Validating Medical Treatment Effects by Projected *F*-tests under High Dimension with a Small Sample Size

Jiajuan Liang[1,2,*]

[1]*Department of Statistics and Data Science, Beijing Normal-Hong Kong Baptist University, 2000 Jintong Road, Tang Jia Wan, Zhuhai 519087, China*

[2]*Guangdong Provincial/Zhuhai Key Laboratory of Interdisciplinary Research and Application for Data Science Beijing Normal-Hong Kong Baptist University, 2000 Jintong Road, Tang Jia Wan, Zhuhai 519087, China*

Abstract: This paper introduces a statistical method for validating treatment effects in high-dimensional medical data with small sample sizes. The method compares multiple multivariate population means under multivariate normality, using spherical matrix distribution theory and principal component analysis (PCA) for dimension reduction. The resulting test statistic follows an exact F-distribution under the null hypothesis of equal means, even when the sample size is smaller than the data dimension. Unlike classical MANOVA, the approach does not require equal covariance matrices across groups, making it more robust for real-world biomedical data where variance-covariance homogeneity rarely holds. Monte Carlo simulations show the test achieves accurate type I error control and favorable power. Application to real medical datasets with high-dimensional biomarkers further demonstrates its practicality and interpretability. This work provides a rigorous and versatile advancement for high-dimensional inference in biomedical research and related fields.

Keywords: Dimension reduction, *F*-test, Monte Carlo study, Multiple mean comparison, Principal component analysis.

1. INTRODUCTION

In medical research, validating treatment effects often involves comparing multiple means across different patient groups based on clinical or biological data. When the number of variables (data dimension) exceeds the total sample size, traditional multivariate analysis of variance (MANOVA) methods fails due to the singularity of the sample covariance matrix. This issue arises frequently in high-dimensional medical studies, such as gene expression analysis and imaging-based diagnostics, where the number of measured features far exceeds the number of available patient samples. To address this challenge, we propose a projected *F*-test approach that combines Principal Component Analysis (PCA) for dimension reduction with the theory of spherical matrix distributions [1], allowing for valid statistical inference in small-sample, high-dimensional settings.

It is well known that the MANOVA framework relies on the assumption that the sample covariance matrix is nonsingular, and the covariance matrices across groups or populations must be identical. However, in high-dimensional settings with limited samples and in many practical situations, these assumptions may not hold, rendering the standard MANOVA approach inapplicable or lack of theoretical justification. While several alternative approaches have been proposed to extend multiple mean comparison methods to high-dimensional data, they often rely on asymptotic approximations or nonparametric resampling techniques, which may be computationally intensive and difficult to implement in practice [2, 3]. To overcome these limitations, researchers have explored random projection methods, shrinkage estimators, and regularized MANOVA techniques [4, 5]. Although these methods improve the stability of covariance matrix estimation, they require additional tuning parameters or large-sample approximations, making their practical implementation challenging. Our PCA-based approach offers an alternative that is computationally efficient and directly applicable using existing statistical tables.

Our proposed method employs PCA to reduce the high-dimensional data to a lower-dimensional subspace that captures most of the variability in the data. By projecting the high-dimensional mean comparison problem onto this subspace, we obtain a test statistic that follows a known distribution under the normal assumption without requiring equal covariance matrices across groups. This avoids the reliance on asymptotic approximations and enables straightforward implementation with the classical *F*-distribution tables. The PCA-based approach extends the classical approaches to high-dimensional multiple mean comparison from two aspects: sample size requirements and covariance homogeneity requirement. The PCA-based projected *F*-test provides an effective alternative to existing methods by leveraging the concentration of high-dimensional data in low-dimensional subspaces [6]. Compared to large-sample-theory approaches, our method remains valid even with small sample sizes and covariance heteroscedasticity across populations, making it particularly suitable for high-dimensional biomedical studies with limited patient data.

*Address correspondence to this author at the Department of Statistics and Data Science, Beijing Normal-Hong Kong Baptist University, 2000 Jintong Road, TangJia Wan, Zhuhai 519087, China; E-mail: jiajuanliang@uic.edu.cn

The remainder of this paper is structured as follows. Section 2 introduces the mathematical formulation of the PCA-based projected *F*-test. Section 3 presents a Monte Carlo study on the power performance by choosing different projection dimensions. Section 4 applies the method to real-world medical datasets. Finally, Section 5 gives some concluding remarks.

2. THE PROJECTED *F*-TEST

Suppose that we want to compare the mean effects from *k* medical treatments $T_1, \ldots, T_k$ based on *k* sets of *p*-dimensional observations:

$$S_l=\{\mathbf{x}_{lj} = (x_{lj1}, \ldots, x_{ljp})^t:\ p\times 1,\ j = 1, \ldots, n\},\ l = 1, \ldots, k \quad (1)$$

where the superscript "*t*" stands for the transpose of a row vector or a matrix. We assume normal samples $\{\boldsymbol{x}_{lj} : j=1, \ldots, n\}$ are i.i.d. *p*-dimensional normal $N_p(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)$ with possible covariance heteroscedasticity. That is, it is possible that $\Sigma_l \neq \Sigma_s$ for some set of (l, s). Here we consider a balanced sample design with an equal sample size across different treatments. Let

$\boldsymbol{\mu}_l$ = the mean treatment effect from treatment l, $l = 1, \ldots, k$. (2)

We want to test the hypothesis

$$H_0 :\ \boldsymbol{\mu}_1 = \ldots = \boldsymbol{\mu}_k \quad (3)$$

against the alternative that at least two means are not equal. It is well known that hypothesis (3) is a classical MANOVA problem when assuming equal covariance matrices $\Sigma_1 = \cdots = \Sigma_k$ [7] and the total sample size $N = kn > p$. When there exists covariance heteroscedasticity or $N = kn \leq p$, the traditional MANOVA approach is no longer applicable. Some large-sample methodologies were proposed to test hypothesis (3) [2, 8, 9]. Our purpose is to develop an exact test for hypothesis (3) under possible covariance heteroscedasticity and $N = kn \leq p$.

Under the assumption of balanced sample design (1), choose a reference sample, say, the last sample without loss of generality, and define

$$X_s = \begin{pmatrix} \boldsymbol{x}'_{\boldsymbol{s1}} \\ \vdots \\ \boldsymbol{x}'_{\boldsymbol{sn}} \end{pmatrix}, n\times p, s = 1, \ldots, k \quad (4)$$

and the random matrix

$$X = \begin{pmatrix} x'_{11} - x'_{k1} & x'_{21} - x'_{k1} & \cdots & x'_{k-1,1} - x'_{k1} \\ x'_{12} - x'_{k2} & x'_{22} - x'_{k2} & \cdots & x'_{k-1,2} - x'_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ x'_{1n} - x'_{kn} & x'_{2n} - x'_{kn} & \cdots & x'_{k-1,n} - x'_{kn} \end{pmatrix} = (5)$$

$(X_1 - X_k, \ldots, X_{k-1} - X_k), n\times(k-1)p.$

Under the null hypothesis (3), the rows of X are independent and have a $(k-1)p$-dimensional normal distribution with zero mean vector and covariance matrix

$$\Sigma = \mathrm{diag}(\Sigma_1, \ldots, \Sigma_{k-1}) + J_{k-1}\otimes\Sigma_k, \quad (6)$$

where $\mathrm{diag}(\Sigma_1, \ldots, \Sigma_{k-1})$ is a block-diagonal matrix with the $(k-1)$ blocks as specified, J_{k-1} stands for the $(k-1)\times(k-1)$ matrix of all elements "1", and "$\otimes$" stands for the Kronecker product of matrices. Under the notations in equations (4)-(6), we have the following theorem.

Theorem 1. Under the notations in (4)-(6) and the null hypothesis (3), we define the eigenvalue-eigenvector problem

$$\left(\frac{1}{n}X'X\right)D = D\Lambda \quad (7)$$

where $\boldsymbol{D} = (\boldsymbol{d}_1, \ldots, \boldsymbol{d}_q)$, $n\times q$, $q=\min\{n,(k-1)p\}-1$. $\boldsymbol{D}$ consists of q eigenvectors $\{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_q\}$ associated with the q positive eigenvalues of the non-negative definite matrix $\frac{1}{n}X'X$. $\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1, \ldots, \lambda_q)$ consists of the eigenvalues $\lambda_1 \geq \ldots \geq \lambda_q > 0$. Let

$$\mathbf{D}_r = (\mathbf{d}_1, \ldots, \mathbf{d}_r):\ n\times r,\ r = 1, \ldots, q.\ \boldsymbol{Z}_r = \mathbf{X}\ \mathbf{D}_r,$$

$$H = Z_r^t\left(\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^t\right)Z_r, G = Z_r^t\left(I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^t\right)Z_r, \quad (8)$$

where $\mathbf{1}_n$ stands for the vector of ones with dimension $n\times 1$, $\boldsymbol{I}_n$ for the identity matrix with dimension $n\times n$, $r = 1, \ldots, \min(n,q)-1$. Define the statistic

$$F_r = \frac{n-r}{r}\mathrm{trace}(HG^{-1}) = \frac{n-r}{nr}\mathbf{1}_n^t Z_r G^{-1} Z_r^t \mathbf{1}_n. \quad (9)$$

Then F_r has an *F*-distribution $F(r, n-r)$.

Proof: under equations (3)-(5), each row of the matrix X has a $(k-1)p$ -dimensional normal distribution with covariance matrix

$$cov(x_{1i} - x_{ki}, x_{2i} - x_{ki}, \ldots, x_{k-1,i} - x_{ki})$$
$$= \begin{pmatrix} \Sigma_1 + \Sigma_k & \Sigma_k & \cdots & \Sigma_k \\ \Sigma_k & \Sigma_2 + \Sigma_k & \cdots & \Sigma_k \\ \vdots & \vdots & \vdots & \vdots \\ \Sigma_k & \Sigma_k & \cdots & \Sigma_{k-1} + \Sigma_k \end{pmatrix}$$
$$= \begin{pmatrix} \Sigma_1 & 0 & \cdots & 0 \\ 0 & \Sigma_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \Sigma_{k-1} \end{pmatrix}$$
$$+ \begin{pmatrix} \Sigma_k & \Sigma_k & \cdots & \Sigma_k \\ \Sigma_k & \Sigma_k & \cdots & \Sigma_k \\ \vdots & \vdots & \vdots & \vdots \\ \Sigma_k & \Sigma_k & \cdots & \Sigma_k \end{pmatrix} =$$
$$= \mathrm{diag}(\Sigma_1, \ldots, \Sigma_{k-1}) + J_{k-1}\otimes\Sigma_k$$

as given by equation (6), for $i = 1,2, \ldots, n$. The rest of the proof can be derived from Theorem 1 in [10].

The idea of using the sample PCA defined by equation (7) for dimension reduction is due to Läuter (1996, [10]) when developing a high-dimensional mean test with a small size and possible cases of $n \leq p$ in comparing medical effects with multiple endpoints. Theorem 1 can be considered as a corollary of Theorem 1 of Läuter (1996, [10]). So we do not repeat the theoretical details that are related to higher level of the theory of spherical matrix distributions [1]. Readers who are not familiar with the matrix-distributional theory can refer to [1, 10]. More discussions on choie of the projection direction matrix D_r in $Z_r=XD_r$ in (8) can be also referred to [11].

The statistic F_r in (9) provides an exact *F*-test for hypothesis (3) for each $r = 1, \ldots, \min(n,q)-1$ for any balanced sample size n across different groups or populations. There is no requirement of equal covariance matrices. Therefore, Theorem 1 provides a new way to compare high-dimensional means under the normal assumption. A large value of the statistic F_r in (9) implies hypothesis (3) should be rejected for any $r = 1, \ldots, \min(n,q)-1$. This is an exact test without requirement of large sample size and covariance homogeneity. The next section gives a Monte Carlo study on the empirical performance of the F_r in (9) for different choices of the projection dimension *r*.

3. A MONTE CARLO STUDY

In this section, we choose the following sampling designs to study the power performance of the statistic F_r in (9) under both homogeneous and heterogeneous covariance assumptions across the populations. Because the null distribution of the F_r-statistic (9) does not depend on the normal means and the covariance matrix under the null hypothesis (3), we choose the following sample designs for the mean vectors and covariance matrices from the populations for multiple mean comparison in our Monte Carlo study.

Design A: k =2 groups and dimension p = 10, 20, 40 with sample sizes $n_1 = n_2 = 20$

With

$$\mu_1 = 1_p, \mu_2 = d1_p, \Sigma_s = 5^{s-1}(\rho_{ij}), p\times p, s = 1,..,k \quad (10)$$

where the constant d controls the difference between the two mean vectors, and

$$\rho_{ij} = \begin{cases} 1, & i = j \\ 0.5, & i \neq j \end{cases} \quad (11)$$

for $i, j = 1, \ldots, p$.

Design B: k =4 groups and dimension p = 10, 20, 40 with sample sizes n_1 =…=

n_4 = 20 with

$$\boldsymbol{\mu}_1 = \mathbf{1}_p, \ \boldsymbol{\mu}_2 = d\mathbf{1}_p \ \boldsymbol{\mu}_3 = 2d\mathbf{1}_p \quad (12)$$

with the same covariance matrices as given in (10).

Design C: k =8 groups and dimension p =10, 20, 40 with sample sizes $n_1 = \ldots =$

n_8 =20 with

$$\boldsymbol{\mu}_i = (i-1)d\mathbf{1}_p, \ i = 1, \ldots, 10. \quad (13)$$

and the same covariance matrices as given in (10).

Design D: k =12 groups and dimension p =10, 20, 40 with sample sizes $n_1 = \ldots =$

n_{12} = 20 with the same mean and covariance structure as in Design 3.

The multivariate normal samples are generated from each of the above designs for 2,000 replications. The empirical power for each design is computed by counting the relative rejection frequency using the significance level α =0.05. Large values of the statistics in (9) imply rejection of hypothesis (3). The power comparison for the above four designs is illustrated by Figures **1-2**, where the exact *F* -tests are from the PCA projected *F* -test given by (9) with different projection dimensions $r_1=[r/4]$, $r_2=[r/3]$, $r_3=[r/2]$, and $r_4=[3r/4]$, here $[\cdot]$ stands for the integer part of a real number. Figures **1-2** show that the projected F-test performs the best with smaller projection dimension like $r_1=[r/4]$ when the number of groups is small; while it performs the best with larger projection dimension like $r_4=[3r/4]$ when the number of groups is large.

The choices of different projection dimensions $r_1=[r/4]$, $r_2=[r/3]$, $r_3=[r/2]$, and $r_4=[3r/4]$ employed in the Monte Carlo study illustrate different empirical power performances. In practical experimental designs, the number of PCA directions can be determined by proportion of variation explanation. For example, let $\lambda_1, \ldots, \lambda_q$ (q=min{$n,(k-1)p$}−1) be the positive eigenvalues from (7). An empirical recommendation for choosing the number of projection directions r was proposed by Jilliffe (2002, [12]): choosing r such that

$$\frac{\sum_{i=1}^{r-1}\lambda_i}{\sum_{i=1}^{q}\lambda_i} < 80\% \ and \ \frac{\sum_{i=1}^{r}\lambda_i}{\sum_{i=1}^{q}\lambda_i} \geq 80\%.$$

Or replace the ratio 80% with any higher percentage if requiring higher explanation of variation.

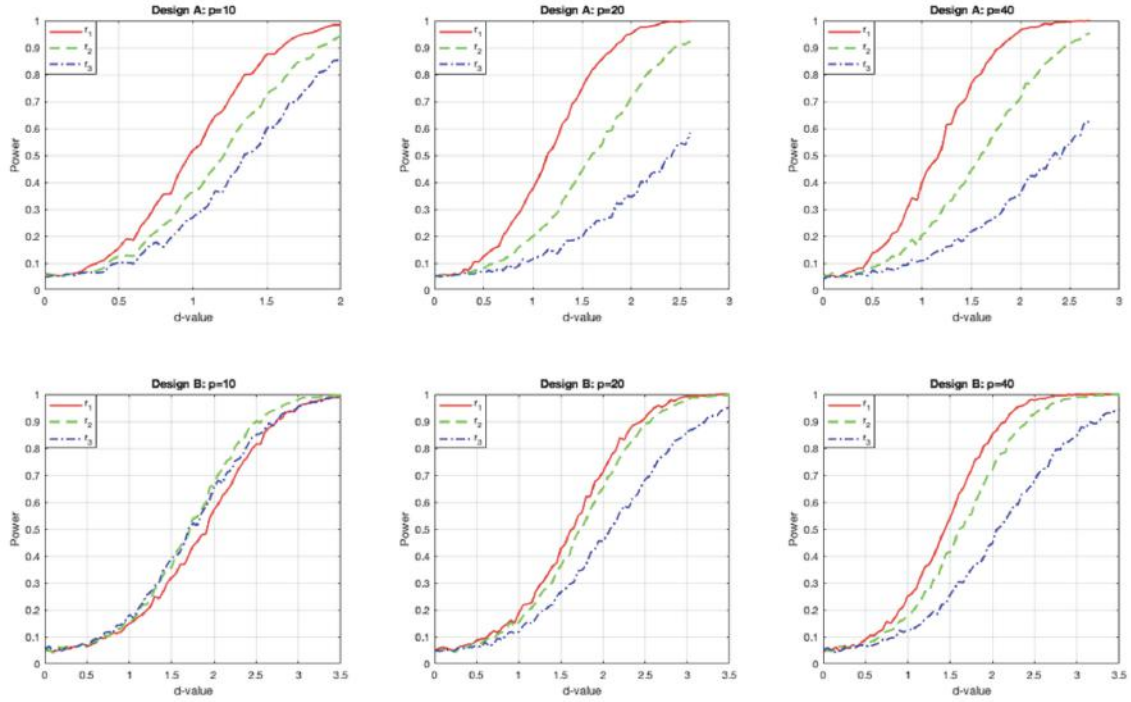

Figure 1: Power comparison among three exact F-tests: $F_1=F_{r1}$, $F_2=F_{r2}$, $F_3=F_{r3}$. Design A contains two groups of data with dimensions p=10, 20, and 40; Design B contains four groups of data with dimensions p=10, 20, and 40.

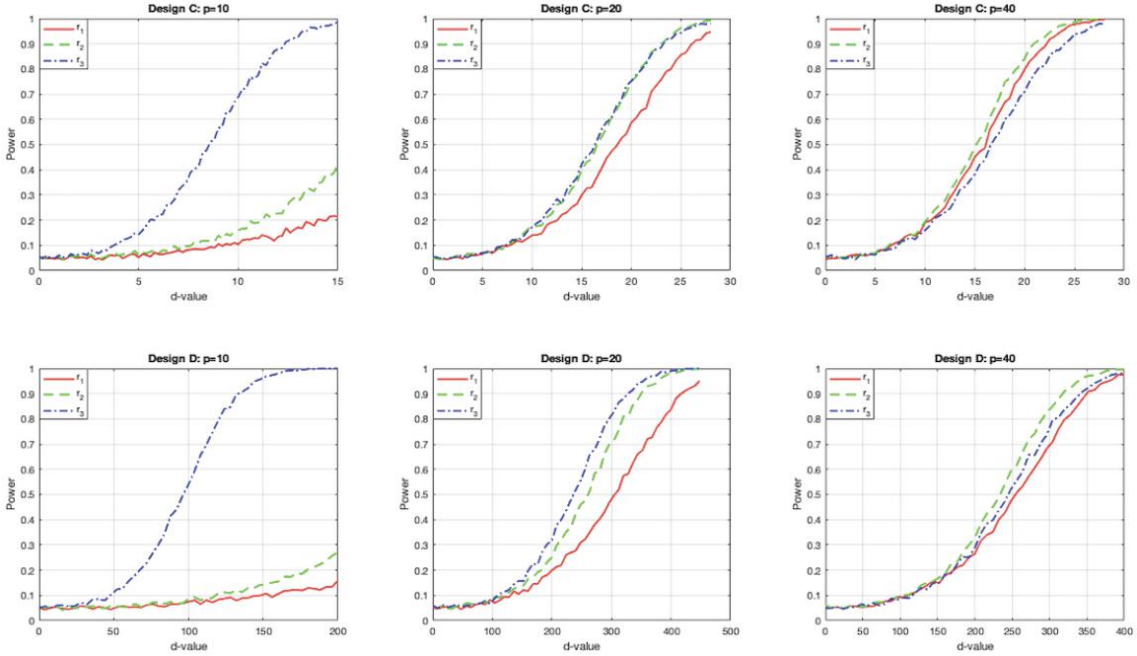

Figure 2: Power comparison among three exact F-tests: $F_1=F_{r1}$, $F_2=F_{r2}$, $F_3=F_{r3}$. Design C contains eight groups of data with dimensions p=10, 20, and 40; Design D contains twelve groups of data with dimensions p=10, 20, and 40.

4. AN ILLUSTRATIVE EXAMPLE

Example. A research project was carried out by Tianjin Medical University, China [13]. Rats were collected for experiment by four different treatments (doses) to see the treatment effects from 46 genes with sample size n = 6 (rats) for each treatment. In the experiment on 6 rats, the ratio of organ wet weight to body weight (organ coefficient) was observed. The purpose is to evaluate organ development during the treatment. Details on the experiment and medical analysis can be found in [13]. Cao *et al*. [14] proposed an exact one-way ANOVA approach to comparing the mean expression levels of individual genes without assuming equal variances across genes. Here we apply the exact F-test in Theorem 1 to test equality of the multivariate mean expression levels among the six treatments. This belongs to multivariate analysis of variance (MANOVA) without assuming equal covariance matrices across different

treatments. Among the observation matrices $\{X_s: n\times p, n=6, p=46, s=1,2,3,4\}$ from the four treatments, we choose the last treatment dataset X_4 as a reference and construct the overall observation matrix as follows.

$$X=(X_1-X_4, X_2-X_4, X_3-X_4): 6\times[(4-1)46]=6\times 138.$$

Equations (5)-(9) are applied to this dataset X and two projection dimensions for *r* in (9) are chosen as $r_1=1$ and $r_2=2$ according to the empirical performance in the Monte Carlo study in section 3 with $r=\left[\frac{q}{3}\right]=\left[\frac{5}{3}\right]=1$ and $r=\left[\frac{q}{2}\right]=\left[\frac{5}{2}\right]=2, q=\min\{n,(k-1)p\}-1$ in Theorem 1, here $n=6, k=4, p=46, q=5$. In order to see that the possibility of treatment difference may come from the paired difference, we also construct the paired difference matrices:

$$X_{12}=X_1-X_2, X_{13}=X_1-X_3, X_{14}=X_1-X_4,$$

$$X_{23}=X_2-X_3, X_{24}=X_2-X_4, X_{34}=X_3-X_4$$

and apply Theorem 1 to these matrices with the data analysis summarized in Tables **1-4** as follows.

Summary conclusions on the high-dimensional mean comparisons from the four datasets in Tables **1-4**:

1. Table **1** shows that there is no significance overall difference between the mean levels of any two treatments or among the four treatments from the male-ARC data although some individual mean difference exists for some of the genes among the 46 genes as identified by Gao *et al*. [13] and Table 5 in [15]. F_{r_2} -test F_{14} in Table **1** has a

Table 1: Projected *F*-Tests for Male-ARC Data from [15]

F-test	F_1	F_{12}	F_{13}	F_{14}	F_{23}	F_{24}	F_{34}
F_{r_1}-value	1.0951	0.0169	4.2697	1.3932	0.9864	0.3982	0.3735
p-value	0.3433	0.9015	0.0937	0.2909	0.3662	0.5557	0.5678
F_{r_2}-value	0.6366	3.1681	1.7080	7.0387	1.3898	0.2480	0.9344
p-value	0.5754	0.1498	0.2909	0.0490	0.3481	0.7916	0.4645

Remark: F_1 for the overall dataset X; F_{12} for X_{12}; F_{13} for X_{13}; F_{14} for X_{14}; F_{23} for X_{23}; F_{24} for X_{24}; F_{34} for X_{34}.

Table 2: Projected *F*-Tests for Male-Neonatal Data from [15]

F-test	F_1	F_{12}	F_{13}	F_{14}	F_{23}	F_{24}	F_{34}
F_{r_1}-value	0.9905	1.5341	2.5197	0.9746	0.2364	0.9835	1.0136
p-value	0.3653	0.2705	0.1733	0.3689	0.6474	0.3669	0.3603
F_{r_2}-value	0.6274	0.6198	1.0670	1.2552	10.880	0.5668	0.6032
p-value	0.5795	0.5828	0.4252	0.3775	0.0241	0.6071	0.5903

Table 3: Projected *F*-Tests for Male-AVPV Data from [15]

F-test	F_1	F_{12}	F_{13}	F_{14}	F_{23}	F_{24}	F_{34}
F_{r_1}-value	0.3216	0.2769	0.0564	0.3786	0.3434	0.0465	1.0969
p-value	0.5952	0.6212	0.8217	0.5653	0.5833	0.8379	0.3429
F_{r_2}-value	0.3491	0.6494	0.6103	0.1522	10.603	0.3170	0.8017
p-value	0.7249	0.5698	0.5870	0.8636	0.0252	0.7451	0.5096

Table 4: Projected *F*-Tests for Male-MPN Data from [15]

F-test	F_1	F_{12}	F_{13}	F_{14}	F_{23}	F_{24}	F_{34}
F_{r_1}-value	5.4993	0.1128	5.1921	10.257	1.6292	1.8910	0.2157
p-value	0.0660	0.7506	0.0717	0.0239	0.2579	0.2275	0.6618
F_{r_2}-value	2.8738	0.1247	13.874	9.2656	2.4821	2.0049	0.1051
p-value	0.1684	0.8861	0.0159	0.0315	0.1991	0.2494	0.9027

p-value=0.0490<0.05, which indicates a marginal significance between treatment 2 and treatment 3 from the male-ARC data.

2. Table **2** shows that there is no significance overall difference among the four treatments from the male-neonatal data although some individual mean difference exists for some of the genes among the 46 genes as identified by Gao *et al*. [13] and Table **4** in [15]. F_{r_2}-test F_{23} in Table **2** has a *p*-value=0.0241<0.05, which indicates a marginal significance between treatment 2 and treatment 3 from the male-neonatal data.

3. Table **3** shows that there is no significance overall difference among the four treatments from the male-AVPV data although some individual mean difference exists for some of the genes among the 46 genes as identified by Gao *et al*. [13] and Table 6 in [15]. F_{r_2}-test F_{23} in Table **3** has a *p*-value=0.0252<0.05, which indicates a marginal significance between treatment 2 and treatment 3 from the male-AVPV data.

4. Table **4** shows that there is no significance overall difference among the four treatments from the male-MPN data although some individual mean difference exists for some of the genes among the 46 genes as identified by Gao *et al*. [13] and Table 7 in [15]. Both F_{r_1}-test and F_{r_2}-test F_{14} in Table **4** have a p-value under the significance level 0.05, showing a significance difference between treatment 1 and treatment 4 from the male-MPN data. The F_{r_2}-test F_{13} in Table **4** has a *p*-value=0.0159, indicating a significant difference between treatment 1 and treatment 3 from the male-MPN data.

Tables **1-4** show the pairwise mean comparison among gene expression levels from four groups of rats. The classical MANOVA method based on the *F*-test requires the equality of covariance matrices across the groups. The *F*-test (9) removes this assumption that implies more flexibility in the MANOVA methodology for high-dimensional mean comparison. It is generally known that rejection of the multiple equal-mean hypothesis does not imply rejection of pairwise mean comparison. It is also true the acceptance (or non-rejection) of the multiple equal-mean hypothesis does not imply the acceptance (or non-rejection) of pairwise mean comparison. The classical MANOVA cannot be applied to the datasets in Tables 1-4 because the total sample size $N = 4\times6 = 24$ with a dimension *p=46*. But our PCA-based *F*-test (9) can be still applied to carry out the overall multiple mean comparison across the four groups of rats and it can be also applied to the pairwise comparisons among the four groups of rats, which show some of the pairwise comparisons are significant at level 0.05. In practical biological or clinical multiple tests of means for validating medical treatments, it is a technical problem about how to calibrate the *p*-values from dependent pairwise tests. This is a relatively big topic related to meta-analysis that cannot be covered in this paper because of limited space. Interested readers can refer to the books [16, 17].

5. CONCLUDING REMARKS

This paper introduces a new and rigorous approach for comparing multiple high-dimensional population means under the framework of multivariate analysis, specifically designed to handle scenarios where the number of variables far exceeds the total number of observations. The proposed test statistic is derived using the theory of spherical matrix distributions and the dimension-reducing properties of principal component analysis. Unlike classical methods such as MANOVA, which require restrictive assumptions like equal covariance matrices and larger sample sizes than the number of variables, the new test overcomes these limitations and remains valid under more realistic conditions.

A principal contribution of this work is the establishment of an exact *F*-distribution for the proposed test statistic under the null hypothesis of equal means across the populations. This exactness is not asymptotic but holds regardless of the data dimension or sample size, provided the populations have equal sample sizes. Moreover, the null distribution is invariant to the assumption of homogeneity of covariances, making the test particularly robust in heterogeneous settings, such as those frequently encountered in medical and biological data analyses. The exact F-test without the assumption of equal population covariance matrices generalizes the exact *F*-test in [14] which was developed under the assumption of equal population covariance matrices.

Through simulation studies, we have shown that the test maintains type I error rates and achieves reasonable power in detecting true mean differences, even in the presence of small sample sizes and unequal covariance structures. The application to real medical datasets further demonstrates the method's practicality and relevance in contemporary research settings.

In summary, this work extends the classical MANOVA framework to high-dimensional data under small sample sizes and covariance heterogeneity across groups, offering a theoretically grounded, computationally feasible, and empirically robust tool for modern statistical inference. It should be pointed out that there are some recent developments in high-dimensional inference on mean comparisons, such as the shrinkage-based MANOVA [18], random projections [19], or regularized covariance testing methods [2, 4, 5], these methods either require generating a series of man-made random matrices, or large sample size, or a high dimension to approximate the null distribution of their test statistics. Our proposed *F*-test (9) does not require these assumptions and has an exact F-distribution under the null hypothesis for any fixed equal sample size n across groups and fixed dimension p no matter $n > p$ or $n \leq p$. Although we are not able to present a comprehensive power comparison between our test and these recently developed tests, we believe that our test possesses some flexibility in the case of any fixed sample size and fixed dimension. Future work may explore extensions to unequal sample sizes, non-Gaussian data, or incorporating sparsity structures in the covariance matrices to further broaden the applicability of the proposed method.

ACKNOWLEDGEMENTS

The author thanks Prof. Na Gao at Tianjin Medical University, China, for providing the real medical datasets in the example of this paper.

Disclosure statement: the author claims no conflict of interest during the preparation of the paper and use of commonly available data for this submission.

Our work was supported by the Guangdong Provincial Key Laboratory of IRADS (2022B1212010006).

REFERENCES

[1] Fang KT, Zhang Y, Generalized Multivariate Analysis, Springer-Verlag and Science Press, Berlin/Beijing 1990.

[2] Bai Z, Saranadasa H, Effect of high dimension: by an example of a two-sample problem. Statistica Sinica 1996; 6: 311-329.

[3] Srivastava MS, Kubokawa T, Tests for multivariate analysis of variance in high dimension under non-normality. Journal of Multivariate Analysis 2013; 115: 204-216. https://doi.org/10.1016/j.jmva.2012.10.011

[4] Cai TT, Liu W, Xia Y, Two-sample test of high dimensional means under dependence. Journal of Royal Statistical Society (Series B) 2014; 76: 349-372. https://doi.org/10.1111/rssb.12034

[5] Chen SX, Qin YL, A two-sample test for high-dimensional data with applications to gene-set testing. Annals of Statistics 2010; 38: 808-835. https://doi.org/10.1214/09-AOS716

[6] Johnstone IM, On the distribution of the largest eigenvalue in principal component analysis. Annals of Statistics 2001; 29: 295-327. https://doi.org/10.1214/aos/1009210544

[7] Anderson TW, An Introduction to Multivariate Statistical Inference (3rd Edition). John Wiley & sons, New Jersey, USA 2003.

[8] Yamadaa T, Himenob T, Testing homogeneity of mean vectors under heteroscedasticity in high-dimension. Journal of Multivariate Analysis 2015; 139: 7-27. https://doi.org/10.1016/j.jmva.2015.02.005

[9] Zhang JT, Guo J, Zhou B, Linear hypothesis testing in high-dimensional one-way MANOVA. Journal of Multivariate Analysis 2017; 155: 200-216. https://doi.org/10.1016/j.jmva.2017.01.002

[10] Läuter J, Exact t and F tests for analyzing studies with multiple endpoints. Biometrics 1996; 52: 964-970. https://doi.org/10.2307/2533057

[11] Läuter J, Glimm E, Kropf S, New multivariate tests for data with an inherent structure. Biometrical Journal 1996; 38(1): 5-23. https://doi.org/10.1002/bimj.4710380102

[12] Jolliffe IT, Principal Component Analysis. 2nd ed. New York: Springer 2002.

[13] Gao N, Hu R, Huang Y, Dao L, Zhang C, Liu Y, Wu L, Wang X, Gore AC, Sun Z, Specific effects of prenatal DEHP exposure on neuroendocrine gene expression in the developing hypothalamus of male rats. Archives of Toxicology 2018; 92: 501-512. https://doi.org/10.1007/s00204-017-2049-z

[14] Cao Y, Liang J, Multiple mean comparison for clusters of gene expression data through the t-SNE plot and PCA dimension reduction, International Journal of Statistics in Medical Research 2025; 14: 1-14. https://doi.org/10.6000/1929-6029.2025.14.01

[15] Cao Y, Liang J, Gao N, Sun Z, A new method for identifying significant genes from gene expression data. Biometrics and Biostatistics International Journal 2022; 11: 140-146. https://doi.org/10.15406/bbij.2022.11.00368

[16] Borenstein M (Ed.), Meta-analysis: A guide to calibrating and combining statistical evidence. Wiley 2024.

[17] Westfall PH, Young SS, Resampling-based multiple testing: Examples and methods for p-value adjustment. John Wiley & Sons 1993.

[18] Srivastava MS, Du M, A test for the mean vector with fewer observations than the dimension. Journal of Multivariate Analysis 2008; 99(3): 386-402. https://doi.org/10.1016/j.jmva.2006.11.002

[19] Srivastava R, Li P, Ruppert D, RAPTT: An exact two-sample test in high dimensions using random projections. Journal of Computational and Graphical Statistics 2016; 25(3): 954-970. https://doi.org/10.1080/10618600.2015.1062771

Received on 19-10-2025 Accepted on 21-11-2025 Published on 09-12-2025

https://doi.org/10.6000/1929-6029.2025.14.73