Journal of Econometrics and Statistics
Vol. 6, Issue 2, 2026, pp. 157-167

A two-sample mean test for high dimensional data

Qingyue Yang and Lihong Wang[a]

School of Mathematics, Nanjing University, Nanjing, China

ABSTRACT
In this article we consider the hypothesis test problem of the mean vector of two multivariate random variables. In the case of high-dimension and small sample size, the Hotelling's T^2 test is no longer applicable because of the singularity of the sample covariance matrix. We propose a procedure to divide the high-dimensional variables into multiple groups, and construct a grouped Hotelling test statistic which averages the T^2 statistics of each group. Numerical simulations and real data analysis show the good performance of the test in identifying the difference between the two population means.

KEYWORDS
Grouped Hotelling's T^2 test, High-dimensional data, Mixed χ^2 distribution, Small sample size.

1. Introduction

High-dimensional data analysis is a central topic of modern statistical science and has been receiving growing attention from both researchers and practitioners. In many applications, high-dimensional data, where the number of features or covariates are larger than the number of samples, are commonly encountered. One such example is the colonoscopy video data set, which contains 76 gastrointestinal lesion samples with 55 benign lesions and 21 malignant lesions, each having 464 pathological features. It is of interest to verify whether there is a significant difference in the mean vector of the pathological features between benign and malignant cases. The two-sample mean testing is a fundamental problem in high-dimensional statistical inference. This is the issue we intend to consider in this article.

Suppose we have independent samples from two multivariate populations X_{i1}, X_{i2}, $\cdots, X_{in_i}$, $i = 1, 2$ with

$$\mathrm{E}(X_{i1}) = \mu_i, \ \ \mathrm{cov}(X_{i1}) = \Sigma, \ \ i = 1, 2,$$

where $X_{ij} = (X_{ij1}, X_{ij2}, \cdots, X_{ijp})'$, $j = 1, \cdots, n_i$, and $p > n = n_1 + n_2$. The primary focus is to test the equality of mean vectors,

$$H_0 : \mu_1 = \mu_2 \longleftrightarrow H_1 : \mu_1 \neq \mu_2.$$

CONTACT Author[a]. Email:lhwang@nju.edu.cn

Article History
Received: 17 May 2026; Revised: 16 June 2026; Accepted: 21 June 2026; Published: 13 July 2026

To cite this paper

Qingyue Yang & Lihong Wang. (2026). A two-sample mean test for high dimensional data. *Journal of Econometrics and Statistics.* 6(2), 157-167.

For normally distributed data, Hotelling's T^2 test is one of the best choices provided $p \leq n$ (c.f. Hotelling[5]),

$$T^2 = \frac{n_1 n_2}{n}(\bar{X}_1 - \bar{X}_2)' \Big(\frac{A_1 + A_2}{n-2}\Big)^{-1}(\bar{X}_1 - \bar{X}_2), \tag{1}$$

where $\bar{X}_1$ and $\bar{X}_2$ are the sample mean vectors of the two samples, and $A_i = \sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)'$, $i = 1, 2$. Under the null hypothesis, T^2 has $T^2(p, n-2)$ distribution. However, the Hotelling's T^2 test can not be applied due to the singularity of $A_1 + A_2$ when $p > n$.

In order to overcome the infeasibility of Hotelling's T^2 statistic in high-dimensional data, various methods have been proposed from different directions (see, e.g., Chen and Qin[1], Dempster [2], Feng et al.[3], Ghosh et al. [4], Qiu et al.[6], Srivastava[8], Srivastava and Du[9], Thulin[10], Wu et al.[11], Zhang and Pan[12], Zhang et al. [13]). For instance, Wu et al.[11] constructed a PCT statistic, which is obtained by summing squared componentwise t-statistics,

$$\text{PCT} = \frac{\sum_{j=1}^{p} t_j^2}{p}, \tag{2}$$

where t_j is the t-statistic for the j-th component, $j = 1, \cdots, p$,

$$t_j = \sqrt{\frac{n_1 n_2}{n}} \frac{|X_{1\cdot j} - X_{2\cdot j}|}{\sqrt{\big\{\sum_{l=1}^{n_1}(X_{1lj} - X_{1\cdot j})^2 + \sum_{l=1}^{n_2}(X_{2lj} - X_{2\cdot j})^2\big\}/(n-2)}},$$

where $X_{i\cdot j}$ is the average of X_{ilj}, $l = 1, \cdots, n_i$, $i = 1, 2$. Wu et al.[11] used $c\chi^2(d)$ as the approximate null distribution of PCT statistic. They also provided the estimators of the parameters c and d.

Motivated by the idea of Wu et al.[11], we introduce a procedure to divide the high-dimensional variables into multiple groups, and propose a grouped Hotelling test statistic which averages the T^2 statistics of each group. That is, we partition p variables into m groups, each having p_k variables, where $\sum_{k=1}^{m} p_k = p$ and $p_k < n$ for $k = 1, \cdots, m$. Then we calculate Hotelling's statistic T_k^2 for each group, and use the average T_H as the test statistic.

The detailed testing procedure is introduced in the next section. In section 3, numerical simulation and empirical analysis are provided to compare the T_H statistic with some other two-sample mean tests for high-dimensional data in the literature. The simulation results demonstrate the good performance of the proposed method in identifying the difference between the two population means.

2. Grouped Hotelling test

The choice of p_k and the grouping procedure play an important role in effectively carrying out the grouped Hotelling test. The p_k's should not be too close to n, as the poor performance of Hotelling's T^2 test in this setting leads to the poor power of the test. However, if p_k is too small, most of the multivariate structure of the data is lost. The key point for the partition is to select the number of variables in each group as many as possible so that the variables having large correlation are placed in the same group. In this way, the correlation between groups can reach a relatively small value, or even be ignored, then we can avoid the calculation of covariance between groups in estimating the variance of the test statistic.

Let $p_1 = \cdots = p_k = p^*$ and $m = p/p^*$ is an integer. In view of the lack of theoretical results on the optimal choice of p^*, we select several different $p^* \in [\lfloor n/4 \rfloor, \lfloor 3n/4 \rfloor]$ in the simulations. The numerical results suggest that p^* around $\lfloor 2n/3 \rfloor$ seems to be an appropriate choice for finite sample sizes, and the power of the test is not sensitive to the small changes of p^* around $\lfloor 2n/3 \rfloor$.

Next we consider the grouping procedure which is mainly based on the sample correlation coefficient matrix $S = (s_{ij})_{p\times p}$. Regardless of the diagonal elements, we first find out the maximum absolute value $|s_{i_0 j_0}|$ of the matrix S, then put the two variables with subscripts i_0 and j_0 into the same group, and set $\mathcal{A} = \{i_0, j_0\}$ and $\mathcal{B} = \{s_{i_0 j_0}\}$. Secondly, we search in the i_0-th row and j_0-th column of the matrix S to find the maximum absolute values $|s_{i_1 j_0}|$ (or $|s_{i_0 j_1}|$) in the complement set of $\mathcal{B}$. It is then added to the set $\mathcal{B}$, and i_1 (or j_1) is also added to the set $\mathcal{A}$. Repeating this selection procedure $p^* - 1$ times, we obtain the first group of p^* variables with highest correlations. Next, the sample correlation coefficient matrix is recalculated for the remaining $p - p^*$ variables, and the second group is selected by using the above steps. Finally the p variables are divided into $m = p/p^*$ groups.

Then, using (1), we calculate the Hotelling's T^2 statistic only using the p^* variables in the k-th group,

$$\begin{aligned} T_k^2 &= \frac{n_1 n_2}{n}(X_{1\cdot k}^* - X_{2\cdot k}^*)'\Big(\frac{A_{1k} + A_{2k}}{n-2}\Big)^{-1}(X_{1\cdot k}^* - X_{2\cdot k}^*), \quad k = 1, \cdots, m, \\ A_{ik} &= \sum_{j=1}^{n_i}(X_{ijk}^* - X_{i\cdot k}^*)(X_{ijk}^* - X_{i\cdot k}^*)', \quad i = 1, 2, \end{aligned} \tag{3}$$

where $X_{i\cdot k}^*$, $k = 1, \cdots, m$, $i = 1, 2$, denotes the mean of the k-th group for the i-th sample. Now the final test statistic is defined as

$$T_H = \frac{1}{m}\sum_{k=1}^{m} T_k^2. \tag{4}$$

The exact null distribution of T_H is difficult to determine, an appropriate approximation is needed. Based on the proposed grouping algorithm, one may assume that the T_k^2's are approximately independent. A normal distribution seems to be a good choice by the central limit theorem.

In Figure 1, we display the histograms of the test T_H under H_0 over 2000 simulation runs with the two samples generated from the p-dimensional normal distribution $N_p(0, \Sigma)$ with $\Sigma = \rho^{|i-j|}$, $p = 1000$, $n_1 = 30$, $n_2 = 50$, and $p^* = 50$. The histogram for $\rho = 0.01$ gives a positive support to the normal approximation of the null distribution of T_H. However, when $\rho = 0.2$ and 0.9, inspection of the histograms reveals that a normal approximation may be inappropriate.

The above grouping algorithm is summarized as follows.

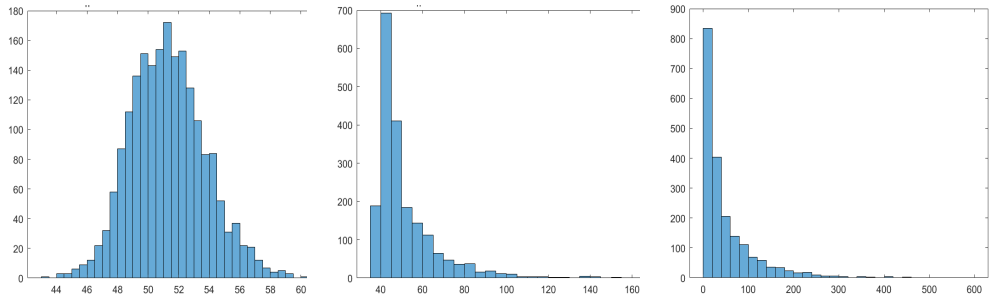

Figure 1.: The histograms of the test T_H under H_0 with $\rho = 0.01, 0.2$ and 0.9 from left to right

Algorithm 1 Grouping algorithm

1: Determine the number of variables p^* in each group, calculate the sample correlation coefficient matrix S, let $k = 1$.
2: **while** $k <= p/p^*$ **do**
3: Set the elements on the diagonal of S to be zero, and let $\mathcal{A}$ to be empty.
4: Find the largest absolute value $|s_{i_0 j_0}|$ in the matrix S, add its subscripts to the set $\mathcal{A}$, and set the value of the (i_0, j_0)-th element in S to be zero.
5: **while** The number of elements in $\mathcal{A}$ is less than p^* **do**
6: Select the element with the largest absolute value from the elements with row and column numbers in $\mathcal{A}$, add its subscripts to the set $\mathcal{A}$, and set the value of this element in S to be zero.
7: **end while**
8: Take the p^* elements in $\mathcal{A}$ as the k-th group, calculate the Hotelling's T^2 statistic by (3) only using the p^* variables of the sample, and store it in T_k^2.
9: Remove the variables in $\mathcal{A}$ from the original sample, and reassign S to be the correlation coefficient matrix of the remaining $p - p^*$ dimensional data.
10: **end while**
11: Using T_k^2, $k = 1, 2, \cdots, m$, to calculate T_H by (4).

In light of the quadratic-form structure of T_H, a scaled chi-squared distribution $c\chi^2(d)$ is a natural candidate to use as an approximation (see, e.g., Wu et al.[11]). Note that, if T_H is approximately $c\chi^2(d)$ distributed, then $c = \text{Var}(T_H)/\{2\text{E}(T_H)\}$ and $d = 2\text{E}^2(T_H)/\text{Var}(T_H)$. To determine the parameters c and d, it suffices to derive $\text{E}(T_H)$ and $\text{Var}(T_H)$ or their estimators.

Under the null hypothesis, $T_k^2 \sim T^2(p^*, n-2)$. Therefore,

$$\begin{aligned} \text{E}(T_H) = \text{E}(T_k^2) &= \frac{p^*(n-2)}{n-p^*-1}\text{E}(F_H) \\ &= \frac{p^*(n-2)}{n-p^*-1} \times \frac{n-p^*-1}{n-p^*-3} = \frac{p^*(n-2)}{n-p^*-3}, \end{aligned} \tag{5}$$

where $F_H \sim F(p^*, n-p^*-1)$.

It is observed that the mean of the test statistic depends only on n and p^*. The computation of the variance of $\text{Var}(T_H)$ is complicated, so we simplify it by assuming the mutual

independence between the groups. Then we have the estimator of $\mathrm{Var}(T_H)$ as

$$\begin{aligned}\widehat{\mathrm{Var}(T_H)} &= \Big(\frac{p^*(n-2)}{n-p^*-1}\Big)^2 \frac{\mathrm{Var}(F_H)}{m} \\ &= \Big(\frac{p^*(n-2)}{n-p^*-1}\Big)^2 \frac{2(n-p^*-1)^2(n-3)}{mp^*(n-p^*-3)^2(n-p-5)} \\ &= \frac{1}{m}\frac{2p^*(n-2)^2(n-3)}{(n-p^*-3)^2(n-p^*-5)}.\end{aligned} \tag{6}$$

(5) and (6) lead to

$$\hat{c} = \frac{\widehat{\mathrm{Var}(T_H)}}{2\mathrm{E}(T_H)}, \quad \hat{d} = \frac{2\mathrm{E}^2(T_H)}{\widehat{\mathrm{Var}(T_H)}}.$$

Let $\chi^2_{1-\alpha}(\hat{d})$ stands for the $(1-\alpha)\times 100\%$-th quantile of the chi-squared distribution with the degrees of freedom $\hat{d}$. We have the rejection region for the hypothesis test (1),

$$\mathcal{X} = \left\{T_H > \hat{c}\chi^2_{1-\alpha}(\hat{d})\right\}.$$

3. Numerical and empirical studies

The purpose of this section is to compare the performance of the proposed approach with several existing ones. We also apply the proposed test for empirical analyses of human gait feature data and the colonoscopy video data.

3.1. Simulation study

In this subsection we investigate the numerical performance of the proposed method through simulation. We generate data from $N_p(\mu_i, \Sigma)$ and consider three types of covariance structures:

1. $\Sigma = D^{1/2}RD^{1/2}$, $D = diag(d_1^2, \cdots, d_p^2)$, $d_i \sim U(2,3)$, $R = (r_{ij})$, $r_{ij} = (-1)^{i+j}\rho^{|i-j|^{0.1}}$.
2. $\Sigma = D^{1/2}RD^{1/2}$, $D = diag(d_1^2, \cdots, d_p^2)$, $d_i \sim U(2,3)$, $R = (r_{ij})$, $r_{ij} = \rho, i \neq j$.
3. $\Sigma = U'AU$, $A = diag(1 + \rho\mathrm{rand}(p,1))$, $U = 1.2\mathrm{orth}(p,p)$, where $\mathrm{rand}(p,1)$ and $\mathrm{orth}(p,p)$ stand for the p-dimensional random number vector and the $p \times p$ random orthonormal matrix, respectively.

In order to investigate the performance of the suggested test, we introduce two benchmark tests. The first test statistic is the PCT defined in (2). The second test statistic T_d is provided by Zhang et al. [13], where the sample covariance matrix in Hotelling's T^2 statistic (1) is replaced by its diagonal matrix,

$$T_d = \frac{n_1 n_2}{n}(\bar{X}_1 - \bar{X}_2)'\hat{D}^{-1}(\bar{X}_1 - \bar{X}_2),$$

where $\hat{D} = diag(\hat{\Sigma})$, $\hat{\Sigma} = \frac{A_1+A_2}{n-2}$. Zhang et al.[13] proved that T_d has an asymptotic $\chi^2(d)/d$

distribution, and d is estimated by $\hat{d} = \frac{p^2}{\widehat{tr(\mathcal{R}^2)}}$, where

$$\widehat{tr(\mathcal{R}^2)} = \frac{(n_1+n_2-2)^2}{(n_1+n_2)(n_1+n_2-3)}\{tr(\hat{\mathcal{R}}^2) - \frac{1}{n_1+n_2-2}tr^2(\hat{\mathcal{R}})\},$$

and $\hat{\mathcal{R}} = \hat{D}^{-1/2}\hat{\Sigma}\hat{D}^{-1/2}$.

Set the nominal significance level α is equal to 5%. Now we report the comparison of the performance of the tests PCT, T_d and T_H. Table 1 summarises the simulation results obtained for Models 1, 2, 3 with different values of p, n_1, n_2 and ρ, when the null hypothesis holds. The empirical size is approximated by an average over 2000 simulation runs for each scenario. We define "ARE" as

$$\text{ARE} = \frac{\sum_{i=1}^{M}|size_i - \alpha|}{M\alpha} \times 100,$$

which is adopted to measure the overall performance of a test in terms of maintaining the nominal size α (see, e.g., Zhang et al.[13]), where $size_i$ is the empirical size of the i-th test, M is the total number of tests conducted. For example, M is equal to 6 in Table 1. Obviously, smaller ARE implies better overall performance of the test. Table 2 exhibits the overall AREs of the tests PCT, T_d and T_H, and their AREs for Models 1, 2, and 3, respectively. It is observed that the overall ARE of T_H is the smallest.

Tables 1 and 2 reveal that, for Models 1 and 2, the ARE values of PCT are the smallest. While T_H outperforms the other tests for Model 3, its ARE is significantly smaller than the other two methods. This is easy to understand from the testing procedure of T_H. The random covariance structure of Model 3 makes the grouping algorithm very efficient. The correlation between groups is very small, so that we can achieve much accurate parameter estimation for the approximated distribution of test statistic T_H. This leads to the best performance of T_H. However, for Model 2 and most cases of Model 1, the correlation between variables is either fixed or relatively large. In this situation the grouping procedure can't effectively reduce the correlation between groups. This results in the poor performance of T_H. Therefore, in terms of the size control, PCT has the best performance for Models 1 and 2, and T_H performs best for Model 3.

Next, we consider the alternative hypothesis. Let $\mu_1 = 0$ and $\mu_2 = 10\delta\mathbf{t}/\|\mathbf{t}\|$, where $\mathbf{t} = (1, 2, \cdots, p)'$ and $\delta = 0.2, 0.4, 0.6, 0.8$. In a similar fashion, the computation results under this setting are presented in Table 3. Figure 2 depicts the empirical powers of PCT, T_d and T_H with $p = 300$ for different δ and ρ values under three covariance matrix structures, respectively.

The simulation results in Table 3 and Figure 2 indicate the best performance of T_H in most cases, especially for Models 1 and 2, while the performance of PCT is the worst. It is also noted that all three tests have poor performance in the case of fixed correlation between variables as in Model 2. Generally speaking, the above simulation evidence reveals that the proposed test T_H is a good alternative to the existing methods.

Table 1.: The empirical sizes (%) for various scenarios under H_0: $\mu_1 = \mu_2 = 0$

Model	p	n_1	n_2	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.8$		
				PCT	T_d	T_H	PCT	T_d	T_H	PCT	T_d	T_H
1	100	20	30	2.15	7.40	5.80	5.30	6.75	5.70	5.05	5.80	6.15
	100	30	40	3.55	7.85	5.20	4.95	5.90	5.70	5.80	6.35	5.80
	100	40	50	4.35	7.70	6.45	5.45	6.65	6.35	5.35	6.00	6.35
	300	20	30	1.95	9.95	5.40	6.00	8.00	7.30	5.70	6.70	7.50
	300	30	40	2.80	9.90	5.90	5.50	6.80	7.20	5.40	6.00	7.80
	300	40	50	3.30	8.40	6.80	7.40	8.40	7.80	4.90	5.90	9.40
		ARE		39.7	70.7	18.5	15.7	28.8	41.2	6.50	21.3	45.2
2	100	20	30	4.90	7.75	6.40	5.65	6.50	5.85	6.25	7.10	6.05
	100	30	40	5.75	8.75	6.80	5.30	6.00	7.15	5.65	6.15	6.95
	100	40	50	4.65	6.15	6.45	6.30	6.65	6.65	5.55	6.15	7.00
	300	20	30	4.40	8.35	6.95	5.35	6.65	6.05	4.50	5.15	7.25
	300	30	40	4.80	8.00	8.20	5.40	6.05	7.45	4.85	5.30	7.80
	300	40	50	5.05	6.70	7.35	5.65	6.15	8.20	4.50	5.00	7.55
		ARE		6.83	52.3	40.5	12.2	26.7	37.8	12.0	16.2	42.0
3	100	20	30	0.30	10.35	5.00	0.50	10.35	5.15	0.45	11.00	5.35
	100	30	40	0.75	9.35	5.25	0.80	8.10	4.95	1.15	8.95	6.55
	100	40	50	1.50	7.90	6.30	0.85	8.10	4.80	1.05	7.50	5.40
	300	20	30	0	14.95	4.60	0	15.80	5.55	0	14.90	5.50
	300	30	40	0	9.75	5.15	0	11.20	5.40	0	10.15	4.75
	300	40	50	0.05	9.60	5.40	0.25	10.40	5.45	0	9.25	4.80
		ARE		91.33	106.3	8.33	92.00	113.2	6.00	91.17	105.8	10.83

Table 2.: The comparison of AREs for three tests

Model	PCT	T_d	T_H
1	20.6	44.68	32.4
2	10.3	31.7	40.1
3	91.50	108.4	8.39
Overall	40.80	61.7	26.96

Table 3.: The empirical powers (%) for various scenarios under H_1

Model	p	n_1	n_2	δ	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.8$		
					PCT	T_d	T_H	PCT	T_d	T_H	PCT	T_d	T_H
1	100	20	30	0.2	5.25	14.20	11.90	6.15	7.70	22.20	5.75	6.50	56.80
				0.4	15.95	45.95	53.35	9.05	12.10	84.20	5.35	6.60	100.0
				0.6	75.15	95.05	96.5	18.60	24.85	99.95	9.50	11.10	100.0
				0.8	99.80	100.0	100.0	57.45	73.45	100.0	17.25	20.10	100.0
	300	20	30	0.2	2.20	12.85	9.70	6.10	7.65	14.65	5.20	6.00	29.80
				0.4	4.00	23.25	30.65	6.45	8.30	57.50	5.95	7.50	98.95
				0.6	10.55	53.30	77.30	9.15	12.25	98.30	6.15	7.20	100.0
				0.8	40.95	92.25	98.05	12.70	17.45	100.0	7.70	9.40	100.0
2	100	20	30	0.2	8.70	13.60	7.10	7.80	9.30	6.15	6.55	7.60	8.00
				0.4	21.65	29.30	13.35	11.00	13.05	8.95	9.55	11.25	11.90
				0.6	45.25	52.60	23.20	22.20	24.75	14.40	14.50	15.95	17.40
				0.8	68.20	76.70	37.20	33.75	36.20	21.30	22.05	24.10	30.55
	300	20	30	0.2	6.20	10.55	8.35	5.50	6.75	8.95	5.30	5.75	9.55
				0.4	9.40	14.95	10.55	7.65	8.80	9.20	6.10	7.20	9.30
				0.6	17.80	25.10	16.15	9.85	11.75	11.55	7.95	8.55	13.50
				0.8	26.35	36.10	25.10	14.60	16.30	15.35	10.45	11.60	15.55
3	100	20	30	0.2	1.45	22.20	10.15	1.50	21.85	10.80	1.55	18.70	8.90
				0.4	21.45	71.00	39.50	16.35	64.55	34.35	13.80	57.00	27.75
				0.6	86.85	99.05	86.25	79.10	98.40	82.00	70.10	96.30	75.05
				0.8	99.95	100.0	99.60	99.70	100.0	99.00	99.10	100.0	98.65
	300	20	30	0.2	0	23.35	8.65	0.10	22.15	8.85	0.10	20.85	7.00
				0.4	0.30	53.70	22.70	0.25	47.05	18.70	0.30	42.35	16.45
				0.6	8.05	90.75	59.45	4.75	85.75	50.45	3.55	80.70	43.65
				0.8	60.80	99.85	92.55	43.95	99.55	87.60	29.40	98.60	80.90

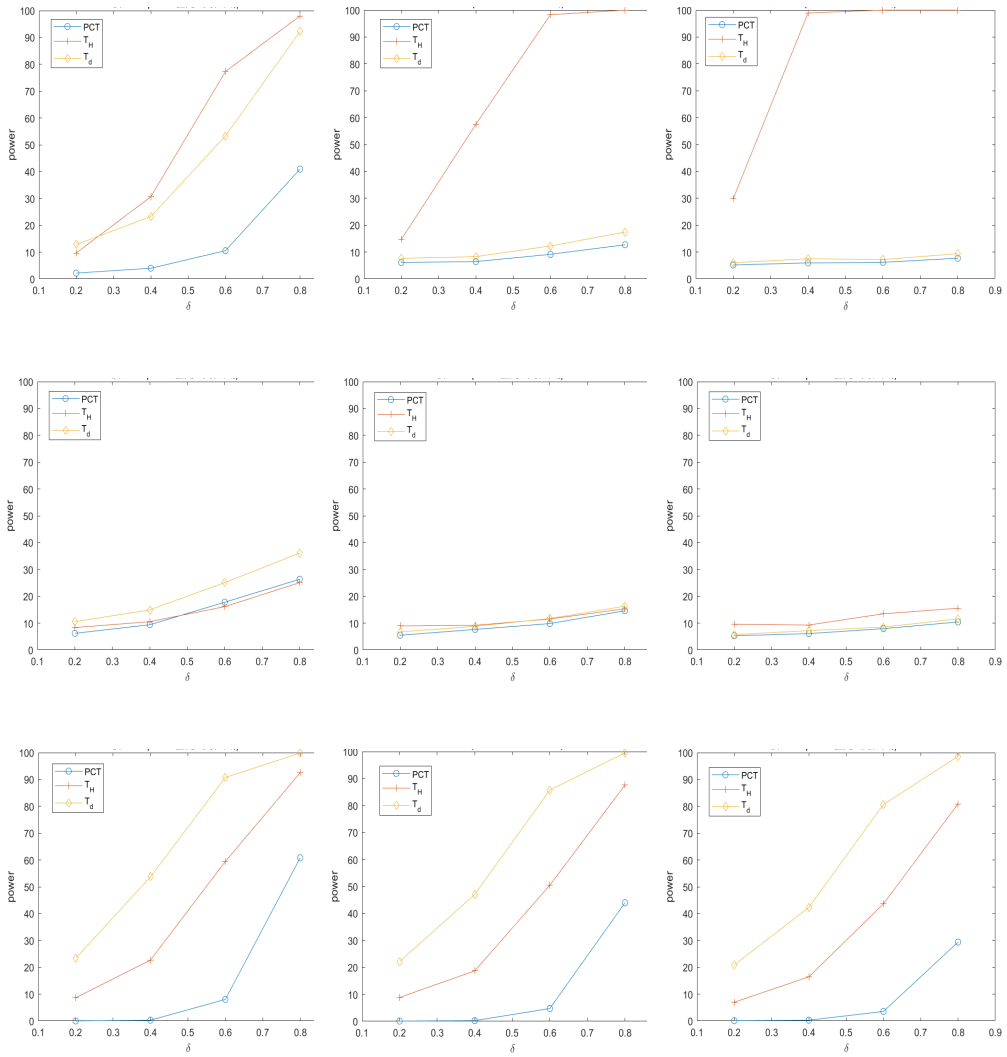

Figure 2.: The empirical powers of PCT, T_d and T_H under different δ values for Model 1(top), Model 2(middle), Model 3(bottom) with $\rho = 0.2, 0.5, 0.8$

3.2. Empirical study

In this subsection we apply the proposed method to two real data sets, which are available at http://archive.ics.uci.edu/ml/datasets.

The first data set is the human gait feature data. The data set includes 320 gait features (basic features, instantaneous features, spatial features, height features, etc.), and is divided into two populations according to gender. The sample size of the first population composed of women is 21, and that of the second population composed of men is 27. The total sample size 48 is far less than 320.

Now we use the three methods PCT, T_d and T_H to test whether there are significant differences in the 320 gait features between the two samples. The results are shown in Table 4. It can be seen that all the tests reject the null hypothesis at the significance level 5% but with different p-values.

Table 4.: Two sample mean tests for the human gait feature data

Method	Test statistic	p-value
PCT	2.04	0.0100
T_d	652.73	0.0011
T_H	45.37	0.0228

The second data set we considered is the colonoscopy video data regarding the characteristics of gastrointestinal lesions. There are 76 samples in this data set with 464 variables for each sample, including texture features, color features, shape features, etc. The original data are divided into two categories: malignant adenoma and benign adenoma. The sample sizes of the first and second category are 55 and 21 respectively. The purpose of analyzing the data is to judge whether the variable information obtained from the colonoscopy video data can be used as the basis for the diagnosis of the disease. To this end, it is necessary to test whether there is significant difference in the mean vectors of these pathological features between benign and malignant cases. The computation results based on three tests are recorded in Table 5. All three tests reject the null hypothesis with significant small p-values. Both real data analyses show that the proposed test is reliable in practice.

Table 5.: Two sample mean tests for the colonoscopy video data

Method	Test statistic	p-value
PCT	6.1105	3.0947e-09
T_d	2835.3	8.4784e-11
T_H	48.6296	0

References

[1] Chen, S.X., 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.

[2] Dempster, A.P., 1958. A high dimensional two sample significance test. The Annals of Mathematical Statistics 29, 995–1010.

[3] Feng, L., Zou, C., Wang, Z. Zhu, L., 2015. Two sample Behrens-Fisher problem for high-dimensional data. Statistica Sinica 25, 1297–1312.

[4] Ghosh, S., Ayyala, D.N., Hellebuyck, R., 2021. Two-sample high dimensional mean test based on prepivots. Computational Statistics and Data Analysis 163, 107284.

[5] Hotelling, H., 1931. The generalization of student's ratio. The Annals of Mathematical Statistics 2, 360–378.

[6] Qiu T., Xu W., Zhu, L., 2021. Two-sample test in high dimensions through random selection. Computational Statistics and Data Analysis 160, 107218.

[7] Srivastava, M.S., 2005. Some tests concerning the covariance matrix in high dimensional data. Journal of the Japan Statistical Society 35, 251–272.

[8] Srivastava, M.S., 2009. A test for the mean vector with fewer observations than the dimension under non-normality. Journal of Multivariate Analysis 100, 518–532.

[9] Srivastava, M.S., Du, M., 2008. A test for the mean vector with fewer observations than the dimension. Journal of Multivariate Analysis 99, 386–402.
[10] Thulin, M., 2014. A high-dimensional two-sample test for the mean using random subspaces. Computational Statistics and Data Analysis 74, 26–38.
[11] Wu, Y., Genton, M.G., Stefanski, L.A., 2006. A multivariate two-sample mean test for small sample size and missing data. Biometrics 62(3), 877–885.
[12] Zhang, J., Pan, M., 2016. A high-dimension two-sample test for the mean using cluster subspaces. Computational Statistics and Data Analysis 97, 87–97.
[13] Zhang, L., Zhu, T., Zhang, J.T., 2020. A simple scale-invariant two-sample test for high-dimensional data. Econometrics and Statistics 14, 131–144.