Some Simple High Dimensional One and Two Sample Tests

Ahlam M. Abid and David J. Olive*
Southern Illinois University

February 28, 2024

Abstract

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.

KEY WORDS: Hotelling's T^2 Test, Paired t Test, Subsampling, U-Statistics.

*David J. Olive is Professor, School of Mathematical & Statistical Sciences, Southern Illinois University, Carbondale, IL 62901, USA.

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 section reviews some tests while the following section gives 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 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 $T_H^2 \xrightarrow{D} \chi_p^2$ 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

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

as $n, p \to \infty$. The next test is attributed to Bai and Saranadasa (1996). Suppose $p/n \to c > 0$. 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 need $p/n \to c > 0$ or $p/n^2 \to 0$. Hence p can not be too big.

For test statistic T_n, let $V(T_n)$ be the variance of T_n and let $s_n^2 = \hat{V}(T_n)$ be a consistent estimator of T_n. Then there are test statistics T_n for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$, 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. 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

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

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 $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ and additional regularity conditions,

$$\frac{T_n}{\sqrt{V(T_n)}} \xrightarrow{D} N(0,1) \quad \text{and} \quad \frac{\mathrm{T_n}}{\mathrm{s_n}} \xrightarrow{D} \mathrm{N}(0,1) \tag{2}$$

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 $\boldsymbol{x}_j$:

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

The T_n in Equation (1) 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 $\boldsymbol{\mu} = \boldsymbol{0}$ iff $\boldsymbol{\mu}^T \boldsymbol{\mu} = 0$ and $E(\boldsymbol{x}_i^T \boldsymbol{x}_j) = \boldsymbol{\mu}^T \boldsymbol{\mu}$ if $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are iid with $E(\boldsymbol{x}_i) = \boldsymbol{\mu}$ and $i \neq j$.

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 $\boldsymbol{z}_i = ss(\boldsymbol{x}_i)$ where the spatial sign function $ss(\boldsymbol{x}_i) = \boldsymbol{0}$ if $\boldsymbol{x}_i = \boldsymbol{0}$, 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 $\boldsymbol{0}$. Let $T_n(\boldsymbol{w})$ denote the statistic T_n computed from an iid sample $\boldsymbol{w}_1, ..., \boldsymbol{w}_n$. Since the $\boldsymbol{z}_i$ are iid if the $\boldsymbol{x}_i$ are iid, use $T_n(\boldsymbol{z})$ to test $H_0 : \boldsymbol{\mu}_{\boldsymbol{z}} = \boldsymbol{0}$ versus $H_A : \boldsymbol{\mu}_{\boldsymbol{z}} \neq \boldsymbol{0}$ where $\boldsymbol{\mu}_{\boldsymbol{z}} = E(\boldsymbol{z}_i)$. In general, $\boldsymbol{\mu}_{\boldsymbol{z}} \neq \boldsymbol{\mu} = \boldsymbol{\mu}_{\boldsymbol{x}} = E(\boldsymbol{x}_i)$, but $\boldsymbol{\mu}_{\boldsymbol{z}} = \boldsymbol{\mu} = \boldsymbol{0}$ 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 $T_n(\boldsymbol{z})$ 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)

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

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 $m = \text{floor(n/2)} = \lfloor \text{n/2} \rfloor$ 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 simple. 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 $\theta = Cov(W_{ij}, W_{id})$ 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 $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ 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), let $W_{ij} = \boldsymbol{x}_i^T \boldsymbol{x}_j = W_{ji}$, and note that

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

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

$$Cov\left(\sum_{i <} \sum_{j} W_{ij}, \sum_{k <} \sum_{d} W_{kd} \right) = \sum_{i <} \sum_{j} \sum_{k <} \sum_{d} Cov(W_{ij}, W_{kd}). \tag{3}$$

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

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

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

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

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

$$\sum_d \sum_t [E(x_{jd})E(x_{kt})\ Cov(x_{id}, x_{it})].$$

Under H_0, $\boldsymbol{\mu} = 0$ and thus $E(x_{jd}) = E(x_{kt}) = 0$. Hence $\theta = 0$. $\square$

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 $\theta = V(Z(\boldsymbol{x}_i)) = V(\boldsymbol{x}_i^T \boldsymbol{\mu}) \geq 0$. 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

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

if $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ 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 $p^\gamma/n \to 0$ where $\gamma > 0.5$ and $\boldsymbol{\mu} = \boldsymbol{0}$. Suppose $\boldsymbol{\mu} = \delta \boldsymbol{1}$ where the constant $\delta > 0$ and $\boldsymbol{1}$ 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

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

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

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 $V(T_m^*) = E([T_m^*]^2) - [T_n]^2 = Cov(T_m^*, T_m^*)$.

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 $\boldsymbol{x}_{1,1}, ..., \boldsymbol{x}_{n_1,1}$ 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 $E(\boldsymbol{x}_i) = \boldsymbol{\mu}_i$ 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 $T_C^2 > m F_{m, n-2, 1-\alpha}$.

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 $\boldsymbol{z}_i = \boldsymbol{x}_{i1} - \boldsymbol{x}_{i2}$ for $i = 1, ..., m$. Then apply the one sample test from Theorem 2 to the $\boldsymbol{z}_i$. In low dimensions, it is known that there are better tests. In high dimensions, the power technique below Theorem 2 may be useful.

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

$$\boldsymbol{y}_i = \boldsymbol{x}_{i1} - \sqrt{\frac{n_1}{n_2}}\boldsymbol{x}_{i2} + \frac{1}{\sqrt{n_1 n_2}} \sum_{j=1}^{n_1} \boldsymbol{x}_{j2} - \overline{\boldsymbol{x}}_2 = \boldsymbol{x}_{i1} - \sqrt{\frac{n_1}{n_2}}\boldsymbol{x}_{i2} + \boldsymbol{a}_{n_1,n_2} - \overline{\boldsymbol{x}}_2$$

for $i = 1, ..., n_1$. Note that $\boldsymbol{y}_i = \boldsymbol{z}_i = \boldsymbol{x}_{i1} - \boldsymbol{x}_{i2}$ 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 $\boldsymbol{y}_i$.

Let T_{j,n_j} denote the one sample test statistic applied to the $\boldsymbol{x}_{ij}$ for $j = 1, 2$. Then the statistic $T_{1n} - T_{2n}$ could be used to test $H_0 : \boldsymbol{\mu}_1^T \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2^T \boldsymbol{\mu}_2$. However, it is possible that $\boldsymbol{\mu}_1^T \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2^T \boldsymbol{\mu}_2$ even if $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$.

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 $vec(\boldsymbol{A})$ stack the columns of matrix $\boldsymbol{A}$ so that $\boldsymbol{c} = vec(\boldsymbol{X}_1^T) = [\boldsymbol{x}_{11}^T, \boldsymbol{x}_{12}^T, ..., \boldsymbol{x}_{1n_1}^T]^T$. 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

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

Note that $\|\boldsymbol{a} - \boldsymbol{b}\|^2 = \boldsymbol{a}^T \boldsymbol{a} + \boldsymbol{b}^T \boldsymbol{b} - 2\boldsymbol{a}^T \boldsymbol{a}$, and let

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

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).

4 SIMULATIONS

In the simulations, we examined four one sample tests. The first two tests used Theorem 2 b) and Equation 2): $T_n/s_n \xrightarrow{D} N(0,1)$ if s_n^2 is a consistent estimator of $V(T_n)$. The first test used $\hat{\sigma}_W^2 = S_W^2$ based on Theorem 1. The second test used

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

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 should be asymptotically equivalent to the Li (2023) test.

These tests computed intervals

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

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 fourth "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 (2023, chapter 2) and Pelawa Watagoda and Olive (2021). All four tests rejected H_0 if 0 was not in the interval. The fourth "test" is ad hoc since it has not yet been proven to have level α. Since $nT_n = n\overline{\boldsymbol{x}}^T\overline{\boldsymbol{x}} - tr(\boldsymbol{S})$, the bootstrap test is also a competitor for the test based on Z_2. The fifth test used the Theorem 2 test applied to the spatial sign vectors with S_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 $\delta = 1$. The simulation used 5000 runs, the 4 $\boldsymbol{x}$ distributions, and the 3 matrices $\boldsymbol{A}$. For the third $\boldsymbol{A}$, we used $\psi = 1/\sqrt{p}$.

5 CONCLUSIONS

The 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 $T_n/V(T_n) \xrightarrow{D} t_k$ as $p \to \infty$ for fixed $n \geq 3$ where $k = n(n-1)/2 - 1$.

The literature for high dimensional one and two sample tests is rather large. Hu, Tong, and Genton (2024) have many references. Two sample tests that need $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ may not work well since the assumption of equal covariance matrices rarely holds. 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).

Simulations were done in *R*. See R Core Team (2020). The collection of Olive (2023) *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 *hdhot3sim* 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 $i = 1, ..., m$ 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 (2023, $\oint$ 1.4.3) and Olive (2017, pp. 120-123).

6 References

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.

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

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

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

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.

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.

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

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.

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

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

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.

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

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.

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

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.

Jiang, Y.L., Wang, X.Q., Wen, C.H., Jiang, Y.K., and Zhang, H.P. (2022), "Nonparametric Two-Sample Tests of High Dimensional Mean Vectors via Random Integration," *Journal of the American Statistical Association,* to appear.

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

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

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

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

Olive, D.J. (2023), *Prediction and Statistical Learning*, online course notes, see (http://parker.ad.siu.edu/Olive/slearnbk.htm).

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.

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.

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.

R Core Team (2020), "R: a Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria, (www.R-project.org).

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.

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.

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.

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

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

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.

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.