SOME SIMPLE HIGH DIMENSIONAL ONE AND TWO SAMPLE TESTS

by

Ahlam Mustafa Abid

B.Sc., Taibah University, 2010
M.S., University of Dayton, 2017

A Dissertation
Submitted in Partial Fulfillment of the Requirements for the
Doctor of Philosophy Degree

Department of Mathematics
in the Graduate School
Southern Illinois University Carbondale
Month, 2025

DISSERTATION APPROVAL

SOME SIMPLE HIGH DIMENSIONAL ONE AND TWO SAMPLE TESTS

By

Ahlam Mustafa Abid

A Dissertation Submitted in Partial

Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in the field of Mathematics

Approved by:

Dr. David Olive, Chair

Dr. Seyed Yaser Samadi

Dr. Michael Sullivan

Dr. Mathew Gluck

Dr. Mohammad Sayeh

Graduate School
Southern Illinois University Carbondale
(Date of Approval)

AN ABSTRACT OF THE DISSERTATION OF

Ahlam M. Abid, for the Doctor of Philosophy degree in Mathematics, presented on Month day, year, at Southern Illinois University Carbondale.

TITLE: SOME SIMPLE HIGH DIMENSIONAL ONE AND TWO SAMPLE TESTS

MAJOR PROFESSOR: Dr. David Olive

Consider testing $H_0 : \boldsymbol{\mu} = \mathbf{0}$ versus $H_A : \boldsymbol{\mu} \neq \mathbf{0}$ using a random sample $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ where the $\boldsymbol{x}_i$ are $p \times 1$ random vectors and p may be much larger than n. Several one sample tests use the same test statistic T_n with different estimators of the variance $V(T_n)$. Rather simple theory from U-statistics is used to find $V(T_n)$, resulting in an estimator that is quick to compute when H_0 is true. Some two sample tests for $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ are also considered.

DEDICATION

ACKNOWLEDGMENTS

I could not have reached this goal without the help of many people in my life. I would like to take this opportunity to thank them for their support.

First and foremost, words cannot express my sincere thanks, deepest gratitude ,and appreciation to my esteemed advisor and chair of my committee Dr. David Olive for his invaluable help, continuous support, infinite patience during my lowest moments. Your immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life. Thank you for being always supportive, helpful and available all of the time for any question or concerns. I am truly grateful for all of your tremendous assistance and insights leading to the writing of this paper. I am thankful, too, for professors of my defense committee, who generously provided support and proper suggestions.

Thereafter, I must express my deepest appreciation for my family. My father and mother, who supported my educational endeavors from the very beginning. Your motivating encouragement, steadfast emotional and financial support, and belief in my potentials have been instrumental in my success since I was a child.

Over and above that, I owe an immense debt of gratitude to my husband, Abdulmonem. Who has walked every step of this journey with me and always pushes me to reach for the stars. Your votes of confidence kept me going when my spirits dipped. I could not have undertaken this journey without you! Your sacrifice, deep understanding, unwavering support, and belief in my academic pursuits have been the foundation of my resilience and success.

Last but not least, I am profoundly grateful to my children, Ziyad, Eyad, Ali, Jawad, and my little princess Lina. Your unconditional love and belief in me has kept my spirits and motivation high during this journey. I would like to acknowledge the emotional support provided by you!

We made it to the top of the mountain! I look forward to celebrating with all of you!

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Consider testing $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ versus $H_A : \boldsymbol{\mu} \neq \boldsymbol{0}$ using independent and identically distributed (iid) $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ where the $\boldsymbol{x}_i$ are $p \times 1$ random vectors and p may be much larger than n. Assume the expected value $E(\boldsymbol{x}_i) = \boldsymbol{\mu}$ and nonsingular covariance matrix $Cov(\boldsymbol{x}_i) = \boldsymbol{\Sigma}$. Replace $\boldsymbol{x}_i$ by $\boldsymbol{w}_i = \boldsymbol{x}_i - \boldsymbol{\mu}_0$ to test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_A : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. This chapter reviews some tests while the following chapter gives simpler large sample theory for some of the tests, including a new test that has very simple large sample theory.

Suppose p is fixed, and consider testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ where a $g \times 1$ statistic T_n satisfies $\sqrt{n}(T_n - \boldsymbol{\theta}) \xrightarrow{D} \boldsymbol{u} \sim N_g(\boldsymbol{0}, \boldsymbol{\Sigma})$. If $\hat{\boldsymbol{\Sigma}}^{-1} \xrightarrow{P} \boldsymbol{\Sigma}^{-1}$ and H_0 is true, then

$$D_n^2 = D^2_{\boldsymbol{\theta}_0}(T_n, \hat{\boldsymbol{\Sigma}}/n) = n(T_n - \boldsymbol{\theta}_0)^T \hat{\boldsymbol{\Sigma}}^{-1}(T_n - \boldsymbol{\theta}_0) \xrightarrow{D} \boldsymbol{u}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{u} \sim \chi^2_g$$

as $n \to \infty$. Then a Wald type test rejects H_0 at significance level δ if $D_n^2 > \chi^2_{g,1-\delta}$ where $P(X \leq \chi^2_{g,1-\delta}) = 1 - \delta$ if $X \sim \chi^2_g$, a chi-square distribution with g degrees of freedom.

It is common to implement a Wald type test using

$$D_n^2 = D^2_{\boldsymbol{\theta}_0}(T_n, \boldsymbol{C}_n/n) = n(T_n - \boldsymbol{\theta}_0)^T \boldsymbol{C}_n^{-1}(T_n - \boldsymbol{\theta}_0) \xrightarrow{D} \boldsymbol{u}^T \boldsymbol{C}^{-1} \boldsymbol{u}$$

as $n \to \infty$ if H_0 is true, where the $g \times g$ symmetric positive definite matrix $\boldsymbol{C}_n \xrightarrow{P} \boldsymbol{C} \neq \boldsymbol{\Sigma}$. Hence $\boldsymbol{C}_n$ is the wrong dispersion matrix, and $\boldsymbol{u}^T \boldsymbol{C}^{-1} \boldsymbol{u}$ does not have a χ^2_g distribution when H_0 is true. Often $\boldsymbol{C}_n$ is a regularized estimator of $\boldsymbol{\Sigma}$, or $\boldsymbol{C}_n^{-1}$ is a regularized estimator of the precision matrix $\boldsymbol{\Sigma}^{-1}$, such as $\boldsymbol{C}_n = diag(\hat{\boldsymbol{\Sigma}})$ or $\boldsymbol{C}_n = \boldsymbol{I}_g$, the $g \times g$ identity matrix.

Rajapaksha and Olive (2024) showed how to bootstrap Wald tests with the wrong dispersion matrix. When $\boldsymbol{C}_n = \boldsymbol{I}_g$, the bootstrap tests often became conservative as g increased to n. For some of these tests, the m out of n bootstrap, which draws a sample of size m without replacement from the n, works better than the nonparametric bootstrap.

When n is much larger than p, the one sample Hotelling (1931) T^2 test is often used to test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_A : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. The sample mean

$$\overline{\boldsymbol{x}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i,$$

and the sample covariance matrix

$$\boldsymbol{S} = \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T = (S_{ij}).$$

That is, the ij entry of $\boldsymbol{S}$ is the sample covariance S_{ij}. If the $\boldsymbol{x}_i$ are iid with expected value $E(\boldsymbol{x}_i) = \boldsymbol{\mu}$ and nonsingular covariance matrix $Cov(\boldsymbol{x}_i) = \boldsymbol{\Sigma}$, then by the multivariate central limit theorem

$$\sqrt{n}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}).$$

If H_0 is true, then

$$T_H^2 = n(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)^T \boldsymbol{S}^{-1} (\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0) \xrightarrow{D} \chi^2_p.$$

The one sample Hotelling's T^2 test rejects H_0 if $T_H^2 > D^2_{1-\delta}$ where $D^2_{1-\delta} = \chi^2_{p,\delta}$ and $P(Y \leq \chi^2_{p,\delta}) = \delta$ if $Y \sim \chi^2_p$. Alternatively, use

$$D^2_{1-\delta} = \frac{(n-1)p}{n-p} F_{p,n-p,1-\delta}$$

where $P(Y \leq F_{p,d,\delta}) = \delta$ if $Y \sim F_{p,d}$. The scaled F cutoff can be used since $T_H^2 \xrightarrow{D} \chi^2_p$ if H_0 holds, and

$$\frac{(n-1)p}{n-p} F_{p,n-p,1-\delta} \to \chi^2_{p,1-\delta}$$

as $n \to \infty$.

The next two high dimensional tests are described in Srivastava and Du (2008). Also see Hu and Bai (2015). Let $tr(\boldsymbol{A})$ be the trace of square matrix $\boldsymbol{A}$. Let $\boldsymbol{R}$ be the sample correlation matrix.

Consider testing $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ versus $H_A : \boldsymbol{\mu} \neq \boldsymbol{0}$. Let $\boldsymbol{D} = diag(\boldsymbol{S})$. Let

$$c_{p,n} = 1 + \frac{tr(\boldsymbol{R}^2)}{p^{3/2}}.$$

Let $n = O(p^\delta)$ where $0.5 < \delta \leq n$. Then under regularity conditions

$$Z_1 = \frac{n\overline{\boldsymbol{x}}^T \boldsymbol{D}^{-1}\overline{\boldsymbol{x}} - \frac{(n-1)p}{n-3}}{\sqrt{2\left(tr(\boldsymbol{R}^2) - \frac{p^2}{n-1}\right) c_{p,n}}} \xrightarrow{D} N(0,1)$$

as $n, p \to \infty$. The next test is attributed to Bai and Saranadasa (1996). Under regularity conditions,

$$Z_2 = \frac{n\overline{\boldsymbol{x}}^T\overline{\boldsymbol{x}} - tr(\boldsymbol{S})}{\left[\frac{2(n-1)n}{(n-2)(n+1)}\left(tr(\boldsymbol{S}^2) - \frac{1}{n}[tr(\boldsymbol{S})]^2\right)\right]^{1/2}} \xrightarrow{D} N(0,1)$$

as $n, p \to \infty$. Both of these test statistics used $p/n \to c > 0$ or $p/n^2 \to 0$.

Note that $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ holds if and only if $\|\boldsymbol{\mu}\|^2 = \boldsymbol{\mu}^T\boldsymbol{\mu} = 0$. The T_n in Equation (1.1) below can be viewed as a modification of $\|\overline{\boldsymbol{x}}\|^2 = \overline{\boldsymbol{x}}^T\overline{\boldsymbol{x}}$ that is a better estimator of $\boldsymbol{\mu}^T\boldsymbol{\mu}$ in high dimensions. Note that $E(\boldsymbol{x}_i^T\boldsymbol{x}_j) = \boldsymbol{\mu}^T\boldsymbol{\mu}$ if $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are iid with $E(\boldsymbol{x}_i) = \boldsymbol{\mu}$ and $i \neq j$. Let $V(T_n)$ be the variance of T_n and let $s_n^2 = \hat{V}(T_n)$ be a consistent estimator of $V(T_n)$.

The following test is due to Chen and Qin (2010). Also see Hu and Bai (2015). Let $\boldsymbol{a} = \sum_{i=1}^n \boldsymbol{x}_i$ and let $\boldsymbol{X} = (x_{ij})$ be the data matrix with ith row $= \boldsymbol{x}_i^T$ and ij element $= x_{ij}$. Let $vec(\boldsymbol{A})$ stack the columns of matrix $\boldsymbol{A}$ so that $\boldsymbol{c} = vec(\boldsymbol{X}^T) = [\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, ..., \boldsymbol{x}_n^T]^T$. Then

$$\boldsymbol{c}^T\boldsymbol{c} = \sum_{i=1}^n \boldsymbol{x}_i^T\boldsymbol{x}_i = \sum_{i=1}^n \|\boldsymbol{x}_i\|^2 = \sum_{i=1}^n\sum_{j=1}^p (x_{ij})^2.$$

Let

$$T_n = \frac{1}{n(n-1)}[\boldsymbol{a}^T\boldsymbol{a} - \boldsymbol{c}^T\boldsymbol{c}] = \frac{1}{n(n-1)}\sum\sum_{i\neq j} \boldsymbol{x}_i^T\boldsymbol{x}_j = \frac{1}{n(n-1)}\sum_{i\neq j} \boldsymbol{x}_i^T\boldsymbol{x}_j. \tag{1.1}$$

The terms in $\boldsymbol{c}^T\boldsymbol{c} = \sum_{i=1}^n \boldsymbol{x}_i^T\boldsymbol{x}_i$ are the terms that cause the restriction on p for asymptotic normality for the previous two tests. Under $H_0 : \boldsymbol{\mu} = \mathbf{0}$ and additional regularity conditions,

$$\frac{T_n}{\sqrt{V(T_n)}} \xrightarrow{D} N(0,1) \;\; \text{and} \;\; \frac{\mathrm{T_n}}{\mathrm{s_n}} \xrightarrow{\mathrm{D}} \mathrm{N}(0,1) \tag{1.2}$$

where s_n is rather hard to compute. Here

$$s_n^2 = \frac{2}{n(n-1)} tr \left[\sum_{i \neq j} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{(i,j)}) \boldsymbol{x}_i^T (\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{(i,j)}) \boldsymbol{x}_j^T \right]$$

is a consistent estimator of $V(T_n)$ where $\overline{\boldsymbol{x}}_{(i,j)}$ is the sample mean computed without $\boldsymbol{x}_i$ or $\boldsymbol{x}_j$:

$$\overline{\boldsymbol{x}}_{(i,j)} = \frac{1}{n-2} \sum_{k \neq i,j} \boldsymbol{x}_k.$$

As noted by Park and Ayyala (2013), $nT_n = n\overline{\boldsymbol{x}}^T\overline{\boldsymbol{x}} - tr(\boldsymbol{S})$. This result holds since

$$T_n = \frac{1}{n(n-1)} \left[\sum_i \sum_j \boldsymbol{x}_i^T \boldsymbol{x}_j - \sum_i \boldsymbol{x}_i^T \boldsymbol{x}_i \right] = \frac{n^2 \overline{\boldsymbol{x}}^T \overline{\boldsymbol{x}} - \sum_i \boldsymbol{x}_i^T \boldsymbol{x}_i}{n(n-1)}.$$

Now

$$\boldsymbol{S} = \frac{1}{n-1} \left[\sum_i \boldsymbol{x}_i \boldsymbol{x}_i^T - n \overline{\boldsymbol{x}}\, \overline{\boldsymbol{x}}^T \right].$$

Thus

$$tr(\boldsymbol{S}) = \frac{1}{n-1} \left[\sum_i tr(\boldsymbol{x}_i \boldsymbol{x}_i^T) - n tr(\overline{\boldsymbol{x}}\, \overline{\boldsymbol{x}}^T) \right] = \frac{1}{n-1} \left[\sum_i \boldsymbol{x}_i^T \boldsymbol{x}_i - n \overline{\boldsymbol{x}}^T \overline{\boldsymbol{x}} \right].$$

Thus

$$n\overline{\boldsymbol{x}}^T\overline{\boldsymbol{x}} - tr(\boldsymbol{S}) = n\overline{\boldsymbol{x}}^T\overline{\boldsymbol{x}} + \frac{n}{n-1}\overline{\boldsymbol{x}}^T\overline{\boldsymbol{x}} - \frac{1}{n-1}\sum_i \boldsymbol{x}_i^T \boldsymbol{x}_i = \frac{n^2 \overline{\boldsymbol{x}}^T \overline{\boldsymbol{x}} - \sum_i \boldsymbol{x}_i^T \boldsymbol{x}_i}{n-1}.$$

We will also consider replacing $\boldsymbol{x}_i$ by $z_i = ss(\boldsymbol{x}_i)$ where the spatial sign function $ss(\boldsymbol{x}_i) = \mathbf{0}$ if $\boldsymbol{x}_i = \mathbf{0}$, and $ss(\boldsymbol{x}_i) = \boldsymbol{x}_i / \|\boldsymbol{x}_i\|$ otherwise. This function projects the nonzero $\boldsymbol{x}_i$ onto the unit p-dimensional hypersphere centered at $\mathbf{0}$. Let $T_n(\boldsymbol{w})$ denote the statistic T_n computed from an iid sample $\boldsymbol{w}_1, ..., \boldsymbol{w}_n$. Since the z_i are iid if the $\boldsymbol{x}_i$ are iid, use $T_n(z)$ to test $H_0 : \boldsymbol{\mu}_{\boldsymbol{z}} = \mathbf{0}$ versus $H_A : \boldsymbol{\mu}_{\boldsymbol{z}} \neq \mathbf{0}$ where $\boldsymbol{\mu}_{\boldsymbol{z}} = E(z_i)$. In general, $\boldsymbol{\mu}_{\boldsymbol{z}} \neq \boldsymbol{\mu} = \boldsymbol{\mu}_{\boldsymbol{x}} = E(\boldsymbol{x}_i)$, but $\boldsymbol{\mu}_{\boldsymbol{z}} = \boldsymbol{\mu} = \mathbf{0}$ can occur if the

$\boldsymbol{x}_i$ have a lot of symmetry about $\boldsymbol{0}$. In particular, $\boldsymbol{\mu}_{\boldsymbol{z}} = \boldsymbol{\mu} = \boldsymbol{0}$ if the $\boldsymbol{x}_i$ are iid from an elliptically contoured distribution with center $\boldsymbol{\mu} = \boldsymbol{0}$. The test based on the statistic $T_n(z)$ can be useful if the second moment of the $\boldsymbol{x}_i$ does not exist, for example if the $\boldsymbol{x}_i$ are iid from a multivariate Cauchy distribution. These results may be useful for understanding papers such as Wang, Peng, and Li (2015)

Chapter 2 considers two estimators s_n^2 of $V(T_n)$ that are easier to compute when H_0 is true, and gives a new test with very simple large sample theory. Chapter 3 considers two sample tests.

CHAPTER 2

ESTIMATING $V(T_N)$

Some notation for the simple test is needed. Assume $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ are iid, $E(\boldsymbol{x}_i) = \boldsymbol{\mu}$ and the variance $V(\boldsymbol{x}_i^T \boldsymbol{x}_j) = \sigma_W^2$ for $i \neq j$. Let m = floor(n/2) = $\lfloor n/2 \rfloor$ be the integer part of $n/2$. So floor(100/2) = floor(101/2) = 50. Let the iid random variables $W_i = \boldsymbol{x}_{2i-1}^T \boldsymbol{x}_{2i}$ for $i = 1, ..., m$. Hence $W_1, W_2, ..., W_m = \boldsymbol{x}_1^T \boldsymbol{x}_2, \boldsymbol{x}_3^T \boldsymbol{x}_4, ..., \boldsymbol{x}_{2m-1}^T \boldsymbol{x}_{2m}$. Note that $E(W_i) = \boldsymbol{\mu}^T \boldsymbol{\mu}$ and $V(W_i) = \sigma_W^2$. Let S_W^2 be the sample variance of the W_i:

$$S_W^2 = \frac{1}{m-1} \sum_{i=1}^m (W_i - \overline{W})^2.$$

If $\sigma_W^2 \propto \tau^2 p$ where $p > n$, then n may not be large enough for the normal approximation to hold. The following theorem follows from the univariate central limit theorem.

Theorem 1. Assume $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ are iid, $E(\boldsymbol{x}_i) = \boldsymbol{\mu}$, and the variance $V(\boldsymbol{x}_i^T \boldsymbol{x}_j) = \sigma_W^2$ for $i \neq j$. Let $W_1, ..., W_m$ be defined as above. Then
a) $\sqrt{m}(\overline{W} - \boldsymbol{\mu}^T \boldsymbol{\mu}) \xrightarrow{D} N(0, \sigma_W^2)$.

$$b) \ \frac{\sqrt{m}(\overline{W} - \boldsymbol{\mu}^T \boldsymbol{\mu})}{S_W} \xrightarrow{D} N(0, 1)$$

as $n \to \infty$.

The following theorem derives $V(T_n)$ under much simpler regularity conditions than those in the literature, and the proof of the theorem is also simpler. For example, Li (2023) finds $V(T_n)$ when H_0 is true, using much stronger regularity conditions than in Theorem 2. In the simulations, we use a variant of the Li (2023) variance estimator $\hat{\sigma}_W^2$, and also use the estimator S_W^2 that is much easier to compute.

Theorem 2. Assume $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ are iid, $E(\boldsymbol{x}_i) = \boldsymbol{\mu}$, and the variance $V(\boldsymbol{x}_i^T \boldsymbol{x}_j) = \sigma_W^2$ for $i \neq j$. Let $W_{ij} = \boldsymbol{x}_i^T \boldsymbol{x}_j$ for $i \neq j$. Let $\theta = Cov(W_{ij}, W_{id}) = \boldsymbol{\mu}^T \boldsymbol{\Sigma} \boldsymbol{\mu}$ where $j \neq d$, $i < j$, and $i < d$. Then

$$a) \ V(T_n) = \frac{2\sigma_W^2}{n(n-1)} + \frac{4(n-2)\theta}{n(n-1)}.$$

b) If $H_0 : \boldsymbol{\mu} = \mathbf{0}$ is true, then $\theta = 0$ and

$$V_0 = V(T_n) = \frac{2\sigma_W^2}{n(n-1)}.$$

Proof. a) To find the variance $V(T_n)$ with T_n from Equation (1.1), let $W_{ij} = \boldsymbol{x}_i^T \boldsymbol{x}_j = W_{ji}$, and note that

$$T_n = \frac{2}{n(n-1)} H_n \ \text{ where } \ \mathrm{H_n} = \sum_{\mathrm{i} <} \sum_{\mathrm{j}} \boldsymbol{x}_{\mathrm{i}}^{\mathrm{T}} \boldsymbol{x}_{\mathrm{j}} = \sum_{\mathrm{i<j}} \boldsymbol{x}_{\mathrm{i}}^{\mathrm{T}} \boldsymbol{x}_{\mathrm{j}}.$$

Then $V(H_n) = Cov(H_n, H_n) =$

$$Cov\left(\sum_{i<} \sum_{j} W_{ij}, \sum_{k<} \sum_{d} W_{kd}\right) = \sum_{i<} \sum_{j} \sum_{k<} \sum_{d} Cov(W_{ij}, W_{kd}). \tag{2.1}$$

Let $V(W_{ij}) = \sigma_W^2$ for $i \neq j$. The covariances are of 3 types. First, if $(ij) = (kd)$ with $i < j$, then $Cov(W_{ij}, W_{kd}) = V(W_{ij}) = \sigma_W^2$. Second, if i, j, k, d are distinct with $i < j$ and $k < d$, then W_{ij} and W_{kd} are independent with $Cov(W_{ij}, W_{kd}) = 0$. Third, there are terms where exactly three of the four subscripts are distinct, which have $Cov(W_{ij}, W_{id}) = \theta$ where $j \neq d$, $i < j$, and $i < d$ or $Cov(W_{ij}, W_{kj}) = \theta$ where $i \neq k$, $i < j$, and $k < j$. These covariance terms are all equal to the same number θ since $W_{ij} = W_{ji}$. The number of ways to get three distinct subscripts is

$$a - b - c = \binom{n}{2}^2 - \binom{n}{2}\binom{n-2}{2} - \binom{n}{2} = n(n-1)(n-2)$$

since a is the number of terms on the right hand side of (2.1), b is the number of terms where i, j, k, d are distinct with $i < j$ and $k < d$, and c is the number of terms where $(ij) = (kd)$ with $i < j$. [Note that $n(n-1)$ terms have i and j distinct. Half of these terms have $i < j$ and half have $i > j$. Similarly, $n(n-1)(n-2)(n-3)$ terms have $ijkd$ distinct, and half of the $n(n-1)$ terms have $i < j$, while half of the $(n-2)(n-3)$ terms have $k < d$.] Thus

$$V(H_n) = 0.5n(n-1)\sigma_W^2 + n(n-1)(n-2)\theta.$$

This calculation was adapted from Lehmann (1975, pp. 336-337). Thus

$$V(T_n) = \frac{4}{[n(n-1)]^2} V(H_n) = \frac{2\sigma_W^2}{n(n-1)} + \frac{4(n-2)\theta}{n(n-1)}.$$

b) Now $\theta = Cov(\boldsymbol{x}_i^T \boldsymbol{x}_j, \boldsymbol{x}_i^T \boldsymbol{x}_k)$ where $\boldsymbol{x}_i, \boldsymbol{x}_j$, and $\boldsymbol{x}_k$ are iid. Hence $\theta =$

$$Cov(\sum_d x_{id} x_{jd}, \sum_t x_{it} x_{kt}) = \sum_d \sum_t Cov(x_{id} x_{jd}, x_{it} x_{kt}) =$$

$$\sum_d \sum_t [E(x_{id} x_{jd} x_{it} x_{kt}) - E(x_{id} x_{jd}) E(x_{it} x_{kt})] =$$

$$\sum_d \sum_t [E(x_{id} x_{it}) E(x_{jd}) E(x_{kt}) - E(x_{id}) E(x_{jd}) E(x_{it}) E(x_{kt})] =$$

$$\sum_d \sum_t [E(x_{jd}) E(x_{kt}) (E(x_{id} x_{it}) - E(x_{id}) E(x_{it}))] =$$

$$\sum_d \sum_t [E(x_{jd}) E(x_{kt})\, Cov(x_{id}, x_{it})] = \boldsymbol{\mu}^T \boldsymbol{\Sigma} \boldsymbol{\mu}.$$

Under H_0, $\boldsymbol{\mu} = \mathbf{0}$ and thus $\theta = 0$. □

Srivastava and Du (2008), Bai and Saranadasa (1996), Chen and Qin (2010), and others use $T_n / \sqrt{\hat{V}(T_n)} \xrightarrow{D} N(0,1)$, while Li (2023) uses $T_n / \sqrt{\hat{V}_0(T_n)} \xrightarrow{D} N(0,1)$. Theorem 2 and the following result show that the second statistic has more power. Adapting an argument from Lehmann (1999, pp. 367-368), let $Z(\boldsymbol{a}) = E(\boldsymbol{a}^T \boldsymbol{x}_j) = \boldsymbol{a}^T \boldsymbol{\mu}$. Then it can be shown that $\theta = V(Z(\boldsymbol{x}_i)) = V(\boldsymbol{x}_i^T \boldsymbol{\mu}) \geq 0$. Also, by Theorem 2, $\theta = \boldsymbol{\mu}^T \boldsymbol{\Sigma} \boldsymbol{\mu} \geq 0$. Let $s_n^2 = \hat{V}$ be a consistent estimator of $V(T_n)$ and let

$$\hat{V}_0 = \frac{2\hat{\sigma}_W^2}{n(n-1)}.$$

The test statistics

$$t_1 = \frac{T_n}{\sqrt{\hat{V}_0}} \xrightarrow{D} N(0,1) \text{ and } \mathrm{t}_2 = \frac{\mathrm{T_n}}{\sqrt{\hat{V}}} \xrightarrow{\mathrm{D}} \mathrm{N}(0,1)$$

if $H_0 : \boldsymbol{\mu} = \mathbf{0}$ is true. However, when H_0 is not true,

$$\hat{V} \approx \hat{V}_0 + \frac{4(n-2)\hat{\theta}}{n(n-1)}$$

where the second term is positive. If H_0 is not true and n and p are such that the second term dominates, then $|t_1|$ tends to be proportional to $\sqrt{n}|t_2|$, greatly increasing the power of the test that uses t_1.

For power, we expect $V_0(T_n) \to 0$ if $p/n^2 \to 0$ as $n \to \infty$. The high dimensional literature often gives very strong regularity conditions where $V(T_n) \to 0$ if $p = p_n = n^\gamma$ where γ is often much larger than 0.5 and $\boldsymbol{\mu} = \mathbf{0}$. Suppose $\boldsymbol{\mu} = \delta\mathbf{1}$ where the constant $\delta > 0$ and $\mathbf{1}$ is the $p \times 1$ vector of ones. Then $\boldsymbol{\mu}^T\boldsymbol{\mu} = \delta^2 p$, and the test using $\hat{V}_0(T_n)$ may have good power for $T_n / \sqrt{\hat{V}_0(T_n)} > 1.96 \approx 2$ or for

$$\frac{\delta^2 \ p}{\sqrt{\frac{2\sigma_W^2}{n(n-1)}}} > 2 \ \text{ or } \ \delta^2 > \frac{2\sqrt{2}\,\sigma_{\mathrm{W}}}{\mathrm{n\ p}}.$$

The above theory can also be applied to the $z_i = ss(\boldsymbol{x}_i)$ to test $H_0 : E(z) = \mathbf{0}$. As noted near the end of Chapter 1, for elliptically contoured distributions, $E(z) = \boldsymbol{\mu}_{\boldsymbol{z}} = \mathbf{0}$ if $E(\boldsymbol{x}) = \boldsymbol{\mu} = \boldsymbol{\mu}_{\boldsymbol{x}} = \mathbf{0}$.

The nonparametric bootstrap draws a bootstrap data set $\boldsymbol{x}_1^*, ..., \boldsymbol{x}_n^*$ with replacement from the $\boldsymbol{x}_i$ and computes T_1^* by applying T_n on the bootstrap data set. This process is repeated B times to get a bootstrap sample $T_1^*, ..., T_B^*$. For the statistic T_n, the nonparametric bootstrap fails in high dimensions because terms like $\boldsymbol{x}_j^T \boldsymbol{x}_j$ need to be avoided, and the nonparametric bootstrap has replicates: the proportion of cases in the bootstrap sample that are not replicates is about $1 - e^1 \approx 2/3 \approx 7/11$. The m out of n bootstrap draws a sample of size m without replacement from the n cases. For $B = 1$, this is a data splitting estimator, and $T_m^* \approx N(0, s_m^2)$ for large enough m and p. Sampling without replacement is also known as subsampling and the delete d jackknife.

Theory for subsampling is given by Politis and Romano (1994) and Wu (1990). Subsampling tends to work well for a large variety of statistics if $m/n \to 0$ with $m \to \infty$. A linear statistic has the form

$$\frac{1}{n} \sum_{i=1}^{n} t(U_i)$$

where $\theta = E[t(U_i)]$ and the U_i are iid. For a linear statistic, subsampling tends to work well if $m/n \to \tau \in [0, 1)$ with $m \to \infty$. For the $W_i = U_i$ in Theorem 1, $t(U_i) = U_i = \boldsymbol{x}_{2i-1}^T \boldsymbol{x}_{2i}$. If different blocks were taken such that the W_i are still iid, then subsampling would still work, but the statistics from the different blocks are estimating the same quantiles. Hence subsampling from all of the data may also work well. That is, subsampling may work well for a U-statistic that is the analog of a linear statistic. Using $m = floor(2n/3)$ worked well in simulations.

Now let W_i be an indicator random variable with $W_i = 1$ if $\boldsymbol{x}_i^*$ is in the sample and $W_i = 0$, otherwise, for $i = 1, ..., n$. The W_i are binary and identically distributed, but not independent. Hence $P(W_i = 1) = m/n$. Let $W_{ij} = W_i W_j$ with $i \neq j$. Again, the W_{ij} are binary and identically distributed. $P(W_{ij} = 1) =$ P(ordered pair $(\boldsymbol{x}_i, \boldsymbol{x}_j)$) was selected in the sample. Hence $P(W_{ij} = 1) = m(m-1)/[n(n-1)]$ since $m(m-1)$ ordered pairs were selected out of $n(n-1)$ possible ordered pairs. Then

$$T_m^* = \frac{1}{m(m-1)} \sum \sum_{k \neq d} \boldsymbol{x}_{i_k}^T \boldsymbol{x}_{i_d} = \frac{1}{m(m-1)} \sum \sum_{i \neq j} W_i W_j \boldsymbol{x}_i^T \boldsymbol{x}_j$$

where the $\boldsymbol{x}_{i_1}, ..., \boldsymbol{x}_{i_m}$ are the m vectors $\boldsymbol{x}_i$ selected in the sample. The first double sum has $m(m-1)$ terms while the second double sum has $n(n-1)$ terms. Hence

$$E(T_m^*) = \frac{1}{m(m-1)} \sum \sum_{i \neq j} E[W_i W_j] \boldsymbol{x}_i^T \boldsymbol{x}_j = T_n.$$

See similar calculations in Buja and Stuetzle (2006). Note that $V(T_m^*) = E([T_m^*]^2) - [T_n]^2 = Cov(T_m^*, T_m^*)$.

CHAPTER 3

TWO SAMPLE TESTS

If $(\boldsymbol{x}_{1i}, \boldsymbol{x}_{2i})$ come in correlated pairs, a high dimensional analog of the paired t test applies the one sample test on $\boldsymbol{z}_i = \boldsymbol{x}_{1i} - \boldsymbol{x}_{2i}$.

Now suppose there are two independent random samples $\boldsymbol{x}_{1,1}, ..., \boldsymbol{x}_{1,n_1}$ and $\boldsymbol{x}_{2,1}, ..., \boldsymbol{x}_{2,n_2}$ 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 $E(\boldsymbol{x}_i) = \boldsymbol{\mu}_i$ are $p \times 1$ vectors. Let $n = n_1 + n_2$. Let $\boldsymbol{S}_i$ be the sample covariance matrix of $\boldsymbol{x}_i$ and let $Cov(\boldsymbol{x}_i) = \boldsymbol{\Sigma}_i$ for $i = 1, 2$.

The classical two sample Hotelling's T^2 test uses

$$T_C^2 = (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)^T \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \hat{\boldsymbol{\Sigma}}_{pool} \right]^{-1} (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)$$

where

$$\hat{\boldsymbol{\Sigma}}_{pool} = \frac{(n_1 - 1)\boldsymbol{S}_1 + (n_2 - 1)\boldsymbol{S}_2}{n - 2}.$$

Then reject H_0 if $T_C^2 > mF_{m,n-2,1-\alpha}$.

The large sample test uses

$$T_L^2 = (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)^T \left(\frac{\boldsymbol{S}_1}{n_1} + \frac{\boldsymbol{S}_2}{n_2} \right)^{-1} (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2).$$

Let $d_n = \min(n_1 - p, n_2 - p)$. Then reject H_0 if $T_L^2 > mF_{m,d_n,1-\alpha}$.

Note that $T_C^2 \approx T_L^2$ if $n_1 \approx n_2 \geq 20p$ and the two tests are asymptotically equivalent if $n_i/n \to 0.5$ as $n_1, n_2 \to \infty$. If the n_i/n are not close to 0.5, then the test based on T_C^2 is useful if $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, a very strong assumption. Rajapaksha and Olive (2024) show how to get a bootstrap test based on T_C^2 where the assumption $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ is not needed.

There are test statistics T_n for testing $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ where p can be much larger than n with

$$\frac{T_n}{s_n} \xrightarrow{D} N(0,1)$$

where T_n is relatively simple to compute while s_n is much harder to compute. A simple test takes $m = \min(n_1, n_2)$ and $\boldsymbol{z}_i = \boldsymbol{x}_{1i} - \boldsymbol{x}_{2i}$ for $i = 1, ..., m$. Then apply the one sample test from Theorem 2 to the $\boldsymbol{z}_i$. This test might work well in high dimensions because of the superior power of the Theorem 2 test, but in low dimensions, it is known that there are better tests.

Let $\boldsymbol{x}_1$ be the $\boldsymbol{x}_i$ that has $n_1 \leq n_2$. Then let

$$\boldsymbol{y}_i = \boldsymbol{x}_{1i} - \sqrt{\frac{n_1}{n_2}}\boldsymbol{x}_{2i} + \frac{1}{\sqrt{n_1 n_2}} \sum_{j=1}^{n_1} \boldsymbol{x}_{2j} - \overline{\boldsymbol{x}}_2 = \boldsymbol{x}_{1i} - \sqrt{\frac{n_1}{n_2}}\boldsymbol{x}_{2i} + \boldsymbol{a}_{n_1,n_2} - \overline{\boldsymbol{x}}_2$$

for $i = 1, ..., n_1$. Note that $\boldsymbol{y}_i = \boldsymbol{z}_i = \boldsymbol{x}_{1i} - \boldsymbol{x}_{2i}$ if $n_1 = n_2$. Anderson (1984, pp. 177-178) proved that $\overline{\boldsymbol{y}} = \overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2$, that $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$ are uncorrelated for $i \neq j$, that $E(\boldsymbol{y}_i) = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$, and that $Cov(\boldsymbol{y}_i) = Cov(\boldsymbol{x}_1) + (n_1/n_2)Cov(\boldsymbol{x}_2)$ for $i = 1, ..., n_1$. Li (2023) showed that $T_n(\boldsymbol{y})/\sqrt{\hat{V}_0(\boldsymbol{y})} \xrightarrow{D} N(0,1)$ where the $\boldsymbol{y}$ denotes that the one sample test was computed using the $\boldsymbol{y}_i$.

Note that $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ holds if and only if $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2 = \boldsymbol{\mu}_1^T\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2^T\boldsymbol{\mu}_2 - 2\boldsymbol{\mu}_1^T\boldsymbol{\mu}_2$. These terms can be estimated by $T_n = T_n(\boldsymbol{x}, \boldsymbol{y}) = T_1 + T_2 - 2T_3$ where T_1 and T_2 are the one sample test statistic applied to samples 1 and 2 and $n_1 n_2 T_3 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \boldsymbol{x}_{1i}^T \boldsymbol{x}_{2j}$. Let $\boldsymbol{a} = \sum_{i=1}^{n_1} \boldsymbol{x}_{1i}$ and let $\boldsymbol{X}_1 = (x_{1ij})$ be the data matrix with ith row $= \boldsymbol{x}_{1i}^T$ and ij element $= x_{1ij}$. Let $\boldsymbol{c} = vec(\boldsymbol{X}_1^T) = [\boldsymbol{x}_{11}^T, \boldsymbol{x}_{12}^T, ..., \boldsymbol{x}_{1n_1}^T]^T$. Then

$$\boldsymbol{c}^T\boldsymbol{c} = \sum_{i=1}^{n1} \boldsymbol{x}_{1i}^T \boldsymbol{x}_{1i} = \sum_{i=1}^{n_1} \|\boldsymbol{x}_{1i}\|^2 = \sum_{i=1}^{n_1} \sum_{j=1}^{p} (x_{1ij})^2.$$

Let $\boldsymbol{b} = \sum_{i=1}^{n_2} \boldsymbol{x}_{2i}$ and let $\boldsymbol{X}_2 = (x_{2ij})$ be the data matrix with ith row $= \boldsymbol{x}_{2i}^T$ and ij element $= x_{2ij}$. Let $\boldsymbol{d} = vec(\boldsymbol{X}_2^T) = [\boldsymbol{x}_{21}^T, \boldsymbol{x}_{22}^T, ..., \boldsymbol{x}_{2n_2}^T]^T$. Then

$$\boldsymbol{d}^T\boldsymbol{d} = \sum_{i=1}^{n_2} \boldsymbol{x}_{2i}^T \boldsymbol{x}_{2i} = \sum_{i=1}^{n_2} \|\boldsymbol{x}_{2i}\|^2 = \sum_{i=1}^{n_2} \sum_{j=1}^{p} (x_{2ij})^2.$$

Thus

$$T_n = T_1 + T_2 - 2T_3 = \frac{1}{n_1(n_1-1)}[\boldsymbol{a}^T\boldsymbol{a} - \boldsymbol{c}^T\boldsymbol{c}] + \frac{1}{n_2(n_2-1)}[\boldsymbol{b}^T\boldsymbol{b} - \boldsymbol{d}^T\boldsymbol{d}] - \frac{2\boldsymbol{a}^T\boldsymbol{b}}{n_1 n_2}.$$

The terms in $\boldsymbol{c}^T\boldsymbol{c}$ and $\boldsymbol{d}^T\boldsymbol{d}$ are the terms that cause the restriction on p for asymptotic normality. Under $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ and additional regularity conditions,

$$\frac{T_n}{s_n} \xrightarrow{D} N(0,1)$$

where s_n is rather hard to compute. See Hu and Bai (2015) and Chen and Qin (2010).

CHAPTER 4

$V(T_N)$ FOR TWO SAMPLE TESTS

Let $n = n_1$ and $m = n_2$. Let $N = n + m$ and assume $n/N \to \pi_1 \in (0,1)$ while $m/N \to \pi_2 \in (0,1)$. In Theorem 2, $V_0(T_n) \propto 1/n^2$ while $V(T_n) \propto 1/n$ if $\boldsymbol{\mu} \neq \mathbf{0}$, resulting in a large increase in power compared to tests that use $V(T_n)$. For the Chen and Qin (2010) two sample test with $T_n = T_n(\boldsymbol{x}, \boldsymbol{y})$, Conjecture 1 suggests that $V_0(T_n) \propto 1/N^2$ and $V(T_n) \propto 1/N$. However, our programs do not simulate well. The Li (2023) test also has $V_0(T_n) \propto 1/N^2$.

Conjecture 1. Assume $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ and $\boldsymbol{y}_1, ..., \boldsymbol{y}_m$ are two independent random samples from two different populations or groups, $E(\boldsymbol{x}_i) = \boldsymbol{\mu_x}$ and $E(\boldsymbol{y}_j) = \boldsymbol{\mu_y}$ and the variances $V(\boldsymbol{x}_i^T \boldsymbol{x}_j) = \sigma_X^2$ for $i \neq j$ and $V(\boldsymbol{y}_i^T \boldsymbol{y}_j) = \sigma_Y^2$ for $i \neq j$ and $V(\boldsymbol{x}_i^T \boldsymbol{y}_j) = \sigma_Z^2$ for $i = 1, ..., n$ and $j = 1, ..., m$. Let $X_{ij} = \boldsymbol{x}_i^T \boldsymbol{x}_j$ for $i \neq j$, $Y_{ij} = \boldsymbol{y}_i^T \boldsymbol{y}_j$ for $i \neq j$ and $Z_{ij} = \boldsymbol{x}_i^T \boldsymbol{y}_j$ for $i = 1, ..., n$ and $j = 1, ..., n$. Let $\theta_1 = Cov(X_{ij}, X_{it}) = \boldsymbol{\mu_x}^T \boldsymbol{\Sigma_x} \boldsymbol{\mu_x}$ where $j \neq t$, $i < j$, and $i < t$, $\theta_2 = Cov(Y_{ij}, Y_{it}) = \boldsymbol{\mu_y}^T \boldsymbol{\Sigma_y} \boldsymbol{\mu_y}$ where $j \neq t$, $i < j$, and $i < t$, $\theta_3 = Cov(Z_{ij}, Z_{it}) = \boldsymbol{\mu_y}^T \boldsymbol{\Sigma_x} \boldsymbol{\mu_y}$ where $j \neq t$, $\theta_4 = Cov(Z_{ij}, Z_{kj}) = \boldsymbol{\mu_x}^T \boldsymbol{\Sigma_y} \boldsymbol{\mu_x}$ where $i \neq k$, $\theta_5 = Cov(X_{ij}, Z_{ab}) = \boldsymbol{\mu_x}^T \boldsymbol{\Sigma_x} \boldsymbol{\mu_y}$ where $a = i$ or $a = j$, and $i < j$, $\theta_6 = Cov(Y_{ij}, Z_{ab}) = \boldsymbol{\mu_y}^T \boldsymbol{\Sigma_y} \boldsymbol{\mu_x}$ where $b = i$ or $b = j$, and $i < j$. Then

$$a)\ V(T_n) = \frac{2\sigma_X^2}{n(n-1)} + \frac{2\sigma_Y^2}{m(m-1)} + \frac{4\sigma_Z^2}{nm} + \frac{4(n-2)\theta_1}{n(n-1)} + \frac{4(m-2)\theta_2}{m(m-1)} + \frac{4(m-1)\theta_3}{nm}$$
$$+ \frac{4(n-1)\theta_4}{nm} - \frac{8\theta_5}{n} - \frac{8\theta_6}{m} \tag{4.1}$$

b) Assume $H_0 : \boldsymbol{\mu_x} = \boldsymbol{\mu_y}$ is true with $m = n$ and $V(T_n) = V_0(T_n)$.

(i) If $\boldsymbol{\mu_x} = \boldsymbol{\mu_y} = \boldsymbol{\mu} \neq \mathbf{0}$, let $\gamma = \boldsymbol{\mu}^T(\boldsymbol{\Sigma_x} + \boldsymbol{\Sigma_y})\boldsymbol{\mu}$. Then

$$V_0(T_n) = \frac{2(\sigma_X^2 + \sigma_Y^2)}{n(n-1)} + \frac{4\sigma_Z^2}{n^2} + \frac{4(1-2n)\gamma}{n^2(n-1)}.$$

(ii) If $\boldsymbol{\mu_X} = \boldsymbol{\mu_Y} = \mathbf{0}$, then $\gamma = 0$ and

$$V_0(T_n) = \frac{2(\sigma_X^2 + \sigma_Y^2)}{n(n-1)} + \frac{4\sigma_Z^2}{n^2}.$$

c) Assume $H_0 : \boldsymbol{\mu_X} = \boldsymbol{\mu_Y}$ is true with $m \neq n$ and $V(T_n) = V_0(T_n)$.

(i) If $\boldsymbol{\mu_X} = \boldsymbol{\mu_Y} = \boldsymbol{\mu} \neq \mathbf{0}$, let $\tau_1 = \frac{1}{(n-1)}\boldsymbol{\mu}^T \boldsymbol{\Sigma_X} \boldsymbol{\mu}$ and $\tau_2 = \frac{1}{(m-1)}\boldsymbol{\mu}^T \boldsymbol{\Sigma_Y} \boldsymbol{\mu}$. Then

$$V_0(T_n) = \frac{2\sigma_X^2}{n(n-1)} + \frac{2\sigma_Y^2}{m(m-1)} + \frac{4\sigma_Z^2}{nm} - \frac{4(N-1)}{nm}(\tau_1 + \tau_2)$$

(ii) If $\boldsymbol{\mu_X} = \boldsymbol{\mu_Y} = \mathbf{0}$, then $\tau_1 = \tau_2 = 0$ and

$$V_0(T_n) = \frac{2\sigma_X^2}{n(n-1)} + \frac{2\sigma_Y^2}{m(m-1)} + \frac{4\sigma_Z^2}{nm}.$$

"Proof." a) Note that $T_n = T_1 + T_2 - 2T_3$, so:

$$V(T_n) = V(T_1 + T_2 - 2T_3) = Cov(T_1 + T_2 - 2T_3, T_1 + T_2 - 2T_3)$$

Thus

$$V(T_n) = V(T_1) + V(T_2) + 4V(T_3) - 4Cov(T_1, T_3) - 4Cov(T_2, T_3) \tag{4.2}$$

To find the variance $V(T_n)$, let $X_{ij} = \boldsymbol{x}_i^T \boldsymbol{x}_j = X_{ji}$, $Y_{ij} = \boldsymbol{y}_i^T \boldsymbol{y}_j = Y_{ji}$, $Z_{ij} = \boldsymbol{x}_i^T \boldsymbol{y}_j = Z_{ji}$ and note that

$$T_1 = \frac{2}{n(n-1)} H_1 \text{ where } \mathrm{H}_1 = \sum_{\mathrm{i} <}^{\mathrm{n}} \sum_{\mathrm{j}}^{\mathrm{n}} \boldsymbol{x}_{\mathrm{i}}^{\mathrm{T}} \boldsymbol{x}_{\mathrm{j}} = \sum_{\mathrm{i<j}}^{\mathrm{n}} \boldsymbol{x}_{\mathrm{i}}^{\mathrm{T}} \boldsymbol{x}_{\mathrm{j}}.$$

$$T_2 = \frac{2}{m(m-1)} H_2 \text{ where } \mathrm{H}_2 = \sum_{\mathrm{i} <}^{\mathrm{m}} \sum_{\mathrm{j}}^{\mathrm{m}} \boldsymbol{y}_{\mathrm{i}}^{\mathrm{T}} \boldsymbol{y}_{\mathrm{j}} = \sum_{\mathrm{i<j}}^{\mathrm{m}} \boldsymbol{y}_{\mathrm{i}}^{\mathrm{T}} \boldsymbol{y}_{\mathrm{j}}.$$

$$T_3 = \frac{1}{nm} H_3 \text{ where } \mathrm{H}_3 = \sum_{\mathrm{i=1}}^{\mathrm{n}} \sum_{\mathrm{j=1}}^{\mathrm{m}} \boldsymbol{x}_{\mathrm{i}}^{\mathrm{T}} \boldsymbol{y}_{\mathrm{j}}.$$

Then from one sample test, we have

$$V(T_1) = \frac{2\sigma_X^2}{n(n-1)} + \frac{4(n-2)\theta_1}{n(n-1)} \tag{4.3}$$

$$V(T_2) = \frac{2\sigma_Y^2}{m(m-1)} + \frac{4(m-2)\theta_2}{m(m-1)}. \tag{4.4}$$

Now consider

$$V(T_3) = Cov(T_3, T_3) = Cov(\frac{1}{nm}H_3, \frac{1}{nm}H_3) = \frac{1}{(nm)^2}Cov(H_3, H_3)$$

$$= \frac{1}{(nm)^2}Cov\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\boldsymbol{x}_i^T\boldsymbol{y}_j, \sum_{k=1}^{n}\sum_{t=1}^{m}\boldsymbol{x}_k^T\boldsymbol{y}_t\right) = \frac{1}{(nm)^2}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{t=1}^{m}Cov(Z_{ij}, Z_{kt}).$$

Let $V(Z_{ij}) = \sigma_Z^2$. The covariances are of 4 types. First, if $i = k$ and $j = t$, then $Cov(Z_{ij}, Z_{kt}) = Cov(Z_{ij}, Z_{ij}) = V(Z_{ij}) = \sigma_Z^2$, there are nm terms. Second, if $i = k$ and $j \neq t$, then $Cov(Z_{ij}, Z_{it}) = \theta_3$, there are $nm(m-1)$ terms. Third, if $i \neq k$ and $j = t$, then $Cov(Z_{ij}, Z_{kj}) = \theta_4$, there are $mn(n-1)$ terms. Fourth, if all four subscripts i, j, k, t are distinct i.e. $i \neq k$ and $j \neq t$, then Z_{ij} and Z_{kt} are independent with $Cov(Z_{ij}, Z_{kt}) = 0$ there are $nm(m-1)(n-1)$ terms. Thus

$$V(T_3) = \frac{1}{nm}[\sigma_Z^2 + (m-1)\theta_3 + (n-1)\theta_4] \tag{4.5}$$

and

$$Cov(T_1, T_3) = Cov\left(\frac{2}{n(n-1)}H_1, \frac{1}{nm}H_3\right) = \frac{2}{mn^2(n-1)}Cov\left(\sum_{i<}^{n}\sum_{j}^{n}\boldsymbol{x}_i^T\boldsymbol{x}_j, \sum_{a=1}^{n}\sum_{b=1}^{m}\boldsymbol{x}_a^T\boldsymbol{y}_b\right)$$

$$= \frac{2}{mn^2(n-1)}\sum_{i<}^{n}\sum_{j}^{n}\sum_{a=1}^{n}\sum_{b=1}^{m}Cov\left(\boldsymbol{x}_i^T\boldsymbol{x}_j, \boldsymbol{x}_a^T\boldsymbol{y}_b\right).$$

The covariances are of 2 types. First, if $a = i$ or $a = j$, then $Cov(X_{ij}, Z_{ab}) = \theta_5$, there are $mn(n-1)$ terms. Second, if $a \neq i$ and $a \neq j$, then i, j, a are distinct and $Cov(X_{ij}, Z_{ab}) = 0$.

Hence,

$$Cov(T_1, T_3) = \frac{2}{n}\theta_5. \tag{4.6}$$

Similarly,

$$Cov(T_2, T_3) = Cov\left(\frac{2}{m(m-1)}H_2, \frac{1}{nm}H_3\right) = \frac{2}{nm^2(m-1)}Cov\left(\sum_{i<}^{m}\sum_{j}^{m}\boldsymbol{y}_i^T\boldsymbol{y}_j, \sum_{a=1}^{n}\sum_{b=1}^{m}\boldsymbol{x}_a^T\boldsymbol{y}_b\right)$$

$$= \frac{2}{nm^2(m-1)}\sum_{i<}^{m}\sum_{j}^{m}\sum_{a=1}^{n}\sum_{b=1}^{m}Cov\left(\boldsymbol{y}_i^T\boldsymbol{y}_j, \boldsymbol{x}_a^T\boldsymbol{y}_b\right).$$

The covariances are of 2 types. First, if $b = i$ or $b = j$, then $Cov(Y_{ij}, Z_{ab}) = \theta_6$, there are $nm(m-1)$ terms. Second, if $b \neq i$ nor $b \neq j$, then i, j, b are distinct and $Cov(Y_{ij}, Z_{ab}) = 0$.

Hence,

$$Cov(T_2, T_3) = \frac{2}{m}\theta_6. \tag{4.7}$$

Therefore, from equations (4.2)-(4.7) we get:

$$V(T_n) = \frac{2\sigma_X^2}{n(n-1)} + \frac{2\sigma_Y^2}{m(m-1)} + \frac{4\sigma_Z^2}{nm} + \frac{4(n-2)\theta_1}{n(n-1)} + \frac{4(m-2)\theta_2}{m(m-1)} + \frac{4(m-1)\theta_3}{nm}$$

$$+\frac{4(n-1)\theta_4}{nm} - \frac{8\theta_5}{n} - \frac{8\theta_6}{m}.$$

□

b) From one sample test, we know that

$$\theta_1 = Cov(X_{ij}, X_{it}) = \boldsymbol{\mu}_{\boldsymbol{x}}^T\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\mu}_{\boldsymbol{x}}$$

and

$$\theta_2 = Cov(Y_{ij}, Y_{it}) = \boldsymbol{\mu_y}^T \boldsymbol{\Sigma_y} \boldsymbol{\mu_y}.$$

Now $\theta_3 = Cov(Z_{ij}, Z_{it}) = Cov(\boldsymbol{x}_i^T \boldsymbol{y}_j, \boldsymbol{x}_i^T \boldsymbol{y}_t)$ where $\boldsymbol{x}_i, \boldsymbol{y}_j$, and $\boldsymbol{y}_t$ are independent. Hence $\theta_3 =$

$$Cov(\sum_a x_{ia} y_{ja}, \sum_b x_{ib} y_{tb}) = \sum_a \sum_b Cov(x_{ia} y_{ja}, x_{ib} y_{tb}) =$$

$$\sum_a \sum_b [E(x_{ia} y_{ja} x_{ib} y_{tb}) - E(x_{ia} y_{ja}) E(x_{ib} y_{tb})] =$$

$$\sum_a \sum_b [E(x_{ia} x_{ib}) E(y_{ja}) E(y_{tb}) - E(x_{ia}) E(y_{ja}) E(x_{ib}) E(y_{tb})] =$$

$$\sum_a \sum_b [E(y_{ja}) E(y_{tb}) (E(x_{ia} x_{ib}) - E(x_{ia}) E(x_{ib}))] =$$

$$\sum_a \sum_b [E(y_{ja}) E(y_{tb}) \; Cov(x_{ia}, x_{ib})] =$$

$$\sum_a \sum_b \sigma_{ab} E(y_{ja}) E(y_{tb}) = \boldsymbol{\mu_y}^T \boldsymbol{\Sigma_x} \boldsymbol{\mu_y}.$$

Similarly, $\theta_4 = Cov(Z_{ij}, Z_{kj}) = Cov(\boldsymbol{x}_i^T \boldsymbol{y}_j, \boldsymbol{x}_k^T \boldsymbol{y}_j)$ where $\boldsymbol{x}_i, \boldsymbol{x}_k$, and $\boldsymbol{y}_j$ are independent. So $\theta_4 =$

$$Cov(\sum_a x_{ia} y_{ja}, \sum_b x_{kb} y_{jb}) = \sum_a \sum_b Cov(x_{ia} y_{ja}, x_{kb} y_{jb}) =$$

$$\sum_a \sum_b [E(x_{ia} y_{ja} x_{kb} y_{jb}) - E(x_{ia} y_{ja}) E(x_{xb} y_{jb})] =$$

$$\sum_a \sum_b [E(x_{ia}) E(x_{kb}) E(y_{ja} y_{jb}) - E(x_{ia}) E(y_{ja}) E(x_{kb}) E(y_{jb})] =$$

$$\sum_a \sum_b [E(x_{ia}) E(x_{kb}) (E(y_{ja} y_{jb}) - E(y_{ja}) E(y_{jb}))] =$$

$$\sum_a \sum_b [E(x_{ia}) E(x_{kb}) \; Cov(y_{ja}, y_{jb})] =$$

$$\sum_a \sum_b \sigma_{ab} E(x_{ia}) E(x_{kb}) = \boldsymbol{\mu_x}^T \boldsymbol{\Sigma_y} \boldsymbol{\mu_x}$$

and $\theta_5 = Cov(X_{ij}, Z_{ab})$ with either $a = i$ or $a = j$. So, $\theta_5 = Cov(\boldsymbol{x}_i^T \boldsymbol{x}_j, \boldsymbol{x}_i^T \boldsymbol{y}_b)$ where $\boldsymbol{x}_i, \boldsymbol{x}_j$, and $\boldsymbol{y}_b$ are independent. Hence

$$\theta_5 = Cov(\sum_d x_{id}x_{jd}, \sum_t x_{it}y_{bt}) = \sum_d \sum_t Cov(x_{id}x_{jd}, x_{it}y_{bt}) =$$

$$\sum_d \sum_t [E(x_{id}x_{jd}x_{it}y_{bt}) - E(x_{id}x_{jd})E(x_{it}y_{bt})] =$$

$$\sum_d \sum_t [E(x_{id}x_{it})E(x_{jd})E(y_{bt}) - E(x_{id})E(x_{jd})E(x_{it})E(y_{bt})] =$$

$$\sum_d \sum_t [E(x_{jd})E(y_{bt})(E(x_{id}x_{it}) - E(x_{id})E(x_{it}))] =$$

$$\sum_d \sum_t [E(x_{jd})E(y_{bt})\ Cov(x_{id}, x_{it})] =$$

$$\sum_d \sum_t \sigma_{dt}E(x_{jd})E(y_{bt}) = \boldsymbol{\mu}_{\boldsymbol{x}}^T\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\mu}_{\boldsymbol{y}}.$$

Likewise, $\theta_6 = Cov(Y_{ij}, Z_{ab})$ with either $b = i$ or $b = j$. So, $\theta_6 = Cov(\boldsymbol{y}_i^T\boldsymbol{y}_j, \boldsymbol{x}_a^T\boldsymbol{y}_i)$ where $\boldsymbol{y}_i, \boldsymbol{y}_j$, and $\boldsymbol{x}_a$ are independent. Hence

$$\theta_6 = Cov(\sum_d y_{id}y_{jd}, \sum_t x_{at}y_{it}) = \sum_d \sum_t Cov(y_{id}y_{jd}, x_{at}y_{it}) =$$

$$\sum_d \sum_t [E(y_{id}y_{jd}x_{at}y_{it}) - E(y_{id}y_{jd})E(x_{at}y_{it})] =$$

$$\sum_d \sum_t [E(y_{id}y_{it})E(y_{jd})E(x_{at}) - E(y_{id})E(y_{jd})E(y_{it})E(x_{at})] =$$

$$\sum_d \sum_t [E(y_{jd})E(x_{at})(E(y_{id}y_{it}) - E(y_{id})E(y_{it}))] =$$

$$\sum_d \sum_t [E(y_{jd})E(x_{at})\ Cov(y_{id}, y_{it})] =$$

$$\sum_d \sum_t \sigma_{dt}E(y_{jd})E(x_{at}) = \boldsymbol{\mu}_{\boldsymbol{y}}^T\boldsymbol{\Sigma}_{\boldsymbol{y}}\boldsymbol{\mu}_{\boldsymbol{x}}.$$

Under H_0, $\boldsymbol{\mu}_{\boldsymbol{x}} = \boldsymbol{\mu}_{\boldsymbol{y}}$, and assume $m = n$. Let $\boldsymbol{\mu}_{\boldsymbol{x}} = \boldsymbol{\mu}_{\boldsymbol{y}} = \boldsymbol{\mu} \neq \boldsymbol{0}$, then $\theta_1 = \theta_3 = \theta_5 = \boldsymbol{\mu}^T\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\mu}$ and $\theta_2 = \theta_4 = \theta_6 = \boldsymbol{\mu}^T\boldsymbol{\Sigma}_{\boldsymbol{y}}\boldsymbol{\mu}$. Substituting these values into equation (4.1) gives

$$V_0(T_n) = \frac{2\sigma_X^2}{n(n-1)} + \frac{2\sigma_Y^2}{n(n-1)} + \frac{4\sigma_Z^2}{n^2} + \left[\frac{4(n-2)}{n(n-1)} + \frac{4(n-1)}{n^2} - \frac{8}{n}\right]\boldsymbol{\mu}^T\boldsymbol{\Sigma_x\mu}$$
$$+ \left[\frac{4(n-2)}{n(n-1)} + \frac{4(n-1)}{n^2} - \frac{8}{n}\right]\boldsymbol{\mu}^T\boldsymbol{\Sigma_y\mu}$$

$$= \frac{2(\sigma_X^2 + \sigma_Y^2)}{n(n-1)} + \frac{4\sigma_Z^2}{n^2} + \left[\frac{4(n-2)}{n(n-1)} + \frac{4(n-1)}{n^2} - \frac{8}{n}\right](\boldsymbol{\mu}^T\boldsymbol{\Sigma_x\mu} + \boldsymbol{\mu}^T\boldsymbol{\Sigma_y\mu})$$

$$= \frac{2(\sigma_X^2 + \sigma_Y^2)}{n(n-1)} + \frac{4\sigma_Z^2}{n^2} + \frac{4(1-2n)}{n^2(n-1)}\boldsymbol{\mu}^T(\boldsymbol{\Sigma_x} + \boldsymbol{\Sigma_y})\boldsymbol{\mu}.$$

Let $\gamma = \boldsymbol{\mu}^T(\boldsymbol{\Sigma_x} + \boldsymbol{\Sigma_y})\boldsymbol{\mu}$. Therefore

$$V_0(T_n) = \frac{2(\sigma_X^2 + \sigma_Y^2)}{n(n-1)} + \frac{4\sigma_Z^2}{n^2} + \frac{4(1-2n)}{n^2(n-1)}\gamma$$

which proves (i).

For (ii), let $\boldsymbol{\mu} = \mathbf{0}$ and thus $\gamma = \boldsymbol{\mu}^T(\boldsymbol{\Sigma_x} + \boldsymbol{\Sigma_y})\boldsymbol{\mu} = 0$ then the result follows. □

c) For (i), let $\boldsymbol{\mu_x} = \boldsymbol{\mu_y} = \boldsymbol{\mu} \neq \mathbf{0}$ and assume $m \neq n$. Use $\theta_1 = \theta_3 = \theta_5 = \boldsymbol{\mu}^T\boldsymbol{\Sigma_x\mu}$ and $\theta_2 = \theta_4 = \theta_6 = \boldsymbol{\mu}^T\boldsymbol{\Sigma_y\mu}$. Substituting these values into equation (4.1) gives

$$V_0(T_n) = \frac{2\sigma_X^2}{n(n-1)} + \frac{2\sigma_Y^2}{m(m-1)} + \frac{4\sigma_Z^2}{nm} + \left[\frac{4(n-2)}{n(n-1)} + \frac{4(m-1)}{nm} - \frac{8}{n}\right]\boldsymbol{\mu}^T\boldsymbol{\Sigma_x\mu}$$
$$+ \left[\frac{4(m-2)}{m(m-1)} + \frac{4(n-1)}{nm} - \frac{8}{m}\right]\boldsymbol{\mu}^T\boldsymbol{\Sigma_y\mu}$$

$$= \frac{2\sigma_X^2}{n(n-1)} + \frac{2\sigma_Y^2}{m(m-1)} + \frac{4\sigma_Z^2}{nm} - \frac{4(n+m-1)}{nm(n-1)}\boldsymbol{\mu}^T\boldsymbol{\Sigma_x\mu} - \frac{4(m+n-1)}{nm(m-1)}\boldsymbol{\mu}^T\boldsymbol{\Sigma_y\mu}$$

$$= \frac{2\sigma_X^2}{n(n-1)} + \frac{2\sigma_Y^2}{m(m-1)} + \frac{4\sigma_Z^2}{nm} - \frac{4(n+m-1)}{nm}[\frac{1}{(n-1)}\boldsymbol{\mu}^T \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\mu} + \frac{1}{(m-1)}\boldsymbol{\mu}^T \boldsymbol{\Sigma}_{\boldsymbol{y}} \boldsymbol{\mu}].$$

Let $\frac{1}{(n-1)}\boldsymbol{\mu}^T \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\mu} = \tau_1$ and $\frac{1}{(m-1)}\boldsymbol{\mu}^T \boldsymbol{\Sigma}_{\boldsymbol{y}} \boldsymbol{\mu} = \tau_2$ and replace $n + m$ by N. Thus

$$V_0(T_n) = \frac{2\sigma_X^2}{n(n-1)} + \frac{2\sigma_Y^2}{m(m-1)} + \frac{4\sigma_Z^2}{nm} - \frac{4(N-1)}{nm}(\tau_1 + \tau_2).$$

For (ii), let $\boldsymbol{\mu}_{\boldsymbol{x}} = \boldsymbol{\mu}_{\boldsymbol{y}} = \boldsymbol{\mu} = \mathbf{0}$ and thus $\tau_1 = \tau_2 = 0$ then the result follows. $\square$

CHAPTER 5

SIMULATIONS

5.0.1 One Sample Tests

In the simulations, we examined four one sample tests. The first "test" used the m out of n bootstrap to compute $T_1^*, ..., T_B^*$ with $B = 100$. We used the shorth bootstrap confidence interval described in Olive (2025, chapter 2) and Pelawa Watagoda and Olive (2021). This "test" has not been proven to have level α. The second test computed the usual t confidence interval

$$[\overline{W} - t_{1-\alpha/2,m-1} S_W/\sqrt{m}, \overline{W} + t_{1-\alpha/2,m-1} S_W/\sqrt{m}]$$

for $\boldsymbol{\mu}^T \boldsymbol{\mu}$ based on the W_i from Theorem 1. The third and fourth tests used Theorem 2 b) and Equation 1.2): $T_n/s_n \xrightarrow{D} N(0,1)$ if s_n^2 is a consistent estimator of $V(T_n)$ when H_0 is true. The third test used

$$s_n^2 = \hat{\sigma}_W^2 = \frac{1}{n(n-1)} \sum \sum_{i \neq j} (\boldsymbol{x}_i^T \boldsymbol{x}_j - T_n)^2 = \frac{1}{n(n-1)} \sum \sum_{i \neq j} (W_{ij} - T_n)^2.$$

If the denominator $n(n-1)$ was replaced by $n(n-1)-1$, this statistic would be the usual sample variance of the W_{ij}, which are not independent. This test is nearly the same as the Li (2023) test. The fourth test used $s_n^2 = S_W^2$ based on Theorem 1. These two tests computed intervals

$$[T_n - t_{1-\alpha/2,m-1} \sqrt{2s_n^2/[n(n-1)]}, T_n + t_{1-\alpha/2,m-1} \sqrt{2s_n^2/[n(n-1)]}].$$

The third test computed the usual t confidence interval

$$[\overline{W} - t_{1-\alpha/2,m-1} S_W/\sqrt{m}, \overline{W} + t_{1-\alpha/2,m-1} S_W/\sqrt{m}]$$

for $\boldsymbol{\mu}^T\boldsymbol{\mu}$ based on the W_i from Theorem 1. The tests 2–4 use the same cutoff $t_{1-\alpha/2,m-1}$ so that the average interval lengths are more comparable. The fifth test used the Theorem 2 test applied to the spatial sign vectors with S^2_W.

The estimator $\hat{\sigma}^2_W$ is easy to code in R. Let $\boldsymbol{X}$ be the $n \times p$ data matrix with ith row $\boldsymbol{x}_i^T$. Then the sum of squares and cross products matrix is $\boldsymbol{C} = \boldsymbol{X}\boldsymbol{X}^T = (c_{ij})$ with ijth element $c_{ij} = \boldsymbol{x}_i^T\boldsymbol{x}_j$. Let $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{X}^T - T_n\boldsymbol{1}\boldsymbol{1}^T = (a_{ij})$ where $\boldsymbol{1}\boldsymbol{1}^T$ is the $n \times n$ matrix of ones. Let matrix $\boldsymbol{V} = (v_{ij})$ where $v_{ij} = a^2_{ij} = (\boldsymbol{x}_i^T\boldsymbol{x}_j - T_n)^2$ is the ijth element of $\boldsymbol{V}$. Thus $n(n-1)\hat{\sigma}^2_W = \sum_{i=1}^n\sum_{j=1}^n v_{ij} - \sum_{i=1}^n v_{ii}$.

```
k <- n*(n-1)
a <- apply(x,2,sum)   #a = n xbar and x is the data matrix
Thd <- (t(a)%*%a - sum(x^2))/k
Thd <- as.double(Thd)   #Thd = Tn
sscp <- x%*%t(x)
ss <- sscp - Thd
ss <- ss^2
vw1 <- (sum(ss) - sum(diag(ss)))/k    #\hat{\sigma}_W^2
```

The simulation used four distribution types where $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{y} + \delta\boldsymbol{1}$ with $E(\boldsymbol{x}) = \delta\boldsymbol{1}$ where $\boldsymbol{1}$ is the $p \times 1$ vector of ones. Type 1 used $\boldsymbol{y} \sim N_p(\boldsymbol{0}, \boldsymbol{I})$, type 2 used a mixture distribution $\boldsymbol{y} \sim 0.6N_p(\boldsymbol{0}, \boldsymbol{I}) + 0.4N_p(\boldsymbol{0}, 25\boldsymbol{I})$, type 3 for a multivariate t_4 distribution, and type 4 for a multivariate lognormal distribution where $\boldsymbol{y} = (y_1, ..., y_p)$ with $w_i = \exp(Z)$ where $Z \sim N(0, 1)$ and $y_i = w_i - E(w_i)$ where $E(w_i) = \exp(0.5)$. The covariance matrix type depended on the matrix $\boldsymbol{A}$. Type 1 used $\boldsymbol{A} = \boldsymbol{I}_p$, type 2 used $\boldsymbol{A} = diag(\sqrt{1}, ..., \sqrt{p})$, and type 3 used $\boldsymbol{A} = \psi\boldsymbol{1}\boldsymbol{1}^T + (1-\psi)\boldsymbol{I}_p$ giving $\text{cor}(x_{ij}, x_{ik}) = \rho$ for $j \neq k$ where $\rho = 0$ if $\psi = 0$, $\rho \to 1/(c+1)$ as $p \to \infty$ if $\psi = 1/\sqrt{cp}$ where $c > 0$, and $\rho \to 1$ as $p \to \infty$ if $\psi \in (0, 1)$ is a constant. We used $\delta = 0$ and $\delta > 0$ chosen so at least one test had good power. The simulation used 5000 runs, the 4 $\boldsymbol{x}$ distributions, and the 3 matrices $\boldsymbol{A}$. For the third $\boldsymbol{A}$, we used $\psi = 1/\sqrt{p}$.

Tables 5.1-5.9 summarize some simulation results. There are two lines for each simulation scenario. The first line gives the simulated power = proportion of times $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ was rejected.

The second line gives the average length of the confidence interval where H_0 is rejected if 0 is not in the confidence interval. When $\delta = 0$, observed coverage between 0.04 and 0.06 suggests coverage = power = level is close to the nominal value 0.05. For larger δ, want the coverage near 1 for good power.

The bootstrap test corresponds to the boot column, the tests using $(\overline{w}, S_W)$, $(T_n, \hat{\sigma}_W)$, and (T_n, S_W) correspond to the next three columns. The last column corresponds to the spatial sign test. This test tends to have much shorter lengths because of the transformation of the data. The test using $(\overline{w}, S_W)$ has simple large sample theory, but large confidence interval length and low power compared to the other methods. The bootstrap test was sometimes conservative with observed coverage < 0.04 when delta=0. For xtype=4 and delta=0, H_0 was not true for the spatial test. Hence the coverage for the spatial test was sometimes higher than 0.06 for this scenario. For delta=0, the test with $(T_n, \hat{\sigma}_W)$ sometimes had coverage less than 0.04, while the test with (T_n, S_W) sometimes had coverage greater than 0.06. In the simulations, the spatial test often performed well, but typically $E(\boldsymbol{z}_i) = \boldsymbol{\mu}_{\boldsymbol{z}} \neq \boldsymbol{\mu}_{\boldsymbol{x}} = E(\boldsymbol{x}_i)$, which makes the spatial test harder to use. For testing $H_0 : \boldsymbol{\mu}_{\boldsymbol{x}} = \boldsymbol{0}$, the test with $(T_n, \hat{\sigma}_W)$ appeared to perform better than the three competitors.

5.0.2 Two Sample Tests

In the simulations, we examined three sample tests. The first "test" used the m out of n bootstrap where $m_i = 2n_i/3$ to bootstrap the Chen and Qin (2010) test that estimates $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2 = \boldsymbol{\mu}_1^T\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2^T\boldsymbol{\mu}_2 - 2\boldsymbol{\mu}_1^T\boldsymbol{\mu}_2$. The second test was the "paired test" with $m = \min(n_1, n_2)$ and $\boldsymbol{z}_i = \boldsymbol{x}_{1i} - \boldsymbol{x}_{2i}$ for $i = 1, ..., m$. Then apply the one sample test from Theorem 2 to the $\boldsymbol{z}_i$. The third test was the Li (2023) test. Both of these tests used S_W^2 applied to the $\boldsymbol{z}_i$ or the $\boldsymbol{y}_i$.

The simulation used four distribution types where $\boldsymbol{x}_1 = \boldsymbol{A}_1\boldsymbol{y}_1 + \delta\boldsymbol{1}$ and $\boldsymbol{x}_2 = \boldsymbol{A}_2\boldsymbol{y}_2$ where $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ had the same distribution, with $E(\boldsymbol{x}_1) = \delta\boldsymbol{1}$ and $E(\boldsymbol{x}_2) = \boldsymbol{0}$. Type 1 used $\boldsymbol{y} \sim N_p(\boldsymbol{0}, \boldsymbol{I})$, type 2 used a mixture distribution $\boldsymbol{y} \sim 0.6N_p(\boldsymbol{0}, \boldsymbol{I}) + 0.4N_p(\boldsymbol{0}, 25\boldsymbol{I})$, type 3 for a multivariate t_4 distribution, and type 4 for a multivariate lognormal distribution where $\boldsymbol{y} = (y_1, ..., y_p)$ with $w_i = \exp(Z)$ where $Z \sim N(0, 1)$ and $y_i = w_i - E(w_i)$ where $E(w_i) = \exp(0.5)$. The covariance matrix type depended on

Table 5.1. One sample tests, covtyp=1, p=100

n	p	psi/xtype	delta	boot	$(\overline{w}, S_W)$	$(T_n, \hat{\sigma}_W)$	(T_n, S_W)	spatial
100	100	0	0	0.0230	0.0580	0.0400	0.0452	0.0444
	len	1		0.6732	5.6520	0.5711	0.5681	0.0057
100	100	0	0.075	0.8160	0.0688	0.9216	0.9176	0.9166
	len	1		0.8081	5.7018	0.5741	0.5731	0.0057
100	100	0	0	0.0236	0.0436	0.0466	0.0776	0.0478
	len	2		7.0590	58.2593	6.0094	5.8553	0.0057
100	100	0	0.15	0.1938	0.0506	0.3128	0.349	0.9988
	len	2		7.5830	58.1417	6.0204	5.8435	0.0057
100	100	0	0	0.0222	0.0466	0.045	0.068	0.0468
	len	3		1.3031	10.6946	1.1140	1.0749	0.0057
100	100	0	0.1	0.7536	0.0544	0.872	0.8714	0.9956
	len	3		1.5563	10.8976	1.1260	1.0953	0.0057
100	100	0	0	0.0206	0.0556	0.0372	0.0656	0.0906
	len	4		3.1105	25.4558	2.6543	2.5584	0.0057
100	100	0	0.17	0.9024	0.0546	0.9622	0.9496	0.7668
	len	4		3.7816	25.5420	2.6708	2.5671	0.0057

Table 5.2. One sample tests, covtyp=1, p=1000

n	p	psi/xtype	delta	boot	$(\overline{w}, S_W)$	$(T_n, \hat{\sigma}_W)$	(T_n, S_W)	spatial
100	1000	0	0	0.0236	0.0482	0.0448	0.0506	0.0506
	len	1		2.1403	17.8302	1.8059	1.7920	0.0018
100	1000	0	0.0415	0.872	0.068	0.9438	0.9398	0.9388
	len	1		2.2771	17.9004	1.8089	1.7991	0.0018
100	1000	0	0	0.0236	0.0448	0.0458	0.0712	0.0558
	len	2		22.4434	185.1105	19.0973	18.6043	0.0018
100	1000	0	0.075	0.142	0.048	0.2222	0.2616	0.9978
	len	2		22.8203	182.6556	18.9772	18.3576	0.0018
100	1000	0	0	0.0214	0.0432	0.0436	0.065	0.045
	len	3		4.1649	34.1708	3.5444	3.4343	0.0018
100	1000	0	0.05	0.6458	0.0558	0.7642	0.777	0.9908
	len	3		4.3708	34.0483	3.5586	3.4220	0.0018
100	1000	0	0	0.0192	0.0544	0.0378	0.0518	0.0484
	len	4		9.9417	82.3953	8.4267	8.2810	0.0018
100	1000	0	0.087	0.843	0.0576	0.9282	0.9242	0.8774
	len	4		10.5664	82.8816	8.4523	8.3299	0.0018

Table 5.3. One sample tests, covtyp=1, p=10000

n	p	psi/xtype	delta	boot	$(\overline{w}, S_W)$	$(T_n, \hat{\sigma}_W)$	(T_n, S_W)	spatial
100	10000	0	0	0.024	0.0474	0.0446	0.0474	0.0462
	len	1		6.7618	56.7383	5.7116	5.7024	0.0006
100	10000	0	0.023	0.8778	0.0652	0.9476	0.946	0.9466
	len	1		6.8718	56.7593	5.7149	5.7045	0.0006
100	10000	0	0	0.021	0.04	0.0386	0.0764	0.0438
	len	2		70.5972	581.8741	60.1418	58.4806	0.0006
100	10000	0	0.05	0.2624	0.0524	0.3738	0.4032	1
	len	2		71.5393	582.2665	60.1618	58.5200	0.0006
100	10000	0	0	0.0224	0.0436	0.0472	0.0778	0.0554
	len	3		13.2420	108.8067	11.2650	10.9355	0.0006
100	10000	0	0.03	0.7824	0.0588	0.8706	0.87	0.9992
	len	3		13.3547	108.5636	11.1969	10.9111	0.0006
100	10000	0	0	0.0272	0.0504	0.0446	0.0516	0.0502
	len	4		31.6188	263.0578	26.67685	26.4383	0.0006
100	10000	0	0.05	0.8958	0.054	0.9606	0.9618	0.953
	len	4		32.3627	263.551	26.6933	26.4879	0.0006

Table 5.4. One sample tests, covtyp=2, p=100

n	p	psi/xtype	delta	boot	$(\overline{w}, S_W)$	$(T_n, \hat{\sigma}_W)$	(T_n, S_W)	spatial
100	100	0	0	0.0212	0.0498	0.038	0.043	0.0414
	len	1		38.9543	329.1668	33.2225	33.0825	0.0065
100	100	0	0.6	0.8966	0.0758	0.956	0.9548	0.9556
	len	1		46.3236	330.7589	33.3672	33.2425	0.0065
100	100	0	0	0.0214	0.0502	0.0398	0.0726	0.0506
	len	2		410.1416	3394.75	350.1749	341.1852	0.0065
100	100	0	1.5	0.5062	0.0526	0.6492	0.662	1
	len	2		455.0242	3396.337	350.6696	341.3447	0.0066
100	100	0	0	0.023	0.041	0.0454	0.0684	0.0474
	len	3		76.2693	629.0579	65.2686	63.2227	0.0065
100	100	0	0.75	0.755	0.06	0.8558	0.8608	0.997
	len	3		88.0646	634.0106	65.49	63.7205	0.0065
100	100	0	0	0.0222	0.0608	0.042	0.0738	0.1156
	len	4		178.6321	1470.551	153.3266	147.7959	0.0064
100	100	0	1.2	0.8532	.0492	0.932	0.9214	0.741
	len	4		207.835	1459.873	154.4866	146.7227	0.0063

Table 5.5. One sample tests, covtyp=2, p=1000

n	p	psi/xtype	delta	boot	$(\overline{w}, S_W)$	$(T_n, \hat{\sigma}_W)$	(T_n, S_W)	spatial
100	1000	0	0	0.0286	0.0476	0.0438	0.0482	0.049
	len	1		1231.498	10344.15	1043.615	1039.626	0.0021
100	1000	0	0.975	0.8472	0.0648	0.9282	0.9204	0.9208
	len	1		1300.17	10379.01	1045.303	1043.129	0.0021
100	1000	0	0	0.0266	0.0386	0.047	0.0784	0.0536
	len	2		12929.72	106330.2	11004.27	10686.59	0.0021
100	1000	0	1.5	0.078	0.0388	0.1286	0.162	0.9474
	len	2		13095.03	106960.8	11016.42	10749.97	0.0021
100	1000	0	0	0.0222	0.0456	0.0446	0.0738	0.0454
	len	3		2387.572	19676.47	2033.522	1977.559	0.0021
100	1000	0	1.25	0.7222	0.0616	0.8276	0.8346	0.9986
	len	3		2514.451	19835.06	2051.272	1993.498	0.0021
100	1000	0	0	0.0268	0.0522	0.0462	0.063	0.0546
	len	4		5747.818	47479.65	4864.88	4771.884	0.0020
100	1000	0	2.15	0.8958	0.054	0.9544	0.9466	0.9198
	len	4		6064.615	47527.19	4876.035	4776.662	0.0021

Table 5.6. One sample tests, covtyp=2, p=10000

n	p	psi/xtype	delta	boot	$(\overline{w}, S_W)$	$(T_n, \hat{\sigma}_W)$	(T_n, S_W)	spatial
100	10000	0	0	0.0272	0.0536	0.045	0.0502	0.0496
	len	1		39006.52	326271.5	32976.41	32791.52	0.0007
100	10000	0	1.69	0.8482	0.0582	0.93	0.9294	0.9286
	len	1		39690.34	327648.8	32994.63	32929.94	0.0007
100	10000	0	0	0.0244	0.0442	0.0486	0.0876	0.0526
	len	2		408860	3330506	347476	334728.5	0.0007
100	10000	0	3	0.1126	0.0488	0.1778	0.2148	0.9952
	len	2		411196.1	3349674	347862.6	336654.9	0.0007
100	10000	0	0	0.0206	0.044	0.0436	0.0632	0.051
	len	3		75976.41	624134.1	64858.9	62727.84	0.0007
100	10000	0	2.5	0.8918	0.0608	0.9462	0.9454	1
	len	3		77389.1	625801.9	64740.62	62895.46	0.0007
100	10000	0	0	0.0236	0.0534	0.038	0.0444	0.0454
	len	4		181871.7	1517807	154052.2	152545.3	0.0007
100	10000	0	3.80	0.8952	0.0578	0.9558	0.9522	0.948
	len	4		185192.4	1518189	154094.1	152583.7	0.0007

Table 5.7. One sample tests, covtyp=3, p=100

n	p	psi/xtype	delta	boot	$(\overline{w}, S_W)$	$(T_n, \hat{\sigma}_W)$	(T_n, S_W)	spatial
100	100	0	0	0.0236	0.0526	0.0406	0.048	0.0448
	len	1		0.6739	5.6555	0.5711	0.5684	0.0057
100	100	0.1	0	0.0058	0.0456	0.0476	0.0492	0.0466
	len	1		6.6184	66.6393	6.7521	6.6975	0.0235
100	100	0	0.075	0.8146	0.0684	0.9172	0.9154	0.9132
	len	1		0.8074	5.6934	0.5744	0.5722	0.0057
100	100	0.1	0.40	0.796	0.1252	0.9572	0.9582	0.9344
	len	1		25.9220	74.8734	7.5906	7.5251	0.0245
100	100	0	0	0.0208	0.0396	0.0432	0.072	0.0496
	len	2		7.0581	58.6427	6.0136	5.8938	0.0057
100	100	0.1	0	0.0068	0.0322	0.0508	0.0688	0.0534
	len	2		70.7733	659.5251	70.1401	66.2847	0.0235
100	100	0	0.15	0.2032	0.051	0.3288	0.3624	0.999
	len	2		7.6090	58.3959	6.0258	5.8690	0.0057
100	100	0.1	0.7	0.228	0.0504	0.5074	0.5228	0.9742
	len	2		146.1211	694.7141	73.0017	69.8214	0.0254
100	100	0	0	0.0232	0.0458	0.0448	0.0712	0.0492
	len	3		1.3148	10.8342	1.1249	1.0889	0.0057
100	100	0.1	0	0.0078	0.0396	0.0576	0.067	0.054
	len	3		13.1361	124.6081	12.9959	12.5236	0.0234
100	100	0	0.1	0.7454	0.0654	0.8616	0.865	0.9962
	len	3		1.5587	10.9085	1.1336	1.0964	0.0057
100	100	0.1	0.5	0.6912	0.109	0.9016	0.9074	0.9808
	len	3		44.5694	137.9813	14.3748	13.8677	0.0250
100	100	0	0	0.0228	0.061	0.0458	0.0778	0.093
	len	4		3.1119	25.5519	2.6567	2.5681	0.0057
100	100	0.1	0	0.0082	0.0468	0.0494	0.049	0.2442
	len	4		30.7926	307.8728	31.3771	30.9424	0.0246
100	100	0	0.17	0.9042	0.048	0.9654	0.957	0.7574
	len	4		3.7701	25.3263	2.6645	2.5454	0.0056
100	100	0.1	0.85	0.7948	0.0976	0.9584	0.9584	0.642
	len	4		115.7032	343.3036	34.9664	34.5033	0.0247

Table 5.8. One sample tests, covtyp=3, p=1000

n	p	psi/xtype	delta	boot	$(\overline{w}, S_W)$	$(T_n, \hat{\sigma}_W)$	(T_n, S_W)	spatial
100	1000	0	0	0.0282	0.049	0.0516	0.056	0.0558
	len	1		2.1401	17.8831	1.8065	1.7973	0.0018
100	1000	0.0316	0	0.0066	0.0426	0.0512	0.0532	0.05
	len	1		58.4898	591.9678	60.0672	59.495	0.0207
100	1000	0	0.04	0.8196	0.061	0.9152	0.9124	0.9128
	len	1		2.2646	17.9067	1.8088	1.7997	0.0018
100	1000	0.0316	0.4	0.8342	0.1524	0.9732	0.974	0.9572
	len	1		241.2136	672.2661	68.1736	67.5653	0.0218
100	1000	0	0	0.0272	0.0438	0.0484	0.082	0.0522
	len	2		22.3855	182.4873	19.0115	18.3407	0.0018
100	1000	0.0316	0	0.0072	0.0306	0.0524	0.0636	0.0502
	len	2		617.6974	5850.04	625.8249	587.9512	0.0208
100	1000	0	0.1	0.3982	0.046	0.533	0.5552	1
	len	2		23.2021	184.5163	19.0359	18.5446	0.0018
100	1000	0.0316	0.7	0.2628	0.0522	0.557	0.5734	0.99
	len	2		1373.899	6128.062	652.5555	615.8934	0.0228
100	1000	0	0	0.0276	0.0464	0.0484	0.0742	0.0536
	len	3		4.1547	34.0314	3.5458	3.4203	0.0018
100	1000	0.0316	0	0.0082	0.043	0.0504	0.061	0.0482
	len	3		114.0482	1097.655	115.6618	110.3185	0.0207
100	1000	0	0.05	0.6502	0.0638	0.7662	0.7722	0.9924
	len	3		4.3608	34.1752	3.5518	3.4347	0.0018
100	1000	0.0316	0.5	0.7432	0.1282	0.9284	0.9294	0.988
	len	3		419.3617	1241.326	129.0976	124.7579	0.0224
100	1000	0	0	0.0252	0.0486	0.0432	0.0568	0.0548
	len	4		9.9698	82.6130	8.4362	8.3029	0.0018
100	1000	0.0316	0	0.0068	0.0448	0.05	0.0512	0.08
	len	4		272.2052	2776.522	281.1257	279.051	0.0210
100	1000	0	0.09	0.8848	0.0614	0.9534	0.9484	0.9128
	len	4		10.5916	82.9419	8.4411	8.3360	0.0018
100	1000	0.0316	0.75	0.7026	0.0962	0.9192	0.9214	0.79
	len	4		978.2186	3071.672	310.7018	308.7147	0.0216

Table 5.9. One sample tests, covtyp=3, p=10000

n	p	psi/xtype	delta	boot	$(\overline{w}, S_W)$	$(T_n, \hat{\sigma}_W)$	(T_n, S_W)	spatial
100	10000	0	0	0.0264	0.0466	0.0472	0.05	0.0504
	len	1		6.7701	56.6313	5.7115	5.6917	0.0006
100	10000	0	0.0233	0.8962	0.0682	0.9552	0.954	0.9542
	len	1		6.8971	56.6364	5.7153	5.6922	0.0006
100	10000	0.01	0	0.0056	0.046	0.0512	0.0528	0.0564
	len	1		564.4415	5680.653	576.1765	570.9271	0.020
100	10000	0.01	0.35	0.7202	0.1126	0.9248	0.9242	0.9026
	len	1		2053.81	6331.873	639.567	636.3772	0.0209
100	10000	0	0	0.024	0.0424	0.0428	0.0756	0.0434
	len	2		70.5595	586.6235	60.1390	58.9579	0.0006
100	10000	0	0.07	0.7832	0.063	0.8812	0.8688	1
	len	2		71.8118	579.2654	60.1122	58.2184	0.0006
100	10000	0.01	0	0.0098	0.0278	0.0536	0.0652	0.0542
	len	2		6023.129	56229.19	5976.323	5651.246	0.020
100	10000	0.01	0.05	0.008	0.0288	0.0538	0.0648	0.0624
	len	2		5909.249	56185.34	5990.066	5646.839	0.0200
100	10000	0	0	0.0228	0.044	0.0468	0.0718	0.0474
	len	3		13.1846	108.8429	11.2381	10.9391	0.0006
100	10000	0	0.031	0.821	0.061	0.9064	0.9024	0.9998
	len	3		13.4349	108.6346	11.2509	10.9182	0.0006
100	10000	0.01	0	0.0094	0.0412	0.051	0.061	0.0514
	len	3		1107.379	10715.91	1117.261	1076.99	0.020
100	10000	0.01	0.5	0.7578	0.1378	0.937	0.9374	0.9936
	len	3		4116.296	12003.91	1244.21	1206.438	0.0217
100	10000	0	0	0.0224	0.0528	0.0406	0.0498	0.048
	len	4		31.5140	262.9498	26.6745	26.4274	0.0006
100	10000	0	0.049	0.8658	0.0562	0.937	0.9342	0.9266
	len	4		32.2037	264.4047	26.6915	26.5737	0.0006
100	10000	0.01	0	0.0054	0.047	0.0476	0.0484	0.0516
	len	4		2596.682	26497.84	2691.244	2663.133	0.020
100	10000	0.01	0.75	0.7112	0.1052	0.9262	0.9268	0.873
	len	4		9507.08	29481.95	2986.855	2963.047	0.0208

Table 5.10. Two sample tests, covtyp=1, p=100

(n_1, n_2, σ, p)	xtype	covtype	delta	boot	pair	Li
(100,100,1,100)	1	1	0	0.0246	0.0494	0.0494
len	1	1	0	1.3426	1.1389	1.1389
(100,100,1,100)	1	1	0.1	0.7224	0.8586	0.8586
len	1	1	0.1	1.5789	1.1417	1.1417
(100,200,1,100)	1	1	0	0.0256	0.0456	0.0462
len	1	1	0	1.0019	1.1360	0.8535
(100,200,1,100)	1	1	0.1	0.9166	0.8602	0.9612
len	1	1	0.1	1.2396	1.1432	0.8609

the matrix $\boldsymbol{A}$.

For the covariance types, $Cov(\boldsymbol{x}_1) = \boldsymbol{I}, Cov(x_2) = \sigma^2 Cov(\boldsymbol{x}_1)$ for covtyp=1. $Cov(\boldsymbol{x}_1) = diag(1, 2, ..., p), Cov(\boldsymbol{x}_2) = \sigma^2 Cov(\boldsymbol{x}_1)$ for covtyp=2. $Cov(\boldsymbol{x}_1) = \boldsymbol{I}, Cov(\boldsymbol{x}_2) = \sigma^2 diag(1, 2, ..., p)$ for covtyp=3. Table 5.10 shows some results. Two lines were used for each simulation scenario, with coverages on the first line and lengths on the second line. When $n_1 = n_2$, the paired test and Li test gave the same results. When n_1/n_2 was not near 1, the Li test had better power and shorter length. Increasing δ could greatly increase the length for the bootstrap test, but the coverage would be 1.

CHAPTER 6

CONCLUSIONS

The one sample test statistic T_n estimates $\boldsymbol{\mu}^T\boldsymbol{\mu}$ and $V(T_n)$ is easy to estimate when $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ is true. Under regularity conditions when H_0 is true, Li (2023) proved that $T_n/\sqrt{V(T_n)} \xrightarrow{D} t_k$ as $p \to \infty$ for fixed $n \geq 3$ where $k = 0.5 * n(n-1) - 1$.

Zhao, Li, Li and Zhang (2024) have an interesting result for the multiple linear regression model

$$Y_i = \alpha + x_{i,1}\beta_1 + \cdots + x_{i,p}\beta_p + e_i = \alpha + \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i \tag{6.1}$$

for $i = 1, ..., n$. Assume that the cases $(\boldsymbol{x}_i^T, Y_i)^T$ are iid with $E(Y) = \mu_Y$, $E(\boldsymbol{x}) = \boldsymbol{\mu}_{\boldsymbol{x}}$ and nonsingular $Cov(\boldsymbol{x}) = \boldsymbol{\Sigma}_{\boldsymbol{x}}$. Let $Cov(\boldsymbol{x}, Y) = \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$. Then testing $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ versus $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ is equivalent to testing $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{0}$ with $\boldsymbol{\mu} = E(z_i) = \boldsymbol{\Sigma}_{\boldsymbol{x}}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)$ where $z_i = (\boldsymbol{x}_i-\boldsymbol{\mu}_{\boldsymbol{x}})(Y_i-\mu_Y-(\boldsymbol{x}_i-\boldsymbol{\mu}_{\boldsymbol{x}})^T\boldsymbol{\beta}_0)$, and the one sample test from Theorem 2 can be applied to $\boldsymbol{w}_i = (\boldsymbol{x}_i-\overline{\boldsymbol{x}})(Y_i-\overline{Y}-(\boldsymbol{x}_i-\overline{\boldsymbol{x}})^T\boldsymbol{\beta}_0)$. Since $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{x}Y}$, using $\boldsymbol{\beta}_0 = \boldsymbol{0}$ gives both a test for $H_0 : \boldsymbol{\beta} = \boldsymbol{0}$ and $H_0 : \boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{0}$. See Olive and Quaye (2025) for applications.

For classification with two groups, let $\boldsymbol{\Sigma}$ be the pooled covariance matrix. Then $\boldsymbol{\beta} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \boldsymbol{0}$ iff $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{0}$, which can be tested with a two sample test. For the importance of $\boldsymbol{\beta}$ in discriminant analysis, see, for example, Wang, Wu, and Wang (2025).

Let the "fail to reject region" be the compliment of the rejection region. Often the fail to reject region is a confidence region for the parameter or parameter vector of interest, where a confidence interval is a special case of a confidence region. For the one sample test, the fail to reject region using V_0 has much more power than using a confidence interval for $\boldsymbol{\mu}^T\boldsymbol{\mu}$. The two sample test statistic $T_N(\boldsymbol{x}, \boldsymbol{y})$ could be used to get a confidence interval for $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2$.

The literature for high dimensional one and two sample tests is rather large. Hu, Tong, and Genton (2024) have many references. Some high dimensional one sample tests include Chen et al. (2011), Feng and Sun (2016), Hyodo and Nishiyama (2017), Park and Ayyala (2013), Srivastava and Du (2008), Wang, Peng, and Li (2015), and Zhao (2017). Hu and Bai (2015) also describes some tests. Chakraborty and Chaudhuri (2017) suggest a method for obtaining a k-sample test of $\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k$ from a one sample test statistic.

Some high dimensional two sample tests include Ahmad (2014), Chen, Li, and Zhong (2019), Feng and Sun (2015), Gregory et al. (2015), Jiang et al. (2022), Xue and Yao (2020), and Zhang et al. (2020). For more on the use of U-statistics for high dimensional methods, see, for example, Xu, Zhu, and Shao (2024).

Several high dimensional two sample tests use the extremely strong assumption that $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$. This assumption is typically stronger than assuming that $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$. See, for example, Huang et al. (2022), Hu and Bai (2015), and Yang, Zheng, and Li (2024).

Simulations were done in *R*. See R Core Team (2024). The collection of Olive (2025) *R* functions *slpack*, available from (http://parker.ad.siu.edu/Olive/slpack.txt), has some useful functions for the inference. The function *hdhot1sim* was used to simulate the four tests, while the function *hdhot1sim2* simulates the first test, which is rather fast. The function *hdhot1sim3* added the test based on sample signs using the fast test. The function *hdhot2sim* simulates the two sample test which applies the fast one sample test on the $\boldsymbol{z}_i = \boldsymbol{x}_{i1} - \boldsymbol{x}_{i2}$ for $i = 1, ..., m$, the Li (2023) test, and the two sample test based on subsampling with $m_i = floor(2n_i/3)$ for i=1,2.

The spatial sign vectors have a some outlier resistance. If the predictor variables are all continuous, the *covmb2* and *ddplot5* functions are useful for detecting outliers in high dimensions. See Olive (2025, $\oint$ 1.4.3) and Olive (2017, pp. 120-123).

REFERENCES

[1] Ahmad, M.R. (2014), "A U-Statistic Approach for a High-Dimensional Two-Sample Mean Testing Problem under Non-Normality and Behrens-Fisher Setting," *The Annals of the Institute of Statistical Mathematics*, 66, 33-61.

[2] Anderson, T.W. (1984), *An Introduction to Multivariate Statistical Analysis,* 2nd ed., Wiley, New York, NY.

[3] Bai, Z.D., and Saranadasa, H. (1996), "Effects of High Dimension: by an Example of a Two Sample Problem," *Statistica Sinica*, 6, 311-329.

[4] Buja, A., and Stuetzle, W. (2006), "Observations on Bagging," *Statistica Sinica*, 16, 323-352.

[5] Chakraborty, A., and Chaudhuri, P. (2017), "Tests for High-Dimensional Data Based on Means, Spatial Signs, and Spatial Ranks," *The Annals of Statistics*, 45, 771-779.

[6] Chen, L.S., Paul, D., Prentice, R.L., and Wang, P. (2011), "A Regularized Hotelling's T^2 Test for Pathway Analysis in Proteomic Studies," *Journal of the American Statistical Association*, 106, 1345-1360.

[7] Chen, S., Li, J., and Zhong, P. (2019), "Two-Sample and ANOVA tests for High Dimensional Means," *The Annals of Statistics*, 47, 1443-1474.

[8] Chen, S.X., and Qin, Y.L. (2010), "A Two Sample Test for High-dimensional Data with Applications to Gene-Set Testing," *The Annals of Statistics*, 38, 808-835.

[9] Feng, L., and Sun, F. (2015), "A Note on High–Dimensional Two–Sample Test," *Statistics & Probability Letters,* 105, 29-36.

[10] Feng, L., and Sun, F. (2016), "Spatial-Sign Based High-Dimensional Location Test," *Electronic Journal of Statistics*, 10, 2420-2434.

[11] Gregory, K.B., Carroll, R.J., Baladandayuthapani, V., and Lahiri, S.N. (2015), "A Two-Sample Test for Equality of Means in High Dimension," *Journal of the American Statistical Association*, 110, 837-849.

[12] Hotelling, H. (1931), "A Generalization of Student's Ratio," *The Annals of Mathematical Statistics*, 2, 360-378.

[13] Hu, J., and Bai, Z. (2015), "A Review of 20 Years of Naive Tests of Significance for High-Dimensional Mean Vectors and Covariance Matrices," *Science China Mathematics*, 55, online.

[14] Hu, Z., Tong, T., and Genton, M.G. (2024), "A Pairwise Hotelling Method for Testing High-Dimensional Mean Vectors," *Statistica Sinica*, 34, 229-256.

[15] Huang, Y., Li, C., Li, R., and Yang, S. (2022), "An Overview of Tests on High- Dimensional Means," *Journal of Multivariate Analysis*, 188, 104813

[16] Hyodo, M., and Nishiyama, T. (2017), "A One-Sample Location Test Based on Weighted Averaging of Two Test Statistics When the Dimension and the Sample Size are Large," *Communications in Statistics: Theory and Methods*, 46, 3526-3541.

[17] Jiang, Y.L., Wang, X.Q., Wen, C.H., Jiang, Y.K., and Zhang, H.P. (2024), "Nonparametric Two-Sample Tests of High Dimensional Mean Vectors via Random Integration," *Journal of the American Statistical Association,* 119, 701-714.

[18] Lehmann, E.L. (1975), *Nonparametrics: Statistical Methods Based on Ranks*, Holden-Day, San Francisco.

[19] Lehmann, E.L. (1999), *Elements of Large–Sample Theory*, Springer, New York, NY.

[20] Li, J. (2023), "Finite Sample *t*-Tests for High-dimensional Means," *Journal of Multivariate Analysis*, 196, 105183.

[21] Olive, D.J. (2017), *Robust Multivariate Analysis*, Springer, New York, NY.

[22] Olive, D.J. (2025), *Prediction and Statistical Learning*, online course notes, see (http:// parker.ad.siu.edu/Olive/slearnbk.htm).

[23] Olive, D.J., and Quaye, P. (2025), "Testing Poisson Regression and Related Models with the One Component Partial Least Squares Estimator," unpublished manuscript at (http://parker.ad.siu.edu/Olive/pphdpois.pdf).

[24] Park, J., and Ayyala, D.N. (2013), "A Test for the Mean Vector in Large Dimension and Small Samples," *Journal of Statistical Planning and Inference*, 143, 929-943.

[25] Pelawa Watagoda, L.C.R., and Olive, D.J. (2021), "Comparing Six Shrinkage Estimators with

Large Sample Theory and Asymptotically Optimal Prediction Intervals," *Statistical Papers*, 62, 2407-2431.

[26] Politis, D.N, and Romano, J.P. (1994), "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions," *The Annals of Statistics*, 22, 2031-2050.

[27] R Core Team (2024), "R: a Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria, (www.R-project.org).

[28] Rajapaksha, K.W.G.D.H., and Olive, D.J. (2024), "Wald Type Tests with the Wrong Dispersion Matrix," *Communications in Statistics: Theory and Methods*, 53, 2236-2251.

[29] Srivastava, M.S., and Du, M. (2008), "A Test for the Mean Vector with Fewer Observations Than the Dimension," *Journal of Multivariate Analysis*, 99, 386-402.

[30] Wang, L., Peng, B., and Li, R. (2015), "A High-Dimensional Nonparametric Multivariate Test for Mean Vector," *Journal of the American Statistical Association*, 110, 1658-1669.

[31] Wang, Y., Wu, Z., and Wang, C. (2025), "High Dimensional Discriminant Analysis under Weak Sparsity," *Communications in Statistics: Theory and Methods*, 54, 2657-2674.

[32] Wu, C.F.J. (1990), "On the Asymptotic Properties of the Jackknife Histogram," *The Annals of Statistics*, 18, 1438-1452.

[33] Xu, T., Zhu, R., and Shao, X. (2024), "On Variance Estimation of Random Forests with Infinite-Order U-Statistics," *Electronic Journal of Statistics*, 18, 2135-2207.

[34] Xue, K., and Yao, F. (2020), "Distribution and Correlation-Free Two-Sample Test of High-Dimensional Mean," *The Annals of Statistics*, 48, 1304-1328.

[35] Yang, S., Zheng, S., and Li, R. (2024), "A New Test for High-Dimensional Two-Sample Mean Problems with Consideration of Correlation Structure," *The Annals of Statistics*, 52, 2217-2240.

[36] Zhang, J.T., Guo, J., Zhou, B., and Cheng, M.Y. (2020), "A Simple Two-Sample Test in High Dimensions Based on L^2-Norm," *Journal of the American Statistical Association*, 115(530):1011-1027.

[37] Zhao, A., Li, C., Li, R., and Zhang, Z. (2024), "Testing High-Dimensional Regression Coef-

ficients in Linear Models," *The Annals of Statistics*, 52, 2034-2058.

[38] Zhao, J. (2017), "A New Test for the Mean Vector in Large Dimension and Small Samples," *Communications in Statistics: Simulation and Computation*, 46, 6115-6128.

VITA

Graduate School
Southern Illinois University

Ahlam Mustafa Abid

ahlam.abid@seu.edu.sa

Taibah University
Bachelor of Science, 2010
University of Dayton
Master of Science, Applied Mathematics, May 2017

Dissertation Paper Title:
Some Simple High Dimensional One and Two Sample Tests

Major Professor: Dr. David Olive

Publications:
Positive solutions of boundary value problems for ordinary differential equations with dependence on higher order derivatives (with Professor Paul Eloe), Communications on Applied Nonlinear Analysis, Volume 24(2017), Number 4, 101-108.