1}]$. Press (2005, p. 262) uses the above model. Then $\boldsymbol{y}_{ij} = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_{ij}$ and

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{\mu}_p^T \\ (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_p)^T \\ (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_p)^T \\ \vdots \\ (\boldsymbol{\mu}_{p-2} - \boldsymbol{\mu}_p)^T \\ (\boldsymbol{\mu}_{p-1} - \boldsymbol{\mu}_p)^T \end{bmatrix}.$$

Then a test statistic for the one way Manova model is $\boldsymbol{w}$ given by Equation (4.1) with $T_i = \hat{\boldsymbol{\mu}}_i = \overline{\boldsymbol{y}}_i$ where it is assumed that $\boldsymbol{\Sigma}_i \equiv \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ for $i = 1, ..., p$.

Large sample theory can be used to derive a better test that does not need the equal population covariance matrix assumption $\boldsymbol{\Sigma}_i \equiv \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$. 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 the $\boldsymbol{\mu}_i = \boldsymbol{\mu}$ and $\sqrt{n_i}(T_i - \boldsymbol{\mu}) \xrightarrow{D} N_m(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$, and $\sqrt{n}(T_i - \boldsymbol{\mu}) \xrightarrow{D} N_m\left(\boldsymbol{0}, \frac{\boldsymbol{\Sigma}_i}{\pi_i}\right)$. Let

$$\boldsymbol{w} = \begin{bmatrix} T_1 - T_p \\ T_2 - T_p \\ \vdots \\ T_{p-2} - T_p \\ T_{p-1} - T_p \end{bmatrix}. \tag{4.1}$$

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

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

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 ith diagonal block entry is $\frac{\hat{\boldsymbol{\Sigma}}_i}{n_i} + \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p}$ for $i = 1, ..., (p-1)$.

Reject H_0 if $t_0 > m(p-1)F_{m(p-1),d_n}(1-\alpha)$ 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 $n_i \geq 20m$ for $i = 1, ..., p$.

Olive (2017b, ch. 10) has the above result where $T_i = \overline{\boldsymbol{y}}_i$ is the sample mean and $\hat{\boldsymbol{\Sigma}}_i = \boldsymbol{S}_i$ is the sample covariance matrix of the ith group. Then $\boldsymbol{\Sigma}_i$ is the population covariance matrix of the ith group. The following theorem gives the general result.

Theorem 4.1. *If*

$$\begin{pmatrix} \sqrt{n_1}\,(T_1 - \boldsymbol{\mu}_1) \\ \vdots \\ \sqrt{n_p}\,(T_p - \boldsymbol{\mu}_p) \end{pmatrix} \xrightarrow{D} N_{mp}\left[\begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{\Sigma}_p \end{pmatrix} \right],$$

then under $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_p$

$$\sqrt{n}\begin{pmatrix} T_1 - T_p \\ \vdots \\ T_{p-1} - T_p \end{pmatrix} \xrightarrow{D} N_{(m-1)p}\left[\begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \frac{\boldsymbol{\Sigma}_1}{\pi_1} + \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \cdots & \frac{\boldsymbol{\Sigma}_p}{\pi_p} \\ \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \frac{\boldsymbol{\Sigma}_2}{\pi_2} + \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \cdots & \frac{\boldsymbol{\Sigma}_p}{\pi_p} \\ \vdots & & \ddots & & \vdots \\ \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \cdots & \frac{\boldsymbol{\Sigma}_{p-1}}{\pi_{p-1}} + \frac{\boldsymbol{\Sigma}_p}{\pi_p} \end{pmatrix} \right].$$

Proof. To simplify large sample theory, assume $n_i = \pi_i n$ for some positive real π_i and $i = 1, \cdots, p$. Let $\hat{\boldsymbol{\Sigma}}_i$ be a consistent nonsingular estimator of $\boldsymbol{\Sigma}_i$. Then

$$\begin{pmatrix} \sqrt{n_1}\,(T_1 - \boldsymbol{\mu}_1) \\ \vdots \\ \sqrt{n_p}\,(T_p - \boldsymbol{\mu}_p) \end{pmatrix} \xrightarrow{D} N_{mp}\left[\begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{\Sigma}_p \end{pmatrix}\right],$$

Under $H_0 : \boldsymbol{\mu}_1 = \cdots \boldsymbol{\mu}_p = \boldsymbol{\mu}$

$$\begin{pmatrix} \sqrt{n}\,(T_1 - \boldsymbol{\mu}) \\ \vdots \\ \sqrt{n}\,(T_p - \boldsymbol{\mu}) \end{pmatrix} \xrightarrow{D} N_{mp}\left[\begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \frac{\boldsymbol{\Sigma}_1}{\pi_1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \frac{\boldsymbol{\Sigma}_p}{\pi_p} \end{pmatrix}\right] = N_{mp}(\mathbf{0}, \boldsymbol{\Sigma}).$$

Let A be a $(m-1)p \times mp$ matrix and

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{I} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & -\boldsymbol{I} \\ \mathbf{0} & \boldsymbol{I} & \mathbf{0} & \cdots & \mathbf{0} & -\boldsymbol{I} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{I} & \mathbf{0} & \cdots & -\boldsymbol{I} \\ \vdots & & & \ddots & & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{I} & -\boldsymbol{I} \end{pmatrix}.$$

Then,

$$\boldsymbol{A}\begin{pmatrix} \sqrt{n}\,(T_1 - \boldsymbol{\mu}) \\ \vdots \\ \sqrt{n}\,(T_p - \boldsymbol{\mu}) \end{pmatrix} = \sqrt{n}\begin{pmatrix} T_1 - T_p \\ \vdots \\ T_{p-1} - T_p \end{pmatrix}$$

and

$$\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T = \begin{pmatrix} \frac{\boldsymbol{\Sigma}_1}{\pi_1} + \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \cdots & \frac{\boldsymbol{\Sigma}_p}{\pi_p} \\ \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \frac{\boldsymbol{\Sigma}_2}{\pi_2} + \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \cdots & \frac{\boldsymbol{\Sigma}_p}{\pi_p} \\ \vdots & & \ddots & & \vdots \\ \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \cdots & \frac{\boldsymbol{\Sigma}_{p-1}}{k_{p-1}} + \frac{\boldsymbol{\Sigma}_p}{\pi_p} \end{pmatrix}.$$

Therefore,

$$\sqrt{n}\begin{pmatrix} T_1 - T_p \\ \vdots \\ T_{p-1} - T_p \end{pmatrix} \xrightarrow{D} N_{(m-1)p}\left[\begin{pmatrix} \mathbf{0} \\ \vdots \\ \vdots \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \frac{\boldsymbol{\Sigma}_1}{\pi_1} + \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \cdots & \frac{\boldsymbol{\Sigma}_p}{\pi_p} \\ \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \frac{\boldsymbol{\Sigma}_2}{\pi_2} + \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \cdots & \frac{\boldsymbol{\Sigma}_p}{\pi_p} \\ \vdots & & \ddots & & \vdots \\ \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \frac{\boldsymbol{\Sigma}_p}{\pi_p} & \cdots & \frac{\boldsymbol{\Sigma}_{p-1}}{\pi_{p-1}} + \frac{\boldsymbol{\Sigma}_p}{\pi_p} \end{pmatrix}\right]$$

□

If $\boldsymbol{T} = (T_1^T, T_2^T, ..., T_P^T)^T$, $\boldsymbol{\theta} = (\boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T, ..., \boldsymbol{\mu}_p^T)^T$, $\boldsymbol{c}$ 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 $H_0 : \boldsymbol{A\theta} = \boldsymbol{c}$ versus $H_1 : \boldsymbol{A\theta} \neq \boldsymbol{c}$ uses

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

When H_0 is true, the statistic

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

The same statistic was used by Zhang and Liu (2013, p. 138) with $T_i = \overline{\boldsymbol{y}}_i$ and $\hat{\boldsymbol{\Sigma}}_i = \boldsymbol{S}_i$.

4.1.1 Test H_0 when $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}}$ is unknown or difficult to estimate.

Since the common covariance matrix assumption $\text{Cov}(\boldsymbol{\epsilon}_k) = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ for $k = 1, ..., p$ is extremely strong, using the prediction region method to test $H_0 : \boldsymbol{LB} = \mathbf{0}$ may be a useful alternative. Take a sample of size n_k with replacement from the n_k cases for each group for $k = 1, 2, ..., p$. Let $\hat{\boldsymbol{B}}_i^*$ be the ith bootstrap estimator of $\boldsymbol{B}$ for $i = 1, ..., B$. Let the $(p-1)m \times 1$ vector $\boldsymbol{w}_i = vec(\boldsymbol{L}\hat{\boldsymbol{B}}_i^*) = ((\hat{\boldsymbol{\mu}}_1^* - \hat{\boldsymbol{\mu}}_p^*)^T, ..., (\hat{\boldsymbol{\mu}}_{p-1}^* - \hat{\boldsymbol{\mu}}_p^*)^T)^T$ for $i = 1, ..., B$, where $vec(\boldsymbol{A})$ stacks columns of a matrix into a vector. For a robust test use $\boldsymbol{w}_i = ((T_1^* - T_p^*)^T, ..., (T_{p-1}^* - T_p^*)^T)^T$ where T_k is a robust location estimator, such as the coordinatewise median or trimmed mean, applied to the cases in the kth treatment group. The prediction region method fails to reject H_0 if $\mathbf{0}$ is in the resulting confidence region. We likely need $n \geq 40mp$, $n \geq (m+p)^2$, and $n_i \geq 40m$.

4.2 POWER COMPARISON AMONG THE TESTS

Figures 4.1, 4.2 and 4.3 try to compare the powers among the tests mentioned above with the classical test. Here $\delta_1 \in \{0.00, 0.04, 0.08, 0.12, 0.16, 0.20, 0.24, 0.30, ...\}$, $\delta_2 = 2 \times \delta_1$ and $\delta_3 = 3 \times \delta_1$. Group i has mean $\boldsymbol{\mu}_i = \delta_i \mathbf{1}$. When δ_1 increases, the distance between the mean vectors increases. Figure 4.1 shows the power curve for clean MVN data with a balanced design where the groups have the same covariance matrices while figure 4.2 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 4.3 uses a mixture distribution. Figure 4.4 is similar to 4.2 except it uses a multivariate t_4 distribution. See the actual simulation results in Chapter 6.

4.3 REAL DATA EXAMPLE

The North Carolina Crime data consists of 630 observations on 24 variables. This data set is available online at https://vincentarelbundock.github.io/Rdatasets/datasets.html. Region is a categorical variable with three categories: Central, West and Other with the number of observations 232, 146 and 245 respectively, and forms the three groups. This example uses "wsta" - weekly wage of state employees, "avgsen" - average sentence days, "prbarr" - 'probability' of arrest,"prbconv" - 'probability' of conviction and "taxpc" - tax revenue per capita as variables. The test with the coordinatewise median had $D_0 = 4.086$ with the cutoff of 4.32 and failed to reject H_0. The classical one-way MANOVA test had a p-value of 0.001 and rejected the null hypothesis.

The DD plots in figure 4.5 reveal a few outliers. Furthermore the boxplots in figure 4.6 and the scatterplot matrix in figure 4.7 shows that the data are highly skewed. Hence the location measures other than the median likely do differ.

See the simulation set up and the simulation results in Chapter 6.

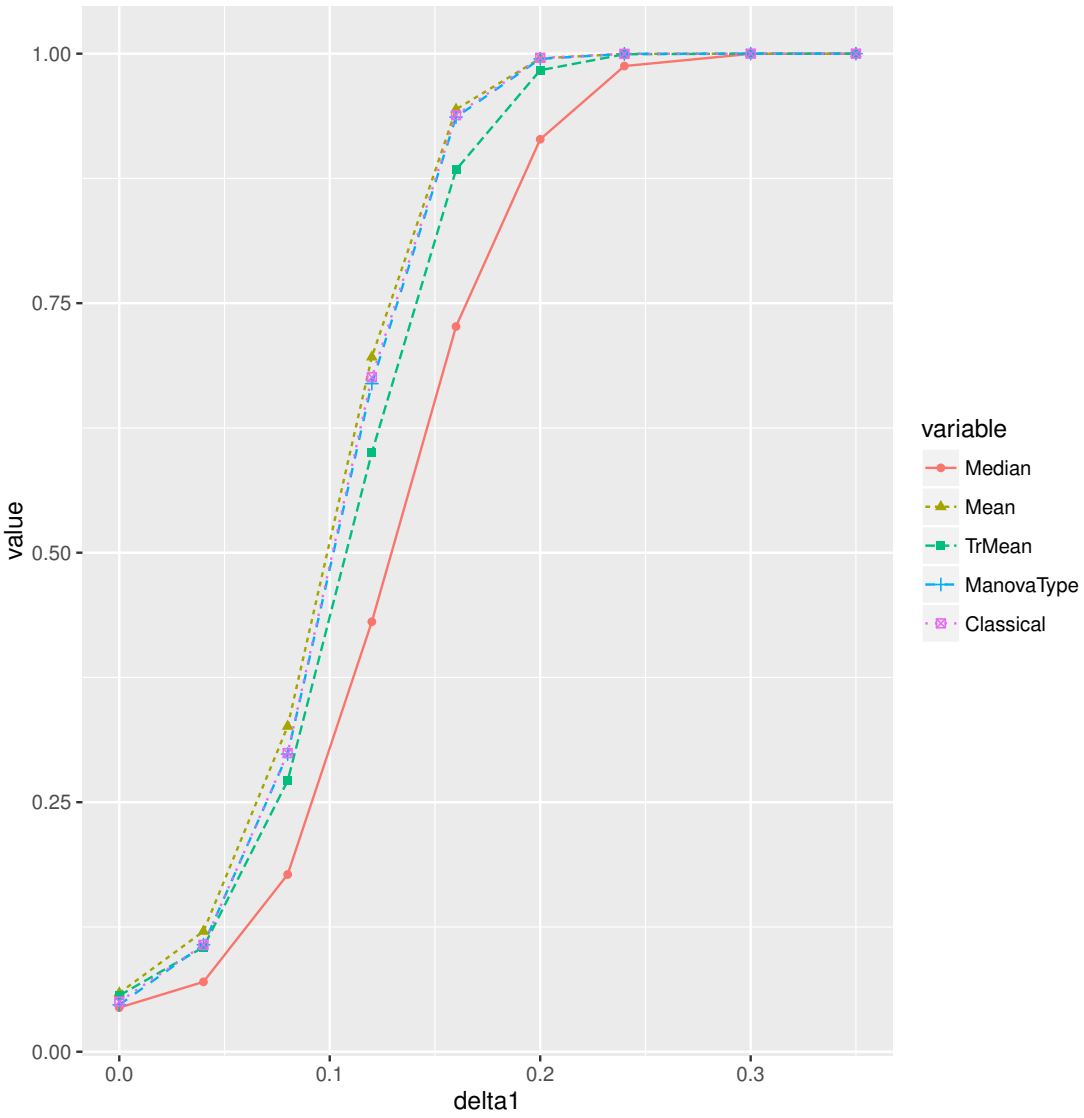

Figure 4.1. Power curve for clean MVN data with $m = 5, \sigma_1 = 1, \sigma_2 = 1, \sigma_3 = 1, n_1 = 200, n_2 = 200$ and $n_3 = 200$

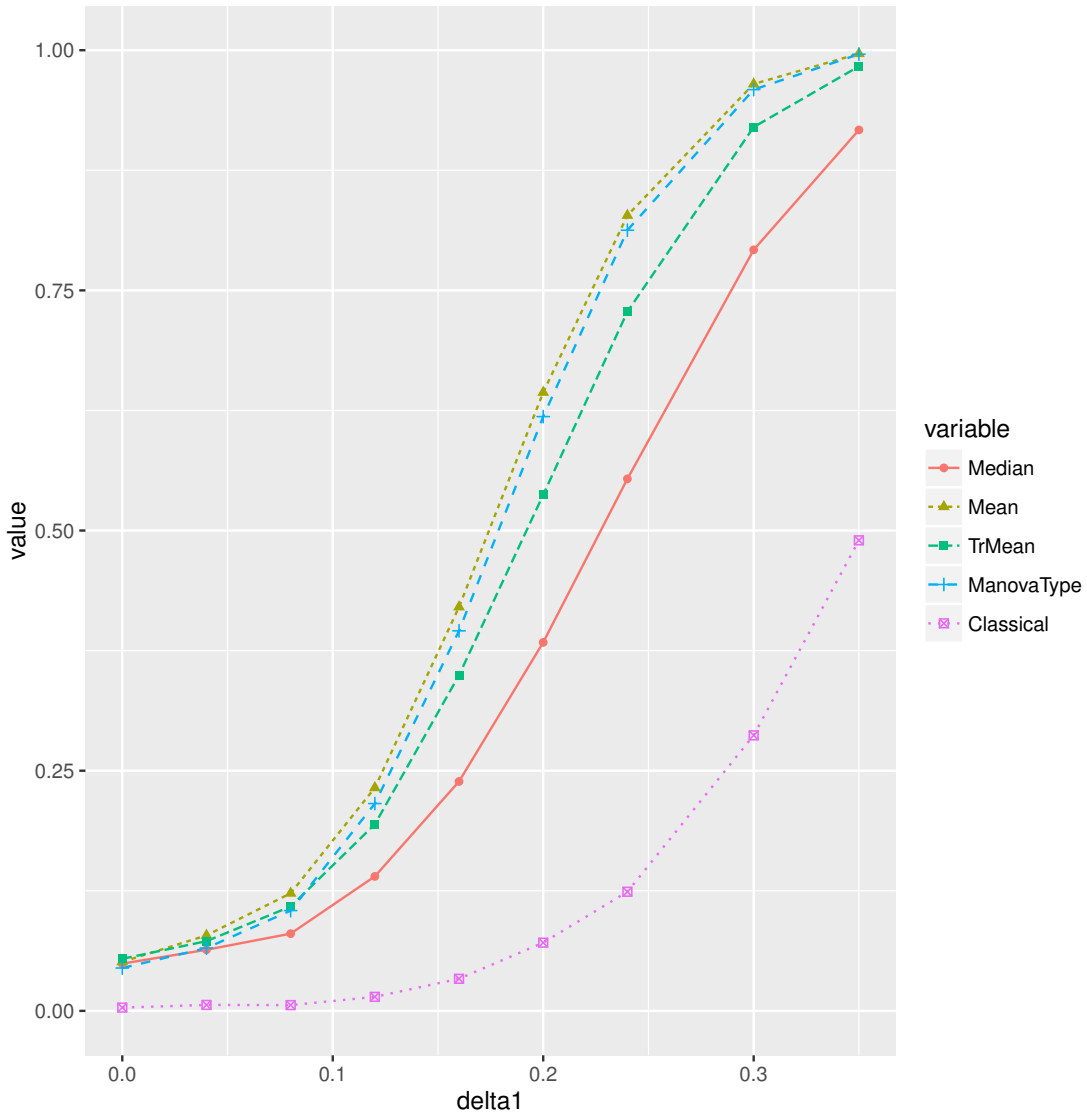

Figure 4.2. Power curve for clean MVN data with $m = 5, \sigma_1 = 1, \sigma_2 = 2, \sigma_3 = 5, n_1 = 200, n_2 = 400$ and $n_3 = 600$

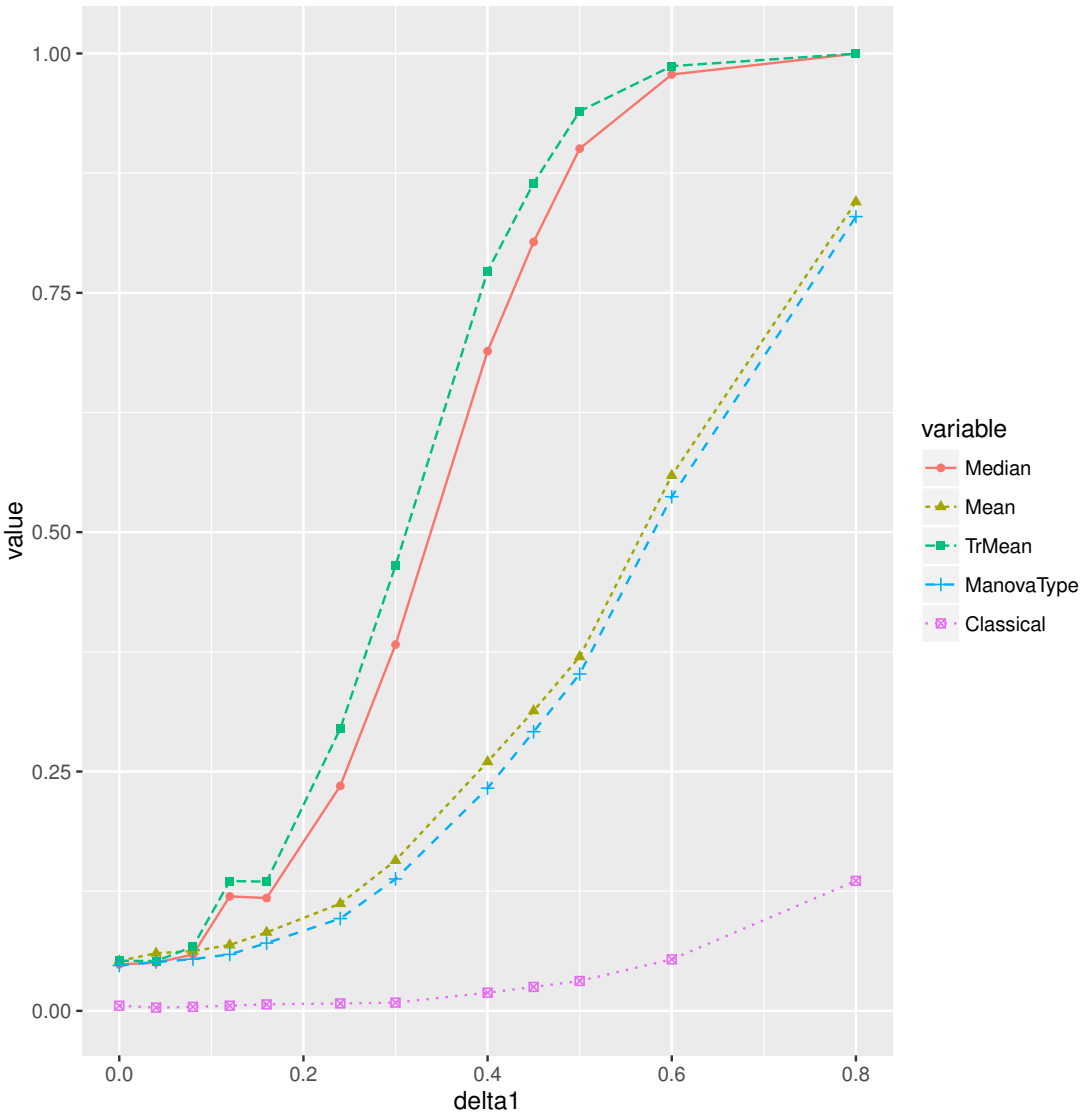

Figure 4.3. Power curve for clean Mixture data with $m = 5, \sigma_1 = 1, \sigma_2 = 2, \sigma_3 = 5, n_1 = 200, n_2 = 400$ and $n_3 = 600$

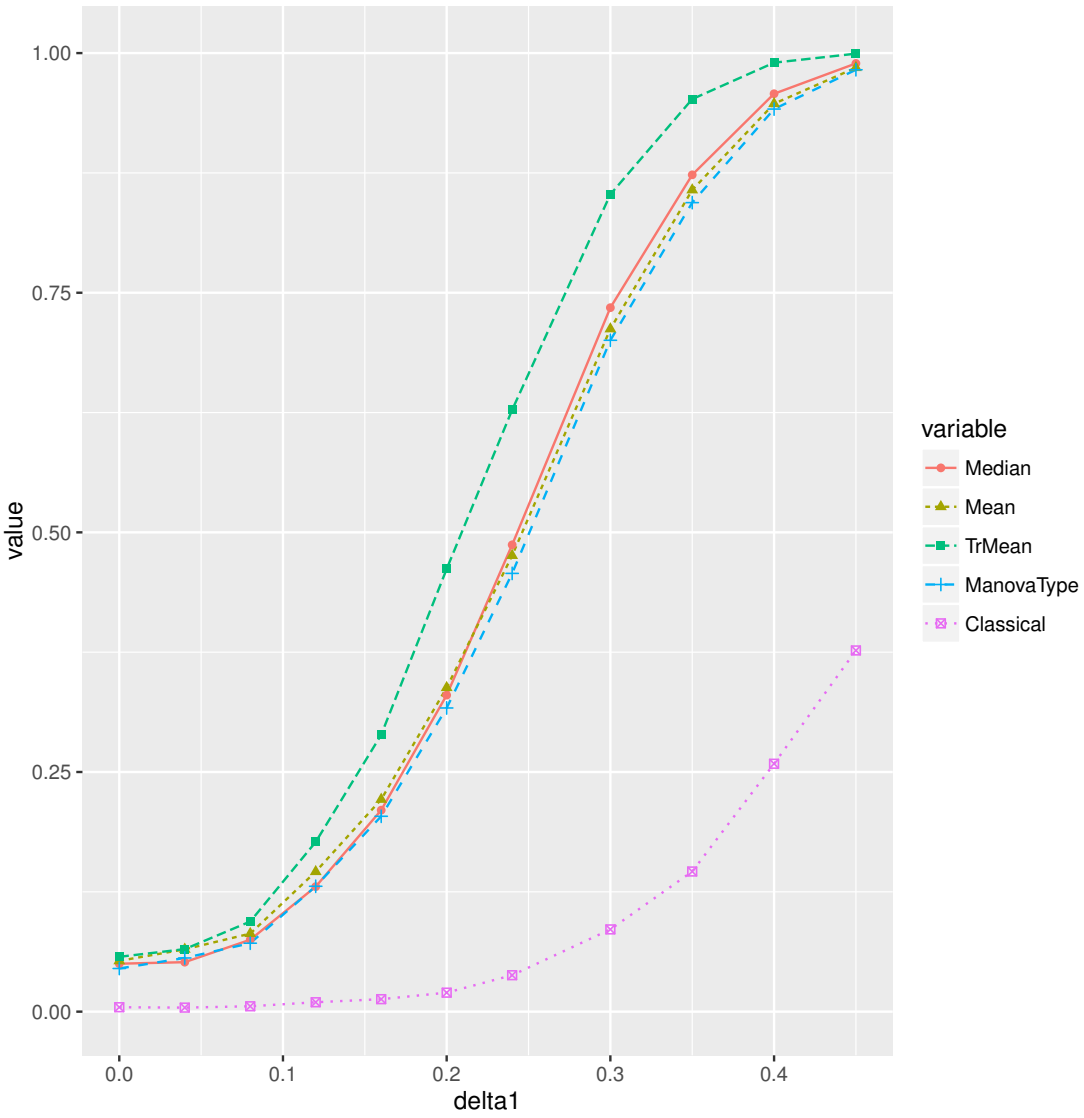

Figure 4.4. Power curve for clean multivariate t_4 data with $m = 5, \sigma_1 = 1, \sigma_2 = 2, \sigma_3 = 5, n_1 = 200, n_2 = 400$ and $n_3 = 600$

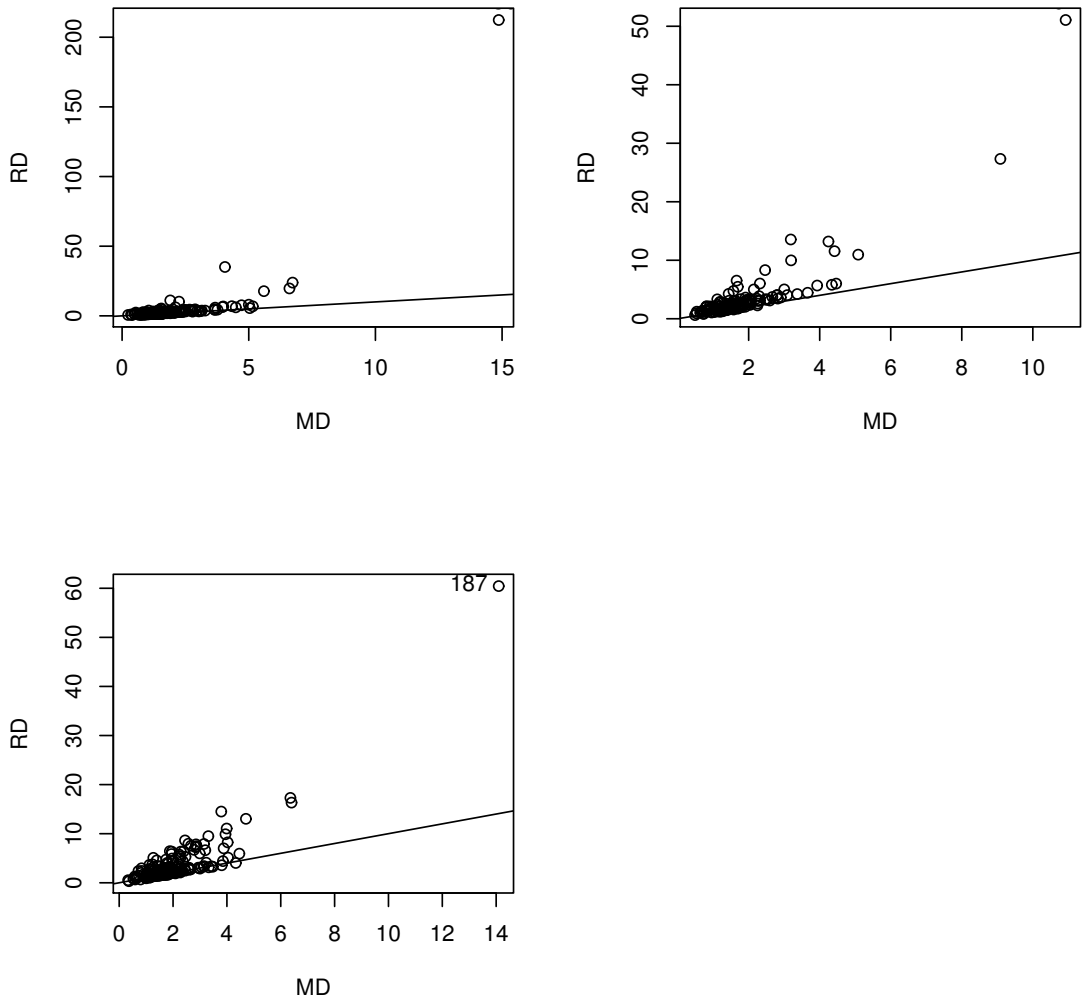

Figure 4.5. DD plots for Crime data

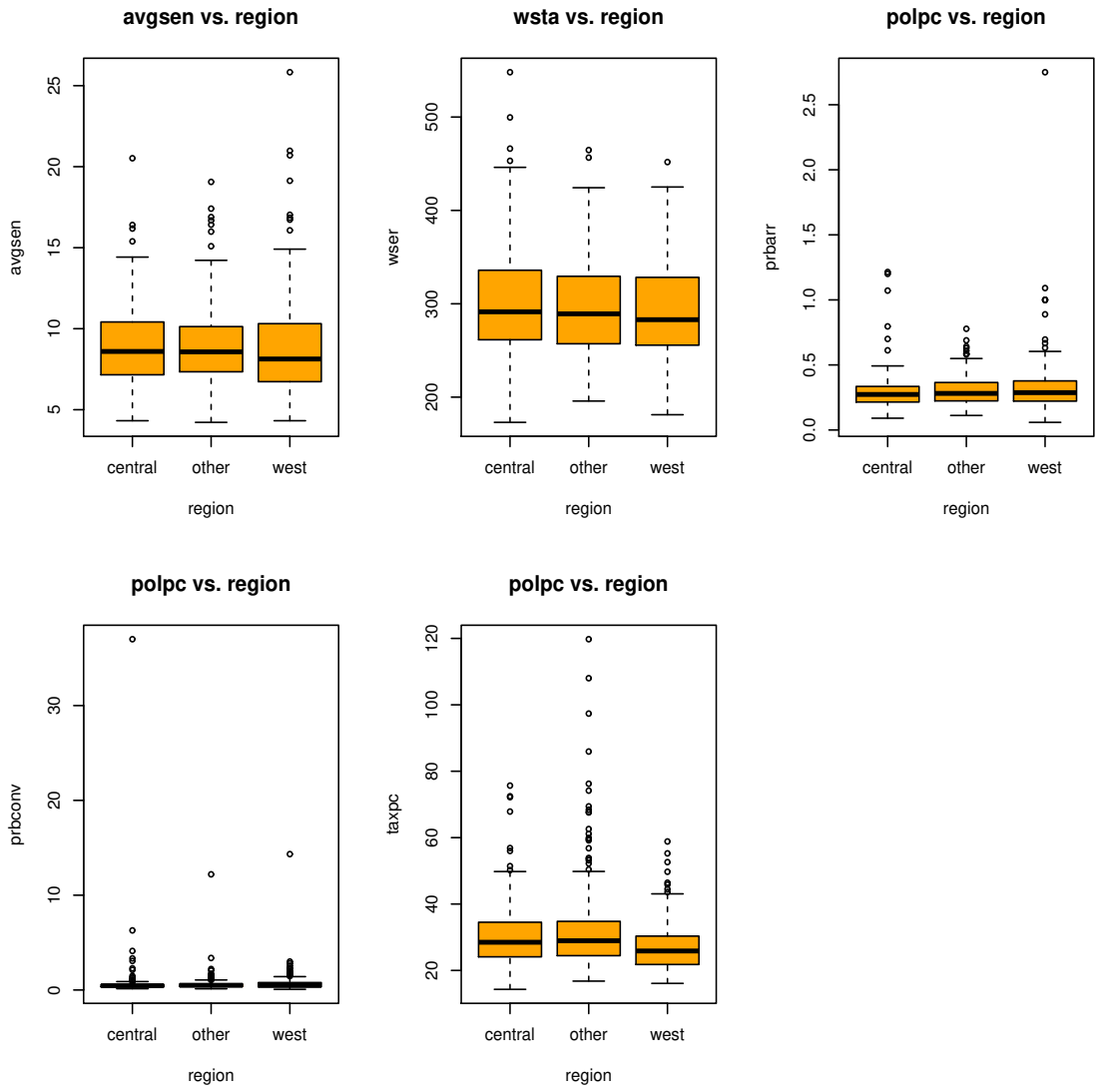

Figure 4.6. Side by side boxplots for Crime data

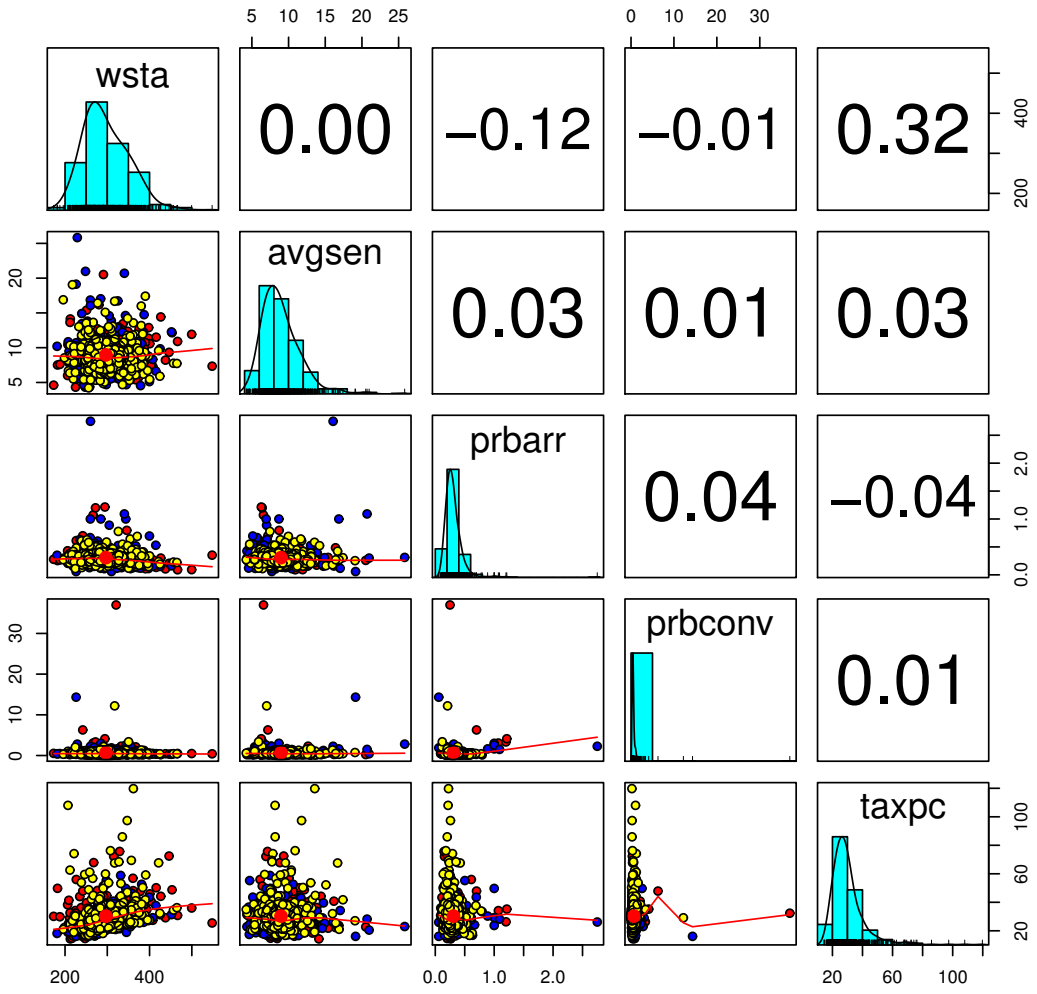

Figure 4.7. Scatterplot matrix for Crime data

CHAPTER 5
SIMULATIONS FOR BOOTSTRAPPING ANALOGS OF THE TWO SAMPLE HOTELLING'S T^2 TEST

5.1 SIMULATION SETUP

The simulation used 5000 runs with B bootstrap samples. Olive (2017bc) suggests that the prediction region method can give good results when the number of bootstrap samples $B \geq 50p$ and if $n \geq 50p$, and the simulation used various values of B. See Rupasinghe Arachchige Don and Pelawa Watagoda (2017).

Four types of data distributions $\boldsymbol{w}_i$ were considered that were identical for $i = 1, 2$. Then $\boldsymbol{x}_1 = \boldsymbol{A}\boldsymbol{w}_1 + \delta\mathbf{1}$ and $\boldsymbol{x}_2 = \sigma\boldsymbol{A}\boldsymbol{w}_2$ where $\mathbf{1} = (1, .., 1)^T$ is a vector of ones and $\boldsymbol{A} = diag(1, \sqrt{2}, ..., \sqrt{p})$. The $\boldsymbol{w}_i$ distributions were:

1. multivariate normal distribution $N_p(\mathbf{0}, \boldsymbol{I})$,
2. multivariate t distribution with 4 degrees of freedom,
3. mixture distribution $0.6N_p(\mathbf{0}, \boldsymbol{I}) + 0.4N_p(\mathbf{0}, 25\boldsymbol{I})$,
4. multivariate lognormal distribution shifted to have nonzero mean $\boldsymbol{\mu} = 0.649\ \mathbf{1}$, but a population coordiatewise median of $\mathbf{0}$.

Note that $\text{Cov}(\boldsymbol{x}_2) = \sigma^2\ \text{Cov}(\boldsymbol{x}_1)$, and for the first three distributions, $E(\boldsymbol{x}_i) = E(\boldsymbol{w}_i) = \mathbf{0}$ if $\delta = 0$.

Adding the same type and proportion of outliers to groups one and two often resulted in two distributions that were still similar. Hence outliers were added to the first group but not the second, making the covariance structures of the two groups quite different. The outlier proportion was $100\gamma\%$. Let $\boldsymbol{x}_1 = (x_{11}, ..., x_{p1})^T$. The five outlier types for group 1 were:

1. type 1: a tight cluster at the major axis $(0, ..., 0, z)^T$,
2. type 2: a tight cluster at the minor axis $(z, 0, ..., 0)^T$,

3. type 3: a mean shift $N((z, ..., z)^T, diag(1, ..., p))$,

4. type 4: x_{1p} replaced by z, and

5. type 5: x_{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.

5.2 SIMULATION OUTPUT

5.2.1 Type I error rates simulation for clean data

Tables 5.1 - 5.6 were for clean elliptically contoured distributions (no outliers present), where H_0 is true and the different location estimators estimate $\boldsymbol{\mu} = \mathbf{0}$, the point of symmetry for the distribution. The chi–square cutoffs when $p = 5$ and $p = 15$ were 11.071 and 24.996, respectively. The *coverages* were often near the nominal value of 0.05, but the RMVN *coverages* were a bit low except for Table 5.4. The classical Hotelling's T^2 test does not use the bootstrap, and performed poorly when H_0 was true and both the sample sizes and the population covariance matrices were different.

For clean multivariate lognormal data, H_0 is true when $\sigma = 1$ (identical distributions for both groups), but H_0 is not true for the population mean when $\sigma = 2$. For $\sigma = 2$, the coordinatewise median had *coverages* near the nominal, while the sample mean had good power with *coverages* near 1. The RMVN coverage was a bit low when $\sigma = 1$ with power that was often less than that of the sample mean when $\sigma = 2$. See Table 5.7 and 5.8. The simulated cutoffs were quite similar to the chi-square cutoffs for Tables 5.1 through 5.8.

Table 5.1. *Coverages* for clean multivariate normal data $p = 5$

p	n_1	n_2	σ	B	Median	Mean	Tr.Mn	RMVN	Class
5	100	100	1	250	0.0418	0.0604	0.0546	0.0172	
				1000	0.0452	0.0678	0.0594	0.0198	
			2	250	0.0502	0.0706	0.0596	0.0258	
				1000	0.0470	0.0684	0.0638	0.0220	
	250	250	1	250	0.0470	0.0554	0.0568	0.0402	0.0560
				1000	0.0440	0.0606	0.0540	0.0414	
			2	250	0.0472	0.0550	0.0574	0.0422	0.0498
				1000	0.0420	0.0568	0.0538	0.0392	
	100	200	1	250	0.0446	0.0670	0.0600	0.0228	
				1000	0.0434	0.0614	0.0582	0.0254	
			2	250	0.0488	0.0610	0.0568	0.0292	
				1000	0.0422	0.0518	0.0532	0.0234	
	250	500	1	250	0.0490	0.0524	0.0496	0.0394	0.0552
				1000	0.0462	0.0588	0.0584	0.0448	
			2	250	0.0460	0.0540	0.0524	0.0436	0.0070
				1000	0.0470	0.0500	0.0534	0.0386	

Table 5.2. *Coverages* for clean multivariate normal data $p = 15$

p	n_1	n_2	σ	B	Median	Mean	Tr.Mn	RMVN	Class
15	300	300	1	750	0.0454	0.0666	0.0622	0.0234	
				1000	0.0378	0.0578	0.0554	0.0256	
			2	750	0.0484	0.0752	0.0674	0.0270	
				1000	0.0576	0.0730	0.0732	0.0296	
	750	750	1	750	0.0462	0.0626	0.0622	0.0466	0.0450
				1000	0.0390	0.0514	0.0470	0.0378	
			2	750	0.0492	0.0598	0.0608	0.0464	0.0516
				1000	0.0474	0.0556	0.0568	0.0446	
	300	600	1	750	0.0424	0.0650	0.0658	0.0286	
				1000	0.0440	0.0638	0.0592	0.0308	
			2	750	0.0438	0.0578	0.0576	0.0376	
				1000	0.0502	0.0620	0.0630	0.0348	
	750	1500	1	750	0.0466	0.0538	0.0550	0.0466	0.0480
				1000	0.0492	0.0556	0.0548	0.0444	
			2	750	0.0424	0.0538	0.0520	0.0454	0.0014
				1000	0.0514	0.0532	0.0542	0.0426	

Table 5.3. *Coverages* for clean $0.6N_p(\mathbf{0}, \boldsymbol{I}) + 0.4N_p(\mathbf{0}, 25\boldsymbol{I})$ data $p = 5$

p	n_1	n_2	σ	B	Median	Mean	Tr.Mn	RMVN	Class
5	100	100	1	250	0.0294	0.0620	0.0388	0.0158	
				1000	0.0390	0.0544	0.0420	0.0130	
			2	250	0.0400	0.0606	0.0416	0.0184	
				1000	0.0422	0.0612	0.0386	0.0162	
	250	250	1	250	0.0420	0.0560	0.0480	0.0394	0.0462
				1000	0.0386	0.0532	0.0464	0.0336	
			2	250	0.0454	0.0550	0.0476	0.0416	0.0476
				1000	0.0370	0.0484	0.0400	0.0368	
	100	200	1	250	0.0364	0.0546	0.0398	0.0190	
				1000	0.0344	0.0632	0.0394	0.0222	
			2	250	0.0372	0.0604	0.0462	0.0238	
				1000	0.0346	0.0616	0.0402	0.0228	
	250	500	1	250	0.0460	0.0542	0.0538	0.0416	0.0470
				1000	0.0368	0.0502	0.0416	0.0404	
			2	250	0.0480	0.0600	0.0474	0.0390	0.0060
				1000	0.0416	0.0598	0.0498	0.0416	

Table 5.4. *Coverages* for clean $0.6N_p(\mathbf{0}, \boldsymbol{I}) + 0.4N_p(\mathbf{0}, 25\boldsymbol{I})$ data $p = 15$

p	n_1	n_2	σ	B	Median	Mean	Tr.Mn	RMVN	Class
15	300	300	1	750	0.0414	0.0598	0.0490	0.0428	
				1000	0.0402	0.0592	0.0484	0.0414	
			2	750	0.0426	0.0620	0.0496	0.0502	
				1000	0.0414	0.0600	0.0448	0.0496	
	750	750	1	750	0.0434	0.0536	0.0540	0.0448	0.0496
				1000	0.0406	0.0598	0.0474	0.0396	
			2	750	0.0468	0.0626	0.0518	0.0456	0.0464
				1000	0.0456	0.0566	0.0490	0.0454	
	300	600	1	750	0.0418	0.0582	0.0464	0.0474	
				1000	0.0430	0.0684	0.0514	0.0466	
			2	750	0.0394	0.0578	0.0466	0.0432	
				1000	0.0356	0.0606	0.0470	0.0422	
	750	1500	1	750	0.0456	0.0584	0.0568	0.0488	0.0502
				1000	0.0426	0.0550	0.0478	0.0438	
			2	750	0.0456	0.0576	0.0508	0.0442	0.0004
				1000	0.0416	0.0572	0.0488	0.0510	

Table 5.5. *Coverages* for clean multivariate t_4 data $p = 5$

p	n_1	n_2	σ	B	Median	Mean	Tr.Mn	RMVN	Class
5	100	100	1	250	0.0478	0.0610	0.0592	0.0178	
				1000	0.0358	0.0548	0.0518	0.0164	
			2	250	0.0514	0.0608	0.0632	0.0238	
				1000	0.0444	0.0512	0.0558	0.0162	
	250	250	1	250	0.0442	0.0574	0.0570	0.0266	0.0456
				1000	0.0426	0.0570	0.0530	0.0282	
			2	250	0.0496	0.0618	0.0614	0.0328	0.0542
				1000	0.0480	0.0558	0.0578	0.0292	
	100	200	1	250	0.0432	0.0556	0.0576	0.0212	
				1000	0.0372	0.0552	0.0522	0.0200	
			2	250	0.0414	0.0586	0.0570	0.0232	
				1000	0.0446	0.0546	0.0568	0.0262	
	250	500	1	250	0.0484	0.0512	0.0540	0.0346	0.0504
				1000	0.0420	0.0488	0.0494	0.0310	
			2	250	0.0408	0.0580	0.0526	0.0348	0.0058
				1000	0.0410	0.0492	0.0510	0.0348	

Table 5.6. *Coverages* for clean multivariate t_4 data $p = 15$

p	n_1	n_2	σ	B	Median	Mean	Tr.Mn	RMVN	Class
15	300	300	1	750	0.0392	0.0546	0.0562	0.0158	
				1000	0.0480	0.0590	0.0662	0.0140	
			2	750	0.0478	0.0572	0.0604	0.0134	
				1000	0.0512	0.0632	0.0640	0.0148	
	750	750	1	750	0.0470	0.0550	0.0562	0.0232	0.0414
				1000	0.0382	0.0526	0.0476	0.0228	
			2	750	0.0472	0.0572	0.0542	0.0248	0.0442
				1000	0.0502	0.0496	0.0556	0.0258	
	300	600	1	750	0.0448	0.0554	0.0598	0.0158	
				1000	0.0458	0.0602	0.0616	0.0184	
			2	750	0.0450	0.0564	0.0558	0.0178	
				1000	0.0400	0.0498	0.0546	0.0196	
	750	1500	1	750	0.0482	0.0556	0.0528	0.0224	0.0446
				1000	0.0464	0.0496	0.0528	0.0254	
			2	750	0.0442	0.0534	0.0502	0.0314	0.0016
				1000	0.0452	0.0508	0.0554	0.0262	

Table 5.7. *Coverages* for clean lognormal data $p = 5$

p	n_1	n_2	σ	B	Median	Mean	Tr.Mn	RMVN	Class
5	100	100	1	250	0.0330	0.0462	0.0470	0.0096	
				1000	0.0290	0.0508	0.0390	0.0088	
			2	250	0.0442	0.6170	0.0600	0.0340	
				1000	0.0368	0.6200	0.0570	0.0352	
	250	250	1	250	0.0408	0.0460	0.0514	0.0274	0.0470
				1000	0.0388	0.0494	0.0474	0.0254	
			2	250	0.0436	0.9816	0.0858	0.1108	0.9968
				1000	0.0398	0.9846	0.0788	0.1168	
	100	200	1	250	0.0346	0.0684	0.0492	0.0204	
				1000	0.0320	0.0548	0.0434	0.0138	
			2	250	0.0400	0.8898	0.0644	0.0692	
				1000	0.0428	0.8962	0.0700	0.0730	
	250	500	1	250	0.0398	0.0540	0.0496	0.0316	0.0472
				1000	0.0368	0.0588	0.0446	0.0292	
			2	250	0.0418	0.9998	0.1192	0.2492	0.9964
				1000	0.0424	0.9994	0.1158	0.2520	

Table 5.8. *Coverages* for clean lognormal data $p = 15$

p	n_1	n_2	σ	B	Median	Mean	Tr.Mn	RMVN	Class
15	300	300	1	750	0.0326	0.0546	0.0478	0.0106	
				1000	0.0364	0.0558	0.0502	0.0120	
			2	750	0.0446	1.0000	0.1408	0.8530	
				1000	0.0474	1.0000	0.1450	0.8680	
	750	750	1	750	0.0402	0.0506	0.0480	0.0216	0.0502
				1000	0.0410	0.0444	0.0490	0.0238	
			2	750	0.0506	1.0000	0.3670	1.0000	1.0000
				1000	0.0510	1.0000	0.3748	1.0000	
	300	600	1	750	0.0422	0.0684	0.0546	0.0188	
				1000	0.0406	0.0736	0.0560	0.0172	
			2	750	0.0396	1.0000	0.2344	0.9984	
				1000	0.0408	1.0000	0.2402	0.9990	
	750	1500	1	750	0.0420	0.0580	0.0514	0.0258	0.0514
				1000	0.0478	0.0558	0.0608	0.0284	
			2	750	0.0446	1.0000	0.6110	1.0000	1.0000
				1000	0.0464	1.0000	0.6256	1.0000	

5.2.2 Type I error rates simulation for contaminated data

Table 5.9 illustrates the simulated results where group 1 had outliers. The coordinatewise median worked with a little higher type I error rate (around 0.08) than the nominal level of 0.05 for the mixture, multivariate t, and multivariate log normal distributions, but failed for the multivariate normal data when $\gamma = 0.4$. The sample mean (classical and bootstrap) and 25% trimmed mean failed to achieve the nominal level with any of the distributions used when H_0 was true for the clean data. The RMVN estimator worked with all four distributions with a better type I error rate compared to the other estimators. The chi–square cutoff was 9.488 since $p = 4$.

The coordinatewise median can achieve better coverages for smaller proportions of outliers with higher values of z (not shown in the tables), i.e. the outliers had to be far from the clean data compared to the RMVN estimator. The RMVN estimator can handle higher proportions of outliers as shown in the Table 5.9.

Table 5.9. *Coverages* and cutoffs with outliers: $p = 4, n_1 = n_2 = 200, B = 200$

Dist.	Otype	γ	z		Med	Mean	Tr.Me	RMVN	Class
MVN	1	0.4	10	Cov	0.6946	1.0000	1.0000	0.0330	1.0000
				cut	10.158	9.769	9.798	10.701	
	2	0.4	20	Cov	0.5232	1.0000	1.0000	0.0382	1.0000
				cut	9.836	9.776	9.809	9.268	
	3	0.4	20	Cov	0.8578	1.0000	1.0000	0.0402	1.0000
				cut	10.214	9.761	9.760	9.288	
	4	0.1	10	Cov	0.0980	0.8654	0.1450	0.0382	0.8684
				cut	9.898	9.771	9.777	9.851	
Mix	2	0.4	20	Cov	0.0828	1.0000	1.0000	0.0144	1.0000
				cut	10.542	9.788	9.878	11.300	
	5	0.1	10	Cov	0.0820	0.5306	0.1228	0.0184	0.5276
				cut	9.933	9.779	9.881	11.056	
MVT	1	0.4	10	Cov	0.0854	0.6700	0.1548	0.0204	1.0000
				cut	10.232	9.799	9.787	10.200	
	5	0.1	20	Cov	0.0864	1.0000	0.1418	0.0304	1.0000
				cut	9.924	9.795	9.795	9.830	
Log	3	0.4	20	Cov	0.0778	1.0000	1.0000	0.0162	1.0000
				cut	13.689	9.822	9.827	12.607	
	4	0.1	10	Cov	0.0842	0.3158	0.1482	0.0234	0.3044
				cut	10.013	9.875	9.872	10.416	

5.2.3 Power Simulation

In the power simulation, $\delta > 0$ was used. Hence for the first three distributions $\boldsymbol{\mu}_2 = \mathbf{0}$ and $\boldsymbol{\mu}_1 = \delta(1, ..., 1)^T$. Then the Euclidean distance between the two means was $\sqrt{p}\delta$, where p is the number of parameters. Therefore the distance increases as p increase. The value of δ had to be fairly small so that the simulated power was not always 1. Also see Table 5.8 with $\sigma = 2$.

For Table 5.10, the sample mean (bootstrap and classical) had the best power while the sample median had the worst power. For Table 5.11, the RMVN estimator had the best power while the sample mean has the worst power. The trimmed mean had the best power for Table 5.12. For Table 5.13, the RMVN estimator had poor power when $p = 5$, $n = 250$, and $\sigma = 2$. No method was always best or worst.

Table 5.10. *Coverages* when H_0 is false for MVN data.

p	$n_1 = n_2$	σ	B	δ	Med	Mean	Tr.Me	RMVN	Class
5	250	1	250	0.35	0.9598	0.9990	0.9928	0.9942	0.9988
			1000	0.35	0.9684	0.9994	0.9970	0.9978	
		2	250	0.35	0.5958	0.8442	0.7672	0.7604	0.8402
			1000	0.35	0.5832	0.8346	0.7438	0.7470	
15	750	1	750	0.15	0.7394	0.9552	0.9012	0.9268	0.9556
			1000	0.15	0.7474	0.9522	0.8984	0.9178	
		2	750	0.15	0.3078	0.5318	0.4550	0.4468	0.5156
			1000	0.15	0.3118	0.5218	0.4430	0.4464	

Table 5.11. *Coverages* when H_0 is false for mixture data.

p	$n_1 = n_2$	σ	B	δ	Med	Mean	Tr.Me	RMVN	Class
5	250	1	250	0.45	0.8826	0.4062	0.9304	0.9938	0.4032
			1000	0.45	0.8858	0.4058	0.9338	0.9948	
		2	250	0.45	0.4458	0.1910	0.5222	0.7454	0.1642
			1000	0.45	0.4656	0.1890	0.5386	0.7626	
15	750	1	750	0.20	0.6204	0.2274	0.7148	0.9492	0.2114
			1000	0.20	0.6316	0.2228	0.7190	0.9494	
		2	750	0.20	0.2318	0.1154	0.2894	0.5034	0.1042
			1000	0.20	0.2438	0.1092	0.2916	0.4980	

Table 5.12. *Coverages* when H_0 is false for multivariate t_4 data.

p	$n_1 = n_2$	σ	B	δ	Med	Mean	Tr.Me	RMVN	Class
5	250	1	250	0.38	0.9642	0.9562	0.9916	0.9878	0.9548
			1000	0.38	0.9728	0.9572	0.9944	0.9880	
		2	250	0.38	0.5958	0.5960	0.7198	0.6488	0.6074
			1000	0.38	0.6188	0.6152	0.7490	0.6636	
15	750	1	750	0.20	0.9418	0.9270	0.9868	0.9714	0.9232
			1000	0.20	0.9422	0.9304	0.9860	0.9724	
		2	750	0.20	0.4934	0.4932	0.6422	0.5384	0.4754
			1000	0.20	0.4842	0.4916	0.6362	0.5252	

Table 5.13. *Coverages* when H_0 is false for lognormal data.

p	$n_1 = n_2$	σ	B	δ	Median	Mean	Tr.Me	RMVN	Class
5	250	1	250	0.45	0.9982	0.8256	0.9994	0.879	0.8208
			1000	0.45	0.9980	0.8324	0.9996	0.883	
		2	250	0.45	0.8210	0.4704	0.6488	0.0914	0.4630
			1000	0.45	0.8378	0.4646	0.6624	0.1038	
15	750	1	750	0.30	1.0000	0.9186	1.0000	0.8514	0.9120
			1000	0.30	1.0000	0.9178	1.0000	0.8544	
		2	750	0.30	0.9436	1.0000	0.5042	0.9438	1.0000
			1000	0.30	0.9484	1.0000	0.5022	0.9424	

CHAPTER 6

SIMULATIONS WITH THREE SAMPLES FOR BOOTSTRAPPING ANALOGS OF THE ONE-WAY MANOVA TEST

6.1 SIMULATION SETUP

The simulation used 5000 runs with B bootstrap samples and $p = 3$ groups. Olive (2017bc) suggests that the prediction region method can give good results when the number of bootstrap samples $B \geq 50m(p-1)$ and if $n \geq 50m(p-1)$, 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 $\boldsymbol{y}_1 = \boldsymbol{A}\boldsymbol{w}_1 + \delta_1\mathbf{1}$, $\boldsymbol{y}_2 = \sigma_2\boldsymbol{A}\boldsymbol{w}_2$, and $\boldsymbol{y}_3 = \sigma_3\boldsymbol{A}\boldsymbol{w}_3 + \delta_3\mathbf{1}$ or $\boldsymbol{y}_3 = \boldsymbol{w}_3$ where $\mathbf{1} = (1, .., 1)^T$ is a vector of ones and $\boldsymbol{A} = diag(1, \sqrt{2}, ..., \sqrt{m})$. The $\boldsymbol{w}_i$ distributions were the multivariate normal distribution $N_m(\mathbf{0}, \boldsymbol{I})$, the mixture distribution $0.6N_m(\mathbf{0}, \boldsymbol{I}) + 0.4N_m(\mathbf{0}, 25\boldsymbol{I})$, the multivariate t distribution with 4 degrees of freedom, and the multivariate lognormal distribution shifted to have nonzero mean $\boldsymbol{\mu} = 0.649\ \mathbf{1}$, but a population coordiatewise median of $\mathbf{0}$. Note that $\mathrm{Cov}(\boldsymbol{y}_2) = \sigma_2^2\ \mathrm{Cov}(\boldsymbol{y}_1)$, and for the first three distributions, $E(\boldsymbol{y}_i) = E(\boldsymbol{w}_i) = \mathbf{0}$ if $\delta_1 = \delta_3 = 0$. If $\boldsymbol{y}_3 = \boldsymbol{w}_3$ then $\mathrm{Cov}(\boldsymbol{y}_3) = c\boldsymbol{I}_m$ for some constant $c > 0$. If $\boldsymbol{y}_3 = \sigma_3\boldsymbol{A}\boldsymbol{w}_3 + \delta_3\mathbf{1}$, then $\mathrm{Cov}(\boldsymbol{y}_3) = \sigma_3^2\ \mathrm{Cov}(\boldsymbol{y}_1)$.

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 $\boldsymbol{y}_1 = (y_{11}, ..., y_{m1})^T$. 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: $N((z, ..., z)^T, diag(1, ..., m))$, type 4: y_{m1} 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.

6.1.1 Simulations for type I error with clean data

Tables 6.1 through 6.8 show simulation results for all for distributions with various covariance settings. We took $\delta_1 = \delta_3 = 0$, B : the number of bootstrap steps used also takes on different values throughout the simulation. Balanced and unbalanced designs have also been considered. For tables 6.5-6.8 $\sigma_2 = \sigma_3 = 1$. According to the tables, the new tests work well with all the distributions and with different covariance settings. The new tests could handle unbalanced designs as well. The classical test works well with the multivariate normal data and when the covariance matrices are the same, but the type I error is higher than the nominal level for different covariance settings. The classical test can be too conservative when the design is unbalanced. Having an unbalanced design and different covariance matrices is the worst case scenario for the classical test regardless of the data distribution.

Table 6.1. Type I error for clean MVN data with cov3I = F

m	n_1	n_2	n_3	B	σ_2	σ_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 6.2. Type I error for clean Mixture data with cov3I = F

m	n_1	n_2	n_3	B	σ_2	σ_3	Median	Mean	Tr.Mn	Class
5	200	200	200	400	1	1	0.0406	0.0526	0.0470	0.0446
				1000	1	1	0.0396	0.0600	0.0476	0.0492
				400	2	3	0.0448	0.0538	0.0484	0.0682
				1000	2	3	0.0418	0.0520	0.0412	0.0692
5	200	400	600	400	1	1	0.0440	0.0562	0.0438	0.0476
				1000	1	1	0.0376	0.0528	0.0486	0.0504
				400	2	3	0.0432	0.0518	0.0502	0.0082
				1000	2	3	0.0392	0.0528	0.0454	0.0060
10	400	400	400	800	1	1	0.0446	0.0604	0.0516	0.0498
				2000	1	1	0.0438	0.0592	0.0496	0.0502
				800	2	3	0.0454	0.0598	0.0478	0.0694
				2000	2	3	0.0460	0.0586	0.0468	0.0664
10	400	800	1200	800	1	1	0.0448	0.0590	0.0458	0.0494
				2000	1	1	0.0412	0.0590	0.0512	0.0532
				800	2	3	0.0490	0.0600	0.0528	0.0036
				2000	2	3	0.0444	0.0524	0.0464	0.0020
20	800	800	800	1600	1	1	0.0476	0.0628	0.0530	0.0472
				4000	1	1	0.0462	0.0606	0.0498	0.0490
				1600	2	3	0.0500	0.0680	0.0522	0.0738
				4000	2	3	0.0468	0.0676	0.0516	0.0720
20	800	1600	2400	1600	1	1	0.0522	0.0618	0.0560	0.0510
				4000	1	1	0.0480	0.0600	0.0504	0.0520
				1600	2	3	0.0488	0.0564	0.0566	0.0004
				4000	2	3	0.0432	0.0590	0.0476	0.0004

Table 6.3. Type I error for clean multivariate t data with cov3I = F

m	n_1	n_2	n_3	B	σ_2	σ_3	Median	Mean	Tr.Mn	Class
5	200	200	200	400	1	1	0.0420	0.0566	0.0524	0.0450
				1000	1	1	0.0388	0.0528	0.0542	0.0488
				400	2	3	0.0534	0.0596	0.0636	0.0666
				1000	2	3	0.0450	0.0564	0.0606	0.0650
5	200	400	600	400	1	1	0.0512	0.0532	0.0578	0.0478
				1000	1	1	0.0432	0.0596	0.0536	0.0526
				400	2	3	0.0458	0.0516	0.0556	0.0062
				1000	2	3	0.0464	0.0564	0.0578	0.0080
10	400	400	400	800	1	1	0.0448	0.0622	0.0588	0.0480
				2000	1	1	0.0490	0.0578	0.0608	0.0488
				800	2	3	0.0498	0.0622	0.0654	0.0680
				2000	2	3	0.0528	0.0576	0.0576	0.0652
10	400	800	1200	800	1	1	0.0470	0.0572	0.0530	0.0496
				2000	1	1	0.0448	0.0626	0.0544	0.0558
				800	2	3	0.0412	0.0508	0.0536	0.0030
				2000	2	3	0.0514	0.0582	0.0568	0.0026
20	800	800	800	1600	1	1	0.0474	0.0620	0.0576	0.0454
				4000	1	1	0.0532	0.0604	0.0634	0.0490
				1600	2	3	0.0534	0.0652	0.0640	0.0714
				4000	2	3	0.0556	0.0644	0.0652	0.0714
20	800	1600	2400	1600	1	1	0.0568	0.0630	0.0618	0.0504
				4000	1	1	0.0492	0.0584	0.0560	0.0532
				1600	2	3	0.0546	0.0570	0.0658	0.0008
				4000	2	3	0.0488	0.0544	0.0612	0.0004

Table 6.4. Type I error for clean lognormal data with cov3I = F

m	n_1	n_2	n_3	B	σ_2	σ_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 6.5. Type I error for clean MVN data with cov3I = T

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

Table 6.6. Type I error for clean Mixture data with cov3I = T

m	n_1	n_2	n_3	B	Median	Mean	Tr.Mn	Class
5	200	200	200	400	0.0424	0.0614	0.0438	0.0570
				1000	0.0446	0.0618	0.0460	0.0550
5	200	400	600	400	0.0524	0.0572	0.0542	0.1284
				1000	0.0434	0.0540	0.0498	0.1270
10	400	400	400	800	0.0422	0.0620	0.0542	0.0598
				2000	0.0450	0.0638	0.0524	0.0642
10	400	800	1200	800	0.0468	0.0558	0.0484	0.2368
				2000	0.0522	0.0548	0.0518	0.2356
20	800	800	800	1600	0.0520	0.0620	0.0560	0.0596
				4000	0.0492	0.0662	0.0518	0.0680
20	800	1600	2400	1600	0.0568	0.0598	0.0546	0.4326
				4000	0.0536	0.0614	0.0524	0.4338

Table 6.7. Type I error for clean Multivariate t data with cov3I = T

m	n_1	n_2	n_3	B	Median	Mean	Tr.Mn	Class
5	200	200	200	400	0.0396	0.0568	0.0538	0.0554
				1000	0.0464	0.0618	0.0552	0.0582
5	200	400	600	400	0.0450	0.0568	0.0500	0.1228
				1000	0.0496	0.0588	0.0552	0.1326
10	400	400	400	800	0.0506	0.0608	0.0616	0.0582
				2000	0.0472	0.0572	0.0602	0.0612
10	400	800	1200	800	0.0542	0.0592	0.0570	0.2294
				2000	0.0492	0.0560	0.0524	0.2372
20	800	800	800	1600	0.0558	0.0684	0.0662	0.0642
				4000	0.0528	0.0608	0.0622	0.0610
20	800	1600	2400	1600	0.0574	0.0576	0.0654	0.4382
				4000	0.0602	0.0652	0.0636	0.4344

Table 6.8. Type I error for clean lognormal data with cov3I = T

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

6.1.2 Simulations for power with clean data

Tables 6.9 through 6.12 show the power simulation results. In the table 6.9, the bootstrap test with the mean works similar to the classical test for multivariate normal data and produced the highest power. The test with the median had the worst power in this case. The classical test had the worst power in table 6.10 while the bootstrap test with the trimmed mean had the best power. In table 6.11, the bootstrap test with the trimmed mean was the best. The classical test was the worst except for the balanced design with the different covariance matrices case where the test with the median was the worst. In table 6.12, one of the three bootstrap tests was the best or the same (mainly the cases with power = 1) compared to the classical test.

Table 6.9. Power for MVN data with $\delta_1 = 0.2$ and $\delta_3 = 0.5$

m	n_1	n_2	n_3	B	σ_2	σ_3	Median	Mean	Tr.Mn	Class
5	200	200	200	400	1	1	0.9942	1.0000	1.0000	1.0000
				1000	1	1	0.9932	1.0000	0.9996	1.0000
				400	2	3	0.4198	0.6798	0.5902	0.8002
				1000	2	3	0.3056	0.5260	0.4394	0.6294
5	200	400	800	400	1	1	1.0000	1.0000	1.0000	1.0000
				1000	1	1	1.0000	1.0000	1.0000	1.0000
				400	2	3	0.8020	0.9738	0.9274	0.9026
				1000	2	3	0.8054	0.9648	0.9276	0.9004
10	400	400	400	800	1	1	1.0000	1.0000	1.0000	1.0000
				2000	1	1	1.0000	1.0000	1.0000	1.0000
				800	2	3	0.5860	0.8624	0.7628	0.9282
				2000	2	3	0.5942	0.8614	0.7740	0.9240
10	400	800	1200	800	1	1	1.0000	1.0000	1.0000	1.0000
				2000	1	1	1.0000	1.0000	1.0000	1.0000
				800	2	3	0.9792	0.9998	0.9984	0.9990
				2000	2	3	0.9794	0.9992	0.9970	0.9982
20	800	800	800	1600	1	1	1.0000	1.0000	1.0000	1.0000
				4000	1	1	1.0000	1.0000	1.0000	1.0000
				1600	2	3	0.9242	0.9970	0.9854	0.9990
				4000	2	3	0.9246	0.9968	0.9876	0.9992
20	800	1600	2400	1600	1	1	1.0000	1.0000	1.0000	1.0000
				4000	1	1	1.0000	1.0000	1.0000	1.0000
				1600	2	3	1.0000	1.0000	1.0000	1.0000
				4000	2	3	1.0000	1.0000	1.0000	1.0000

Table 6.10. Power for Mixture data $\delta_1 = 0.2$ and $\delta_3 = 0.5$

m	n_1	n_2	n_3	B	σ_2	σ_3	Median	Mean	Tr.Mn	Class
5	200	200	200	400	1	1	0.7742	0.3168	0.8548	0.2912
				1000	1	1	0.7954	0.3126	0.8654	0.2868
				400	2	3	0.1344	0.0876	0.1530	0.1040
				1000	2	3	0.1392	0.0824	0.1552	0.0992
5	200	400	800	400	1	1	0.9994	0.7694	1.0000	0.7646
				1000	1	1	0.9996	0.7598	1.0000	0.7550
				400	2	3	0.3950	0.1566	0.4742	0.0258
				1000	2	3	0.4038	0.1582	0.4776	0.0244
10	400	400	400	800	1	1	0.9928	0.6144	0.9976	0.5868
				2000	1	1	0.9910	0.6104	0.9968	0.5758
				800	2	3	0.2520	0.1166	0.3020	0.1500
				2000	2	3	0.2474	0.1222	0.2934	0.1568
10	400	800	1200	800	1	1	1.0000	0.9728	1.0000	0.9714
				2000	1	1	1.0000	0.9740	1.0000	0.9720
				800	2	3	0.6626	0.2240	0.7466	0.0574
				2000	2	3	0.6640	0.2218	0.7498	0.0496
20	800	800	800	1600	1	1	1.0000	0.9156	1.0000	0.9004
				4000	1	1	1.0000	0.9214	1.0000	0.9074
				1600	2	3	0.5058	0.1830	0.5920	0.2308
				4000	2	3	0.5030	0.1884	0.5914	0.2396
20	800	1600	2400	1600	1	1	1.0000	0.9998	1.0000	0.9998
				4000	1	1	1.0000	1.0000	1.0000	1.0000
				1600	2	3	0.9594	0.4310	0.9838	0.0830
				4000	2	3	0.9606	0.4208	0.9864	0.0758

Table 6.11. Power for Multivariate t data $\delta_1 = 0.2$ and $\delta_3 = 0.5$

m	n_1	n_2	n_3	B	σ_2	σ_3	Median	Mean	Tr.Mn	Class
5	200	200	200	400	1	1	0.9824	0.9790	0.9982	0.9770
				1000	1	1	0.9832	0.9806	0.9980	0.9778
				400	2	3	0.2564	0.2760	0.3450	0.3448
				1000	2	3	0.2622	0.2750	0.3538	0.3490
5	200	400	800	400	1	1	1.0000	1.0000	1.0000	1.0000
				1000	1	1	1.0000	1.0000	1.0000	1.0000
				400	2	3	0.7328	0.7132	0.8656	0.4634
				1000	2	3	0.7382	0.7210	0.8604	0.4620
10	400	400	400	800	1	1	1.0000	1.0000	1.0000	0.9998
				2000	1	1	1.0000	1.0000	1.0000	1.0000
				800	2	3	0.5048	0.5276	0.6646	0.6324
				2000	2	3	0.5154	0.5266	0.6750	0.6290
10	400	800	1200	800	1	1	1.0000	1.0000	1.0000	1.0000
				2000	1	1	1.0000	1.0000	1.0000	1.0000
				800	2	3	0.9576	0.9494	0.9908	0.8522
				2000	2	3	0.9574	0.9468	0.9892	0.8538
20	800	800	800	1600	1	1	1.0000	1.0000	1.0000	1.0000
				4000	1	1	1.0000	1.0000	1.0000	1.0000
				1600	2	3	0.8670	0.8698	0.9582	0.9288
				4000	2	3	0.8686	0.8660	0.9600	0.9272
20	800	1600	2400	1600	1	1	1.0000	1.0000	1.0000	1.0000
				4000	1	1	1.0000	1.0000	1.0000	1.0000
				1600	2	3	1.0000	0.9996	1.0000	0.9974
				4000	2	3	1.0000	0.9998	1.0000	0.9972

Table 6.12. Power for lognormal data $\delta_1 = 0.2$ and $\delta_3 = 0.5$

m	n_1	n_2	n_3	B	σ_2	σ_3	Median	Mean	Tr.Mn	Class
5	200	200	200	400	1	1	0.9928	0.7360	0.9974	0.6936
				1000	1	1	0.9954	0.7500	0.9986	0.7086
				400	2	3	0.2712	1.0000	0.5182	0.9998
				1000	2	3	0.2810	1.0000	0.5320	1.0000
5	200	400	800	400	1	1	1.0000	0.9934	1.0000	0.9940
				1000	1	1	1.0000	0.9942	1.0000	0.9944
				400	2	3	0.7906	1.0000	0.9922	1.0000
				1000	2	3	0.7956	1.0000	0.9940	1.0000
10	400	400	400	800	1	1	1.0000	0.9774	1.0000	0.9670
				2000	1	1	1.0000	0.9752	1.0000	0.9662
				800	2	3	0.5744	1.0000	0.9402	1.0000
				2000	2	3	0.5632	1.0000	0.9404	1.0000
10	400	800	1200	800	1	1	1.0000	1.0000	1.0000	1.0000
				2000	1	1	1.0000	1.0000	1.0000	1.0000
				800	2	3	0.9780	1.0000	1.0000	1.0000
				2000	2	3	0.9808	1.0000	1.0000	1.0000
20	800	800	800	1600	1	1	1.0000	1.0000	1.0000	1.0000
				4000	1	1	1.0000	0.9998	1.0000	0.9998
				1600	2	3	0.9182	1.0000	0.9998	1.0000
				4000	2	3	0.9212	1.0000	1.0000	1.0000
20	800	1600	2400	1600	1	1	1.0000	1.0000	1.0000	1.0000
				4000	1	1	1.0000	1.0000	1.0000	1.0000
				1600	2	3	1.0000	1.0000	1.0000	1.0000
				4000	2	3	1.0000	1.0000	1.0000	1.0000

6.1.3 Simulations for type I error with contaminated data

For the tables 6.13 through 6.16 the data was contaminated by using the five types of outliers mentioned earlier, and $\gamma = 10\%$ or 5% was used. In table 6.13 the test with the median works reasonably well (close to the nominal coverage) with 10% of outliers with all the distribution and for all the outlier types with the exception of outlier type 3. All the other tests including the classical test failed. Table 6.14 uses 10 parameters with 5% of outliers and results are similar to those in table 6.13. Tables 6.15 and 6.16 uses 20 parameters with 10% and 5% outliers respectively.

Tables 6.17 through 6.20 show the simulations used to get the power curves in section 4.2. It is clear from the graphs that classical test is worst when there is an unbalanced design with different covariance matrices. Also see figures 4.2-4.4. Figure 4.1 uses a balanced design with same covariance matrices. Then the bootstrap test with the mean, the large sample MANOVA type test and the classical test are best and all there of these test almost overlap. In figure 4.2, the bootstrap test with the mean does best while in 4.3 and 4.4 the bootstrap test with the trimmed mean does best.

Table 6.13. Type I error with contaminated data: $m = 5$, $\gamma = 0.1$

Dist.	m	$n_1 = n_2 = n_3$	B	outlier	Median	Mean	Tr.Mn	Class
1	5	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	5	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	5	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	5	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

Table 6.14. Type I error with contaminated data: $m = 10$, $\gamma = 0.05$

Dist.	m	$n_1 = n_2 = n_3$	B	outlier	Median	Mean	Tr.Mn	Class
1	10	800	1000	1	0.0274	0.9720	0.1318	0.9956
				2	0.0192	1.0000	0.1132	1.0000
				3	0.4710	1.0000	0.6760	1.0000
				4	0.0872	0.9660	0.1264	0.9974
				5	0.0756	1.0000	0.1084	1.0000
2	10	800	1000	1	0.0272	0.2586	0.1246	0.2734
				2	0.0152	0.9182	0.0992	0.9764
				3	0.4288	0.9976	0.5954	1.0000
				4	0.0852	0.2454	0.1102	0.2658
				5	0.0644	0.9032	0.0898	0.9750
3	10	800	1000	1	0.0252	0.8384	0.1266	0.9178
				2	0.0158	0.9996	0.1160	1.0000
				3	0.4596	1.0000	0.6544	1.0000
				4	0.0954	0.8254	0.1270	0.9158
				5	0.0804	1.0000	0.1054	1.0000
4	10	800	1000	1	0.0232	1.0000	0.6856	1.0000
				2	0.0146	1.0000	0.5036	1.0000
				3	0.4130	1.0000	0.9962	1.0000
				4	0.0806	1.0000	0.7362	1.0000
				5	0.0714	1.0000	0.6184	1.0000

Table 6.15. Type I error with contaminated data: $m = 20$, $\gamma = 0.1$

Dist.	m	$n_1 = n_2 = n_3$	B	outlier	Median	Mean	Tr.Mn	Class
1	20	800	2000	1	0.1522	0.8752	0.3994	0.8906
				2	0.0960	1.0000	0.2958	1.0000
				3	0.9986	1.0000	1.0000	1.0000
				4	0.2170	0.8612	0.3684	0.8800
				5	0.1740	1.0000	0.2890	1.0000
2	20	800	2000	1	0.1256	0.1496	0.3378	0.1342
				2	0.0988	0.9434	0.2560	0.9654
				3	0.9762	1.0000	0.9924	1.0000
				4	0.1994	0.1426	0.3028	0.1210
				5	0.1578	0.9438	0.2392	0.9664
3	20	800	2000	1	0.1438	0.5946	0.3928	0.5666
				2	0.0904	1.0000	0.2900	1.0000
				3	0.9970	1.0000	0.9998	1.0000
				4	0.2184	0.5888	0.3598	0.5716
				5	0.1546	1.0000	0.2804	1.0000
4	20	800	2000	1	0.2928	1.0000	0.8634	1.0000
				2	0.0892	1.0000	0.9490	1.0000
				3	0.9912	1.0000	1.0000	1.0000
				4	0.1756	1.0000	0.9986	1.0000
				5	0.1512	1.0000	0.9920	1.0000

Table 6.16. Type I error with contaminated data: $m = 20$, $\gamma = 0.05$

Dist.	m	$n_1 = n_2 = n_3$	B	outlier	Median	Mean	Tr.Mn	Class
1	20	800	1000	1	0.0046	0.7012	0.1330	0.7932
				2	0.0010	0.9974	0.1148	1.0000
				3	0.5950	0.9998	0.7496	1.0000
				4	0.0886	0.6820	0.1218	0.7758
				5	0.0878	0.9972	0.1138	1.0000
2	20	800	1000	1	0.0044	0.1468	0.1064	0.1338
				2	0.0014	0.7946	0.0960	0.9004
				3	0.5116	0.9776	0.6054	1.0000
				4	0.0800	0.1434	0.1038	0.1294
				5	0.0696	0.7718	0.0850	0.8908
3	20	800	1000	1	0.0032	0.4786	0.1346	0.5006
				2	0.0020	0.9908	0.1076	1.0000
				3	0.5762	0.9992	0.7190	1.0000
				4	0.0816	0.4524	0.1136	0.4848
				5	0.0784	0.9888	0.1020	1.0000
4	20	800	1000	1	0.0016	1.0000	0.9664	1.0000
				2	0.0014	1.0000	0.9216	1.0000
				3	0.4778	1.0000	1.0000	1.0000
				4	0.0822	1.0000	0.9892	1.0000
				5	0.0720	1.0000	0.9768	1.0000

Table 6.17. Power curve for MVN data with $m = 5, \sigma_1 = 1, \sigma_2 = 1, \sigma_3 = 1, n_1 = 200, n_2 = 200$ and $n_3 = 200$

δ_1	δ_2	δ_3	Med	Mean	Tr.Me	Clas	MaTyp
0	0	0	0.0442	0.0588	0.0562	0.0502	0.0470
0.04	0.08	0.12	0.0700	0.1206	0.1048	0.1070	0.1072
0.08	0.16	0.24	0.1776	0.326	0.2714	0.2992	0.2982
0.12	0.24	0.36	0.4308	0.696	0.6002	0.6758	0.6694
0.16	0.32	0.48	0.7266	0.9442	0.8836	0.9384	0.9364
0.20	0.40	0.60	0.9142	0.9952	0.9834	0.9958	0.9948
0.24	0.48	0.72	0.9876	0.9998	0.9994	0.9998	0.9998
0.30	0.60	0.90	0.9998	1	1	1	1
0.35	0.70	1.05	1	1	1	1	1

Table 6.18. Power curve for MVN data with $m = 5, \sigma_1 = 1, \sigma_2 = 2, \sigma_3 = 5, n_1 = 200, n_2 = 400$ and $n_3 = 600$

δ_1	δ_2	δ_3	Med	Mean	Tr.Me	Clas	MaTyp
0	0	0	0.0492	0.0510	0.0542	0.0034	0.0446
0.04	0.08	0.12	0.0636	0.0786	0.0726	0.0062	0.0654
0.08	0.16	0.24	0.0804	0.1224	0.1086	0.0060	0.1044
0.12	0.24	0.36	0.1400	0.2322	0.1942	0.0146	0.2158
0.16	0.32	0.48	0.2388	0.4204	0.3494	0.0334	0.3956
0.20	0.4	0.6	0.3836	0.6438	0.5372	0.0710	0.6186
0.24	0.48	0.72	0.5538	0.8280	0.7278	0.1240	0.8126
0.30	0.60	0.90	0.7922	0.9648	0.9200	0.2868	0.9590
0.35	0.70	1.05	0.9170	0.9964	0.9832	0.4898	0.9958

Table 6.19. Power curve for Mixture data with $m = 5, \sigma_1 = 1, \sigma_2 = 2, \sigma_3 = 5, n_1 = 200, n_2 = 400$ and $n_3 = 600$

δ_1	δ_2	δ_3	Med	Mean	Tr.Me	Clas	MaTyp
0	0	0	0.0486	0.0522	0.0524	0.0474	0.0054
0.04	0.08	0.12	0.0504	0.0602	0.0520	0.0510	0.0034
0.08	0.16	0.24	0.0588	0.0624	0.0670	0.0540	0.0042
0.12	0.24	0.36	0.1196	0.0688	0.1356	0.059	0.0054
0.16	0.32	0.48	0.1178	0.0820	0.1350	0.0708	0.0068
0.20	0.40	0.60	0.3846	0.6354	0.5432	0.6148	0.0644
0.24	0.48	0.72	0.2350	0.1120	0.2948	0.0964	0.0078
0.30	0.60	0.90	0.3826	0.1570	0.4650	0.1378	0.0086
0.40	0.80	1.20	0.689	0.2602	0.7726	0.2326	0.0190
0.45	0.90	1.35	0.8032	0.3134	0.8646	0.2916	0.0250
0.50	1.00	1.50	0.9006	0.3698	0.9396	0.3518	0.0312
0.60	1.20	1.80	0.9780	0.5592	0.9868	0.5370	0.0538
0.80	1.60	2.40	0.9998	0.8450	0.9998	0.8296	0.1358

Table 6.20. Power curve for Multivariate t_4 data with $m = 5, \sigma_1 = 1, \sigma_2 = 2, \sigma_3 = 5, n_1 = 200, n_2 = 400$ and $n_3 = 600$

δ_1	δ_2	δ_3	Med	Mean	Tr.Me	Clas	MaTyp
0	0	0	0.0500	0.0530	0.0572	0.0046	0.0450
0.04	0.08	0.12	0.0516	0.0650	0.0652	0.0042	0.0560
0.08	0.16	0.24	0.0750	0.0814	0.0940	0.0056	0.0714
0.12	0.24	0.36	0.1304	0.1460	0.1770	0.0098	0.1308
0.16	0.32	0.48	0.2102	0.2214	0.2888	0.013	0.2038
0.20	0.40	0.60	0.3302	0.338	0.4618	0.0198	0.3168
0.24	0.48	0.72	0.487	0.4758	0.6278	0.0380	0.4572
0.30	0.60	0.90	0.7344	0.7122	0.8526	0.0858	0.7004
0.35	0.70	1.05	0.873	0.8572	0.9520	0.1462	0.8440
0.40	0.80	1.20	0.9576	0.9472	0.9900	0.2586	0.9416
0.45	0.90	1.35	0.9892	0.9848	0.9994	0.3768	0.9826

CHAPTER 7
DISCUSSION

The first part of this dissertation suggests a practical method to perform a multivariate two sample test when the asymptotic covariance matrix of the statistic T_i is difficult to estimate. Such tests may be useful when the data distribution is unknown or outliers are present. The method was illustrated with the coordinatewise median, sample mean, 25% trimmed mean, and RMVN estimators. All four estimators work well when the prediction region method was applied to the clean data, although care needs to be taken with the multivariate lognormal distribution where the four estimators T_i are estimating different parameters $\boldsymbol{\mu}_{T_i}$.

Both the sample mean and the 25% trimmed mean failed to achieve the nominal coverage when H_0 is true with the contaminated data. The coordinatewise median could handle up to 10% outliers, while the RMVN estimator could handle up to 40% outliers. Both estimators were robust to the equal covariance assumption.

This work also gives practical alternatives for the classical one way MANOVA test. One novel large sample test was given. This test does not use the unreasonably strong equal covariance assumption and worked well in the simulations. Another practical method to bootstrap analogs of the one way MANOVA test was also suggested. Such tests may be useful when the data distribution is unknown or outliers are present. The method was illustrated with the coordinatewise median, sample mean, 25% trimmed mean. All three estimators work well when the prediction region method was applied to the clean data. The classical test does not work well when the assumptions are not met. The classical test is worst in terms of type I error and power when working with an unbalanced design and the equal covariance assumption is violated.

The bootstrap test with the coordinatewise median is recommended when the outliers are present. To make the bootstrap test more efficient with outliers, a robust location estimator can be used (e.g. RMVN location estimator) as in the two sample case.

Konietschke, Bathke, Harrar, and Pauly (2015) suggest a method for bootstrapping

the MANOVA model, and Willems, Pison, Rousseeuw, and Van Aelst (2002) suggest a robust one sample Hotelling's T^2 type test. References for robust one way MANOVA tests are in Finch and French (2013), Todorov and Filzmoser (2010), Van Aelst and Willems (2011), and Wilcox (1995).

The R software was used in the simulation. See R Core Team (2016). Programs are in the Olive (2017b) collection of R functions *mpack.txt* available from (http://lagrange.math.siu.edu/Olive/mpack.txt). The function `hot2sim` was used to simulate the two sample tests of hypotheses, The Curran (2013) R package `Hotelling` was used to perform the classical 2 sample Hotelling's T^2 test. The function `manbtsim2` was used to simulate the bootstrapped one way MANOVA tests, `predreg` computes the confidence region given the bootstrap values from `rhot2boot`, and `manbtsim4` simulates the large sample one way MANOVA type test and to produce the power curves.

APPENDICES

R CODES FOR THE SIMULATIONS

APPENDIX I

The following is the R code for the simulations for Bootstrapping Analogs of the Two Sample Hotelling's T^2 Test.

```
source("http://lagrange.math.siu.edu/Olive/mpack.txt")
library(MASS)


hot2sim<-function(n = 100, n2 = 100, p = 2, csteps = 5, B= 100, gam = 0.4,
nruns = 100, xtype = 1, outliers = 0, z = 10, sig = 1, eps = 0.4,
dd=4, delta = 0){
# This R function simulates the two sample Hotelling's T^2 test based on the
#bootstrap. Here x2 is clean, x = x1 may have outliers.
# Uses the coordinatewise mean, median, 25% trimmed mean, and RMVN location
estimators.
# Need p > 1. Want n > 20p, n2 > 20p, B > 20p.
# Multiply x by A where xtype = 1 for MVN Np(0,I),
# 2, (with delta = eps) for (1 - delta) Np(0,I) + delta Np(0, 25 I)
# 3 for multivariate t_d with d = dd
# 4 for lognormal.
# outliers = 0 for no outliers and X~N(0,diag(1,...,p)),
# 1 for outliers a tight cluster at major axis (0,...,0,z)'
# 2 for outliers a tight cluster at minor axis  (z,0, ...,0)'
# 3 for outliers X~N((z,...,z)',diag(1,...,p))
# 4 for outliers X[i,p] = z
# 5 for outliers X[i,1] = z
# For the point mass outlier types 4 and 5, need gam much smaller than 0.4.
# Power can be estimated by increasing delta so mu = delta(1,...,1)
# and mu2 = (0, ..., 0).
```

```
# Cov(x) = diag(1,2,...,p), Cov(x2) = sig^2 Cov(x) for clean data.
# For outliers=0, want hquant and rquant approx 1.
    A <- sqrt(diag(1:p))
    munot <- 0 * (1:p)
    mu <- delta * (1 + munot)
    val <- floor(gam * n)
        indx <- 1:n
        indx2 <- 1:n2
        medmus <- matrix(0,nrow=B,ncol=p)
        mnmus <- medmus
        tmnmus <- medmus
        rmvnmus <- medmus
        medcv <- 0
        mncv <- 0
        tmncv <- 0
        rmvncv <- 0
        chisqcut <- qchisq(p=0.95,df=p)
        cutoffs <- matrix(0,nrow=nruns,ncol=4)
        for(i in 1:nruns) {
#make data
        x <- matrix(rnorm(n * p), ncol = p, nrow = n)
               x2 <- matrix(rnorm(n2 * p), ncol = p, nrow = n2)
        if(xtype == 2) {
            zu <- runif(n)
            x[zu < eps,  ] <- x[zu < eps,  ] * 5
                        zu <- runif(n2)
            x2[zu < eps,  ] <- x2[zu < eps,  ] * 5
        }
        if(xtype == 3) {
            zu <- sqrt(rchisq(n, dd)/dd)
```

```
    x <- x/zu
                zu <- sqrt(rchisq(n2, dd)/dd)
    x2 <- x2/zu
}
if(xtype == 4){  #Want pop coord med(x) = 0.
    x <- exp(x)
                x <- x - 1
                x2 <- exp(x2)
                x2 <- x2 - 1
        }
x <- x %*% A
        x2 <- x2 %*% A
        x2 <- sig * x2
if(outliers == 1) {
            x[1:val,  ] <- matrix(rnorm(val * p, sd = 0.01), ncol
         = p, nrow = val)
    x[1:val, p] <- x[1:val, p] + z
}
if(outliers == 2) {
    x[1:val,  ] <- matrix(rnorm(val * p, sd = 0.01), ncol
         = p, nrow = val)
    x[1:val, 1] <- x[1:val, 1] + z
}
if(outliers == 3) {
    tem <- z + 0 * 1:p
    x[1:val,  ] <- x[1:val,  ] + tem
}
if(outliers == 4) {
    x[1:val, p] <- z
}
```

```
if(outliers == 5) {
    x[1:val, 1] <- z
}
x <- mu + x
  #clean x has mean mu =delta(1,...,1)^T, x2 has mean (0,...,0)^T
        #get bootstrapped Tx - Tx2 for various statistics T
        for(j in 1:B){
        tem <- sample(indx,n,replace=T)
        medx <- apply(x[tem,],2,median)
        mnx <- apply(x[tem,],2,mean)
        tmnx <- apply(x[tem,],2,tmn)
        rmvnx <- covrmvn(x[tem,])$center
        tem <- sample(indx2,n2,replace=T)
        medx2 <- apply(x2[tem,],2,median)
        mnx2 <- apply(x2[tem,],2,mean)
        tmnx2 <- apply(x2[tem,],2,tmn)
        rmvnx2 <- covrmvn(x2[tem,])$center
        medmus[j,] <- medx-medx2
        mnmus[j,] <- mnx-mnx2
        tmnmus[j,] <- tmnx-tmnx2
        rmvnmus[j,] <- rmvnx-rmvnx2
 }
        outmed<-predreg(medmus)
        medcv <- medcv + outmed$inr
        outmn <- predreg(mnmus)
        mncv <- mncv + outmn$inr
        outtmn <- predreg(tmnmus)
        tmncv <- tmncv + outtmn$inr
        outrmvn <- predreg(rmvnmus)
      rmvncv <- rmvncv + outrmvn$inr
```

```
                cutoffs[i,]<-
                c(outmed$cuplim,outmn$cuplim,outtmn$cuplim,outrmvn$cuplim)^2
    }
    medcv <- 1 - medcv/nruns #prop of times Ho is rejected
        mncv  <- 1 - mncv/nruns
        tmncv  <- 1 - tmncv/nruns
        rmvncv <- 1 - rmvncv/nruns
        mncut <- apply(cutoffs,2,mean)
    list(chisqcut = chisqcut, mncut=mncut, medcv = medcv,
              mncv = mncv,tmncv=tmncv,rmvncv=rmvncv)
}


tmn<-function(x, tp = 0.25)
{# computes 100tp% trimmed mean
    mean(x, trim = tp)
}
```

The following is the R code for the simulations for Bootstrapping Analogs of the One-Way MANOVA Test.

```
library(MASS)
source("http://lagrange.math.siu.edu/Olive/mpack.txt")

manbtsim2<-function(n = 100, n2 = 100, n3 = 100, m = 2, csteps = 5, B= 100,
gam = 0.4, nruns = 100, ytype = 1, outliers = 0, z = 10, sig2 = 1, sig3 = 1,
eps = 0.4, dd=4, delta = 0, delta3 = 0, cov3I = F){
# This R function simulates a one way Manova type bootstrap test for
# Ho: mu_1 = mu_2 = mu_3 where p = g = 3 = number of groups.
#Here y2 and y3 are clean, y = y1 may have outliers, and y is m by 1.
# Uses the coordinatewise mean, median, and 25% trimmed mean.
## Does not use the slow RMVN location estimators.
# Need m > 1. Want n > 20m, n2 > 20m, n3 > 20m, B > 20m.
# Multiply y by A where ytype = 1 for MVN Nm(0,I),
# 2  for (1 - eps) Nm(0,I) + eps Nm(0, 25 I),
# 3 for multivariate t_d with d = dd,
# 4 for lognormal.
# outliers = 0 for no outliers and Y~N(0,diag(1,...,m)),
# 1 for outliers a tight cluster at major axis (0,...,0,z)'
# 2 for outliers a tight cluster at minor axis  (z,0, ...,0)'
# 3 for outliers Y~N((z,...,z)',diag(1,...,m))
# 4 for outliers Y[i,m] = z
# 5 for outliers Y[i,1] = z
# For the point mass outlier types 4 and 5, need gam much smaller than 0.4.
# Power can be estimated by increasing delta so mu1 = delta(1,...,1)
# and mu2 = (0, ..., 0), mu3 = delta3(1,...,1).
# Cov(y) = diag(1,2,...,m), Cov(y2) = sig^2 Cov(y) for clean data.
# Cov(y3) = sig^3 Cov(y) for clean data if cov3I = F,
```

```
# or Cov(y3) = cI_3 if cov3I = T.
# For outliers=0, want hquant and rquant approx 1.
A <- sqrt(diag(1:m))
        munot <- 0 * (1:m)
mu <- delta * (1 + munot)
        mu3 <- delta3 * (1 + munot)
val <- floor(gam * n)
        indx <- 1:n
        indx2 <- 1:n2
        indx3 <- 1:n3
        medmus <- matrix(0,nrow=B,ncol=2*m)
        mnmus <- medmus
        tmnmus <- medmus
        medcv <- 0
        mncv <- 0
        tmncv <- 0
crej <- 0
gp <- 0
grp <- 0
pval <- 0
out <- 0

        chisqcut <- qchisq(p=0.95,df=2*m)
        cutoffs <- matrix(0,nrow=nruns,ncol=3)
        for(i in 1:nruns) {
#make data
y <- matrix(rnorm(n * m), ncol = m, nrow = n)
               y2 <- matrix(rnorm(n2 * m), ncol = m, nrow = n2)
               y3 <- matrix(rnorm(n3 * m), ncol = m, nrow =n3)
if(ytype == 2) {
```

```
zu <- runif(n)
y[zu < eps,  ] <- y[zu < eps,  ] * 5
                        zu <- runif(n2)
y2[zu < eps,  ] <- y2[zu < eps,  ] * 5
                        zu <- runif(n3)
y3[zu < eps,  ] <- y3[zu < eps,  ] * 5
}
if(ytype == 3) {
zu <- sqrt(rchisq(n, dd)/dd)
y <- y/zu
                        zu <- sqrt(rchisq(n2, dd)/dd)
y2 <- y2/zu
                        zu <- sqrt(rchisq(n3, dd)/dd)
y3 <- y3/zu
}
if(ytype == 4){  #Want pop coord med(y) = 0.
y <- exp(y)
                        y <- y - 1
                        y2 <- exp(y2)
                        y2 <- y2 - 1
                        y3 <- exp(y3)
                        y3 <- y3 - 1
                }
y <- y %*% A
                y2 <- y2 %*% A
                y2 <- sig2 * y2
                if( cov3I != T){
                y3 <- y3 %*% A
                y3 <- sig3 * y3}
if(outliers == 1) {
```

```
                    y[1:val,  ] <- matrix(rnorm(val * m, sd = 0.01), ncol
 = m, nrow = val)
y[1:val, m] <- y[1:val, m] + z
}
if(outliers == 2) {
y[1:val,  ] <- matrix(rnorm(val * m, sd = 0.01), ncol
 = m, nrow = val)
y[1:val, 1] <- y[1:val, 1] + z
}
if(outliers == 3) {
tem <- z + 0 * 1:m
y[1:val,  ] <- y[1:val,  ] + tem
}
if(outliers == 4) {
y[1:val, m] <- z
}
if(outliers == 5) {
y[1:val, 1] <- z
}
y <- mu + y
                y3 <- mu3 + y3
          #clean y has mean mu =delta(1,...,1)^T,
          #y2 has mean (0,..,0)^T and y3 has mean delta3 (1,...,1)^T
                #get bootstrapped Ty - Ty3, Ty2 - Ty3 for various statistics T
                for(j in 1:B){
                tem <- sample(indx,n,replace=T)
                medy <- apply(y[tem,],2,median)
                mny <- apply(y[tem,],2,mean)
                tmny <- apply(y[tem,],2,tmn)
                tem <- sample(indx2,n2,replace=T)
```

```
                medy2 <- apply(y2[tem,],2,median)
                mny2 <- apply(y2[tem,],2,mean)
                tmny2 <- apply(y2[tem,],2,tmn)
                tem <- sample(indx3,n3,replace=T)
                medy3 <- apply(y3[tem,],2,median)
                mny3 <- apply(y3[tem,],2,mean)
                tmny3 <- apply(y3[tem,],2,tmn)
                medmus[j,] <- c(medy-medy3,medy2-medy3)
                mnmus[j,] <- c(mny-mny3,mny2-mny3)
                tmnmus[j,] <- c(tmny-tmny3,tmny2-tmny3)
            }
                outmed<-predreg(medmus)
                medcv <- medcv + outmed$inr
                outmn <- predreg(mnmus)
                mncv <- mncv + outmn$inr
                outtmn <- predreg(tmnmus)
                tmncv <- tmncv + outtmn$inr
                cutoffs[i,]<-c(outmed$cuplim,outmn$cuplim,outtmn$cuplim)^2

#Get the classical test coverage
yall<-rbind(y,y2,y3)
yall<-as.matrix(yall)
  gp <- c(rep(1, n),rep(2, n2),rep(3, n3))
grp<-factor(gp)
out<-manova(yall~grp)

pval <- summary(out,test="Hotelling-Lawley")$stats[1,6]
    #pvalue for Hotelling-Lawley's test

if(pval < 0.05){
```

```
crej <- crej +1
}

}
medcv <- 1 - medcv/nruns #prop of times Ho is rejected
        mncv  <- 1 - mncv/nruns
        tmncv  <- 1 - tmncv/nruns
        mncut <- apply(cutoffs,2,mean)
ccv <- crej/nruns


list(chisqcut = chisqcut, mncut=mncut, medcv = medcv,
             mncv = mncv,tmncv=tmncv,ccv=ccv)
}
```

The following is the R code for the power curves for Bootstrapping Analogs of the One-Way MANOVA Test and the large sample test from the section 4.1.

```
manbtsim4 <- function(n1 = 100, n2 = 100, n3 = 100, m = 2, csteps = 5, B= 100,
gam = 0.4, nruns = 100, ytype = 1, outliers = 0, z = 10, sig2 = 1, sig3 = 1,
eps = 0.4, dd=4, delta1 = 0, delta2 = 0, delta3 = 0, cov3I = F){
#needs library(MASS)
# This R function simulates  one way Manova type  tests for
# Ho: mu_1 = mu_2 = mu_3 where p = g = 3 = number of groups.
#Can vary the mean vectors mui better than in manbtsim2.
#Has one more test than manbtsim3.
#Here y2 and y3 are clean,  y1 may have outliers, and y1 is m by 1.
# Uses the coordinatewise mean, median, and 25% trimmed mean.
#Also does the classical Hotelling Lawley one way MANOVA test.
#Partially written by Hasthika S. Rupasinghe Arachchige Don.
## Does not use the slow RMVN location estimators.
# Need m > 1. Want n > 20m, n2 > 20m, n3 > 20m, B > 20m.
# Multiply y by A where ytype = 1 for MVN Nm(0,I),
# 2  for (1 - eps) Nm(0,I) + eps Nm(0, 25 I),
# 3 for multivariate t_d with d = dd,
# 4 for lognormal.
# outliers = 0 for no outliers and Y~N(0,diag(1,...,m)),
# 1 for outliers a tight cluster at major axis (0,...,0,z)'
# 2 for outliers a tight cluster at minor axis  (z,0, ...,0)'
# 3 for outliers Y~N((z,...,z)',diag(1,...,m))
# 4 for outliers Y[i,m] = z
# 5 for outliers Y[i,1] = z
# For the point mass outlier types 4 and 5, need gam much smaller than 0.4.
# Power can be estimated by using unequal deltai so mu1 = delta1(1,...,1)
```

```
# and mu2 = delta2(1, ..., 1), mu3 = delta3(1,...,1).
# Cov(y1) = diag(1,2,...,m), Cov(y2) = sig^2 Cov(y1) for clean data.
# Cov(y3) = sig^3 Cov(y1) for clean data if cov3I = F,
# or Cov(y3) = cI_3 if cov3I = T.
# For outliers=0, want hquant and rquant approx 1.
        A <- sqrt(diag(1:m))
        munot <- 0 * (1:m)
        mu1 <- delta1 * (1 + munot)
        mu2 <- delta2 * (1 + munot)
        mu3 <- delta3 * (1 + munot)
        val <- floor(gam * n1)
        indx1 <- 1:n1
        indx2 <- 1:n2
        indx3 <- 1:n3
        medmus <- matrix(0,nrow=B,ncol=2*m)
        mnmus <- medmus
        tmnmus <- medmus
        medcv <- 0
        mncv <- 0
        tmncv <- 0
        mantcov <- 0
        crej <- 0
        gp <- 0
        grp <- 0
        pval <- 0
        out <- 0

        chisqcut <- qchisq(p=0.95,df=2*m)
        cutoffs <- matrix(0,nrow=nruns,ncol=3)
        for(i in 1:nruns) {
```

```
#make data
               y1 <- matrix(rnorm(n1 * m), ncol = m, nrow = n1)
               y2 <- matrix(rnorm(n2 * m), ncol = m, nrow = n2)
               y3 <- matrix(rnorm(n3 * m), ncol = m, nrow = n3)
                if(ytype == 2) {
                        zu <- runif(n)
                        y1[zu < eps,  ] <- y1[zu < eps,  ] * 5
                        zu <- runif(n2)
                        y2[zu < eps,  ] <- y2[zu < eps,  ] * 5
                        zu <- runif(n3)
                        y3[zu < eps,  ] <- y3[zu < eps,  ] * 5
                }
                if(ytype == 3) {
                        zu <- sqrt(rchisq(n1, dd)/dd)
                        y1 <- y1/zu
                        zu <- sqrt(rchisq(n2, dd)/dd)
                        y2 <- y2/zu
                        zu <- sqrt(rchisq(n3, dd)/dd)
                        y3 <- y3/zu
                }
                if(ytype == 4){  #Want pop coord med(y) = 0.
                        y1 <- exp(y1)
                        y1 <- y1 - 1
                        y2 <- exp(y2)
                        y2 <- y2 - 1
                        y3 <- exp(y3)
                        y3 <- y3 - 1
                }
                y1 <- y1 %*% A
                y2 <- y2 %*% A
```

```
      y2 <- sig2 * y2
      if( cov3I != T){
      y3 <- y3 %*% A
      y3 <- sig3 * y3}
      if(outliers == 1) {
          y1[1:val,  ] <- matrix(rnorm(val * m, sd = 0.01), ncol
                       = m, nrow = val)
              y1[1:val, m] <- y1[1:val, m] + z
      }
      if(outliers == 2) {
              y1[1:val,  ] <- matrix(rnorm(val * m, sd = 0.01), ncol
                       = m, nrow = val)
              y1[1:val, 1] <- y1[1:val, 1] + z
      }
      if(outliers == 3) {
              tem <- z + 0 * 1:m
              y1[1:val,  ] <- y1[1:val,  ] + tem
      }
      if(outliers == 4) {
              y1[1:val, m] <- z
      }
      if(outliers == 5) {
              y1[1:val, 1] <- z
      }
      y1 <- mu1 + y1
      y2 <- mu2 + y2
      y3 <- mu3 + y3
#clean y1 has mean mu1 =delta1(1,...,1)^T,
#y2 has mean delta2(1,..,1)^T and y3 has mean delta3 (1,...,1)^T
      #get bootstrapped Ty - Ty3, Ty2 - Ty3 for various statistics T
```

```
        for(j in 1:B){
        tem <- sample(indx1,n1,replace=T)
        medy <- apply(y1[tem,],2,median)
        mny <- apply(y1[tem,],2,mean)
        tmny <- apply(y1[tem,],2,tmn)
        tem <- sample(indx2,n2,replace=T)
        medy2 <- apply(y2[tem,],2,median)
        mny2 <- apply(y2[tem,],2,mean)
        tmny2 <- apply(y2[tem,],2,tmn)
        tem <- sample(indx3,n3,replace=T)
        medy3 <- apply(y3[tem,],2,median)
        mny3 <- apply(y3[tem,],2,mean)
        tmny3 <- apply(y3[tem,],2,tmn)
        medmus[j,] <- c(medy-medy3,medy2-medy3)
        mnmus[j,] <- c(mny-mny3,mny2-mny3)
        tmnmus[j,] <- c(tmny-tmny3,tmny2-tmny3)
    }
        outmed<-predreg(medmus)
        medcv <- medcv + outmed$inr
        outmn <- predreg(mnmus)
        mncv <- mncv + outmn$inr
        outtmn <- predreg(tmnmus)
        tmncv <- tmncv + outtmn$inr
        cutoffs[i,]<-c(outmed$cuplim,outmn$cuplim,outtmn$cuplim)^2

#Get the classical test coverage
yall<-rbind(y1,y2,y3)
yall<-as.matrix(yall)
gp <- c(rep(1, n1),rep(2, n2),rep(3, n3))
grp<-factor(gp)
```

```
        out<-manova(yall~grp)

        pval <- summary(out,test="Hotelling-Lawley")$stats[1,6]  #pvalue for
         #Hotelling-Lawley's test

                if(pval < 0.05){
                crej <- crej +1
                }
        #large sample Manova type test based on the sample mean
        pval <- mantyp(y=yall,p=3,group=grp)$pval
        if(pval < 0.05) mantcov <- mantcov + 1
        }
        medcv <- 1 - medcv/nruns #prop of times Ho is rejected
        mncv  <- 1 - mncv/nruns
        tmncv  <- 1 - tmncv/nruns
        mncut <- apply(cutoffs,2,mean)
        ccv <- crej/nruns
        mantcov <- mantcov/nruns

        list(chisqcut = chisqcut, mncut=mncut, medcv = medcv,
             mncv = mncv,tmncv=tmncv,ccv=ccv,mantcov=mantcov)
}
```

REFERENCES

[1] 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, 91-112.

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

[3] 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.

[4] Curran, J. M. (2013). Hotelling: Hotelling's T-squared test and variants. *R Package version 1.0-2,* (https://cran.r-project.org/package=Hotelling).

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

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

[7] 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.

[8] 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.

[9] Johnson, R. W. (1996). Fitting percentage of body fat to simple body measurements. *Journal of Statistics Education 4*, online at (www.amstat.org/publications/jse/).

[10] 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.

[11] 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.

[12] Olive, D. J. (2004). A resistant estimator of multivariate location and dispersion. *Computational Statistics & Data Analysis,* 46:99-102.

[13] Olive, D. J. (2013). Asymptotically optimal regression prediction intervals and prediction regions for multivariate data. *International Journal of Statistics and Probability,* 2:90-100.

[14] Olive, D. J. (2017a). *Applications of hyperellipsoidal prediction regions.* Statistical Papers. To appear.

[15] Olive, D. J. (2017b). *Robust Multivariate Analysis.* New York: Springer. To appear.

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

[17] Olive, D. J., Hawkins, D. M. (2010). Robust multivariate location and dispersion. Unpublished Manuscript available from (http://lagrange.math.siu.edu/Olive /pphbmld.pdf).

[18] Olive, D.J., Pelawa Watagoda, L.C.R., and Rupasinghe Arachchige Don, H.S. (2015), Visualizing and Testing the Multivariate Linear Regression Model, *International Journal of Statistics and Probability,* 4, 126-137.

[19] Press, S.J. (2005), *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*, 2nd ed., Dover, Mineola, NY.

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

[21] Ren, J.-J. (1991), On Hadamard Differentiability of Extended Statistical Functional, *Journal of Multivariate Analysis,* 39, 30-43.

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

[23] Rupasinghe Arachchige Don, H.S., and Pelawa Watagoda, L.C.R. (2017), Bootstrapping Analogs of the Two Sample Hotelling's T^2 Test, *Preprint at* (http://lagrange. math.siu.edu/Olive/stwosample.pdf).

[24] Todorov, V., Filzmoser, P. (2010). Robust statistics for the one-way MAVOVA. *Com-*

putational Statistics & Data Analysis, 54:37-48.

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

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

[27] Willems, G., Pison, G., Rousseeuw, P. J., Van Aelst, S. (2002). A robust "Hotelling" test. *Metrika,* 55:125-138.

[28] Zhang, J.-T., and Liu, X. (2013), A Modified Bartlett Test for Heteroscedastic One-Way MANOVA, *Metrika,* 76, 135–152.

[29] Zhang, J., Olive, D. J., Ye, P. (2012). Robust covariance matrix estimation with canonical correlation analysis. *International Journal of Statistics and Probability,* 1:119-136.

VITA

Graduate School
Southern Illinois University

Hasthika S. Rupasinghe Arachchige Don Date of Birth: August 15, 1985

1457 West Lake Road, Murphysboro, Illinois 62966

hasthika@siu.edu (hasthika@appstate.edu)

Southern Illinois University at Carbondale
Master of Science, Mathematics, August 2013

Special Honors and Awards: Dissertation Research Assistantship Award 2017.
John M. H. Olmsted Award for Outstanding Teaching Performance in the Department of Mathematics at SIU 2014.

Research Paper Title:
Bootstrapping Analogs of the One Way Manova Test

Major Professor: Dr. David J. Olive

Publications:

1. "Visualizing and Testing the Multivariate Linear Regression Model", International Journal of Statistics and Probability, January 22, 2015, with David J Olive, Lasanthi Watagoda.
2. "Bootstrapping analogs of the Hotelling's T^2 test", Communications in Statistics Theory and Methods, submitted, with Lasanthi Watagoda.
3. "Bootstrapping Analogs of the One Way Manova Test", work in progress, with David J Olive.