



A New F -Test Applicable to Large-Scale Data

Mohsen Salehi^{1,2}, Adel Mohammadpour^{2,*}, Kerrie Mengersen³

¹Department of Statistics, University of Qom, Qom, Iran

²Department of Statistics, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran

³Science and Engineering Faculty, Queensland University of Technology, Brisbane, Australia

ARTICLE INFO

Article History

Received 25 May 2017

Accepted 06 Jun 2018

Keywords

Microarray data

Permutation test

Null statistic

F -test

ABSTRACT

In large-scale multiple testing, the permutation test based on making a null statistic has been widely employed in the literature. Because it enables us to use the null permuted samples and estimate the p -value more accurately. Some test statistics, which can be modified to a null statistic, have been proposed for two independent groups. In this paper, we propose a new statistic and corresponding F -test, which can be applied for three or more groups. The simulation results demonstrate that the proposed method has a smaller interquartile range (IQR) than the usual permutation method and estimates p -values more accurately than its rival.

© 2019 The Authors. Published by Atlantis Press SARL.

This is an open access article distributed under the CC BY-NC 4.0 license (<http://creativecommons.org/licenses/by-nc/4.0/>).

1. INTRODUCTION

In some experimental scientific fields such as microarray and deep sequencing (or next generation sequencing) methods in biology, the numbers of features (probe sets) is huge while the number of samples is limited to perhaps only two or three. Nearly all features are numeric, showing the values of gene expressions between control and other groups, and finding differentially expressed genes between those groups is of great interest. Such data refer to large-scale data.

Traditional approaches to multiple testing under a null hypothesis, the equality of means in different groups, can lead to inaccurate decisions, particularly in the context of many tests and small sample sizes. These poor decisions can arise for a number of reasons, including the detection of spurious significant differences, asymmetry in the distribution of alternatives [1], and failure of assumptions about the normality of the underlying random variables. In this situation, employing the permutation method to estimate the p -values using the permutation samples has been recommended [2–11]. In addition, in some multiple hypothesis testing, the distribution of alternatives is not symmetric, and we need some alternative methods in this situation, see [1].

However, as the sample size is small, the number of permutation samples is reduced and the estimation of an appropriate p -value seems unreachable. In addition, using the permutation samples, the samples in two types of groups, which are measured under two conditions (i.e., time point), are being mixed together. Then, multiple testing under the null hypothesis is done, which we know that the null hypothesis is not satisfied for many of them. In fact, the permuted samples may not permit resampling from the null distribution of the test statistic, which is called the null permuted sample. Therefore, the usual permutation method cannot be used to estimate the p -value directly.

To overcome this obstacle one can use the null statistic. The null statistic is a modified test statistic from the initial test statistic with the same distribution. Some methods based on deriving a null statistic have been constructed for the analysis of two groups, e.g., see [4,12]. However, in many cases, interest may center on three or more groups. For instance, in microarray analysis introduced by [13–15], three groups were used to describe the expression value of many genes.

Some researchers have proposed three-stage filtrations, which reduce the number of comparisons and then use one-way analysis of variance (ANOVA) to identify the differentially expressed genes on filtered genes or features, see [13], and also [16,17]. Applying one-way ANOVA and calculating the p -values and subjecting them to a new approach, namely a Beta Contamination model, has also been reported; see [16]. In brief, in [13,16] the p -values estimated by a traditional F -test are used as the starting point of a subsequent statistical analysis, whereas in [17] ANOVA is employed to identify differentially expressed genes from next generation sequencing and microarray in transcriptome

*Corresponding author. Email: adel@aut.ac.ir

profiling of activated T cells. In addition, some widely used tools in these fields like RMAexpress, which compute the summary of values of gene expression, using the obtained p -values from ANOVA for screening huge numbers of differentially expressed genes, see [18–20]. This literature motivates the development of a new test statistic to facilitate estimation of more accurate p -values in multiple hypothesis testing with three or more groups.

We organize the paper as follows: In Section 2, we review the Fisher test, the usual permutation test for the one-way ANOVA, and the estimation of the p -value using the null statistic. We propose a new test statistic, which is modified to a null statistic in Section 3. To investigate the robustness of the proposed test statistic and accuracy of the estimated p -values, some simulation studies are presented in Section 4. In this section, two real data sets are also considered. Finally, the conclusions are given in Section 5.

2. F-TEST AND PERMUTATION SOLUTION

When we have three independent groups or more and we are going to test the null hypothesis, equality of means, versus the alternative hypothesis, different means, we usually use the well-known one-way ANOVA method, under the assumption of normally distributed, independent random samples in each group. When the normality assumption is not satisfied, we can use the permutation method instead. In this section, we review these methods briefly. In addition, we present a method for estimating the p -value based on the permutation method in large-scale data.

2.1. Multiple Testing Problems

In this paper, we deal with multiple hypothesis testing with k independent groups. For ease of exposition, we consider a single test. Let X_1, X_2, \dots, X_k be the underlying random variables in k groups, and $\mu_1, \mu_2, \dots, \mu_k$ be the corresponding means, respectively with common variance σ^2 . Let $X_i = (X_{i1}, X_{i2}, \dots, X_{in_i})$, $i = 1, 2, \dots, k$ is the vector of random samples from X_i 's. Take $Z = (X_1, X_2, \dots, X_k)$. We are going to test the following hypotheses:

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \dots = \mu_k \\ H_1 : \mu_i \neq \mu_j, \quad \exists i \neq j. \end{cases} \tag{1}$$

A common test statistic, which is used for testing the hypotheses (1) is the Fisher test, [21], as follows:

$$F(Z) = \frac{\sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2 / (k - 1)}{\sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 / (N - k)}, \tag{2}$$

where

$$\bar{X}_i = \frac{\sum_{j=1}^{n_i} X_{ij}}{n_i}, \quad \bar{X} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}}{N},$$

and $N = \sum_{i=1}^k n_i$. If we consider the normality assumption for the underlying variables in each group, the distribution of F in (2), under the null hypothesis will be F distribution with $k - 1$ and $N - k$ degrees of freedom.

In the usual permutation approach described by [22], one constructs (2) by randomly permuting

$$Z = (X_1, X_2, \dots, X_k) = (X_{11}, \dots, X_{1n_1}, X_{21}, \dots, X_{2n_2}, \dots, X_{k1}, \dots, X_{kn_k}).$$

Denote the permuted elements of Z by Z^* and let X_{ij}^* denote the j th permutation sample in the i th group. Then, X_1^* and X_k^* are the first n_1 and the last n_k random samples of Z^* . Clearly, the F statistic with permutation samples denoted as $F^* = F(Z^*)$, has also the F distribution.

2.2. Estimated p -Value

Suppose that we decide to reject the null hypothesis H_0 when the test statistic $F(Z)$ is sufficiently large. Therefore, the p -value can be defined by

$$p = Pr(F(Z) \geq f | H_0),$$

where f is the observed value of $F(\mathbf{Z})$. If we consider the B permutation samples and denote \mathbf{Z}_b^* as a b th permutation sample of \mathbf{Z} , then the estimator of the p -value can be defined as follows:

$$\hat{p} = \frac{1}{B} \sum_{b=1}^B I(F(\mathbf{Z}_b^*) \geq f), \tag{3}$$

where $I(D)$ denotes the indicator function: $I(D)$ is one if D is satisfied, and is zero if is not.

In large-scale data like microarray data, we can use the information of other genes to estimate the p -value for a special gene with higher accuracy. A good alternative for the p -value estimator is defined by

$$\hat{p}(F(\mathbf{Z}^*); f_g) = \frac{1}{BG} \sum_{b=1}^B \sum_{g'=1}^G I(F(\mathbf{Z}_{bt}^*) \geq f_g), \tag{4}$$

where \mathbf{Z}_{bt}^* and f_g are the b th permutation sample at the t th gene and the observed value of the test statistic of the g th gene, respectively.

(3) and (4) involve the use of information about a lot of genes, and it is unknown if the null hypothesis holds for them. [11] and [4] suggest the usage of the null statistic instead of the main test statistic. The null statistic, $F^{\text{null}}(\mathbf{Z})$, is a test statistic with the same distribution as the initial test statistic, $F(\mathbf{Z})$, such that

$$p(F(\mathbf{Z}); f) = Pr(F(\mathbf{Z}) \geq f | H_0) = Pr(F^{\text{null}}(\mathbf{Z}) \geq f). \tag{5}$$

Therefore, one can estimate the p -value more accurately as follows:

$$\hat{p}(F^{\text{null}}(\mathbf{Z}^*); f_g) = \frac{1}{BG} \sum_{b=1}^B \sum_{g'=1}^G I(F^{\text{null}}(\mathbf{Z}_{bt}^*) \geq f_g). \tag{6}$$

3. PROPOSED METHOD

We propose a new permutation test statistic for hypothesis testing (1), such that it can be modified to a null statistic. The p -values are estimated by (6).

Constructing the test statistic and its null statistic: Consider $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{in_i})$ for $i = 1, 2, \dots, k$. We divide the sample size of the i th group into two partitions, with n_{i1} and n_{i2} elements,

$$\mathbf{X}_{i(1)} = (X_{i1}, X_{i2}, \dots, X_{in_i}), \quad \mathbf{X}_{i(2)} = (X_{i(n_i+1)}, X_{i(n_i+2)}, \dots, X_{in_i}), \tag{7}$$

such that, $n_i = n_{i1} + n_{i2}$. Take

$$\bar{X}_{i(1)} = \frac{\sum_{j=1}^{n_{i1}} X_{ij}}{n_{i1}}, \quad \bar{X}_{i(2)} = \frac{\sum_{j=n_i+1}^{n_i} X_{ij}}{n_{i2}}.$$

We want to test the hypotheses (1).

Theorem 3.1. *Under the normality assumption for X_i , where $i = 1, 2, \dots, k$, and under the null hypothesis, the test statistic for testing (1) given by*

$$F_s(\mathbf{Z}) = \frac{\sum_{i=1}^k n'_i \left(\bar{X}_{i(1)} + \bar{X}_{i(2)} - \frac{\sum_{i=1}^k n'_i (\bar{X}_{i(1)} + \bar{X}_{i(2)})}{\sum_{i=1}^k n'_i} \right)^2}{\sum_{i=1}^k \sum_{l=1}^2 (n_{il} - 1) S_{i(l)}^2 / (N - 2k)}, \tag{8}$$

where

$$\frac{1}{n'_i} = \frac{1}{n_{i1}} + \frac{1}{n_{i2}}, \quad S_{i(l)}^2 = \frac{1}{n_{il} - 1} \sum_{j=1}^{n_{il}} (X_{ij} - \bar{X}_{i(l)})^2$$

has an F distribution with $k - 1$ and $N - 2k$ degrees of freedom, respectively.

Proof. Let $k = 2$, $\bar{U}_1 = \bar{X}_{1(1)} + \bar{X}_{1(2)}$ and $\bar{U}_2 = \bar{X}_{2(1)} + \bar{X}_{2(2)}$. Assuming a normal distribution for X_1 and X_2 , the distribution of $\bar{U}_1 - \bar{U}_2$ is given by

$$\bar{U}_1 - \bar{U}_2 \sim N\left(2\mu_1 - 2\mu_2, \left(\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}\right)\sigma^2\right).$$

Since $X_i, i = 1, 2$, are independent, then $X_{i(1)}, X_{i(2)}$ are also independent. On the other hand, if we define S_i^2 as the variance of the random sample in the i th group, we know that $\frac{(n_i-1)S_i^2}{\sigma^2}$ has a Chi-squared distribution with $(n_i - 1)$ degrees of freedom. Then, $\frac{(n_{il}-1)S_{i(l)}^2}{\sigma^2}$, for $i = 1, 2$, and $l = 1, 2$, are also independent and have a Chi-square distribution with the $(n_{il} - 1)$ degrees of freedom. Let $S^{*2} = \sum_{i=1}^2 \sum_{l=1}^2 (n_{il} - 1) S_{i(l)}^2$, then $\frac{(N-4)S^{*2}}{\sigma^2}$ has a Chi-squared distribution with $N - 4$ degrees of freedom. Thus, the distribution of $\frac{\bar{U}_1 - \bar{U}_2}{\sigma\sqrt{\frac{1}{n'_1} + \frac{1}{n'_2}}}$ is the standard normal distribution under the

null hypothesis $\mu_1 = \mu_2$. We can rewrite $\frac{\bar{U}_1 - \bar{U}_2}{\sqrt{\frac{1}{n'_1} + \frac{1}{n'_2}}}$ as follows:

$$\begin{aligned} \left(\frac{\bar{U}_1 - \bar{U}_2}{\sqrt{\frac{1}{n'_1} + \frac{1}{n'_2}}}\right)^2 &= \left(\frac{\bar{U}_1 - \bar{U}_2}{\sqrt{\frac{N'}{n'_1 n'_2}}}\right)^2, \\ &= n'_2 \left(\frac{n'_1}{N'} (\bar{U}_1 - \bar{U}_2)\right)^2 + n'_1 \left(\frac{n'_2}{N'} (\bar{U}_1 - \bar{U}_2)\right)^2, \\ &= n'_2 \left(-\bar{U}_2 + \frac{n'_1 \bar{U}_1 + n'_2 \bar{U}_2}{N'}\right)^2 + n'_1 \left(\bar{U}_1 - \frac{n'_1 \bar{U}_1 + n'_2 \bar{U}_2}{N'}\right)^2, \end{aligned}$$

where $N' = n'_1 + n'_2$. Suppose that $\bar{\bar{U}} = \frac{n'_1 \bar{U}_1 + n'_2 \bar{U}_2}{N'}$, then

$$\left(\frac{\bar{U}_1 - \bar{U}_2}{\sqrt{\frac{1}{n'_1} + \frac{1}{n'_2}}}\right)^2 = \sum_{i=1}^2 n'_i (\bar{U}_i - \bar{\bar{U}})^2. \tag{9}$$

It can be shown that $\bar{U}_1 - \bar{U}_2$ and S^{*2} are independent [4], so (9) is independent of S^{*2} . Thus, $\frac{\bar{U}_1 - \bar{U}_2}{S^* \sqrt{\frac{1}{n'_1} + \frac{1}{n'_2}}}$ has a t -distribution with $N - 4$ degrees of freedom. Therefore,

$$\frac{\sum_{i=1}^2 n'_i (\bar{U}_i - \bar{\bar{U}})^2}{S^{*2} / (N - 4)},$$

has a Fisher distribution with $2 - 1$ and $N - 2 \times 2$ degrees of freedom. In fact, \bar{U}_i and $\bar{\bar{U}}$ are the weighted means of the i th group and total sample, respectively. Therefore,

$$F_s(\mathbf{Z}) = \frac{\sum_{i=1}^k n'_i \left(\bar{X}_{i(1)} + \bar{X}_{i(2)} - \frac{\sum_{i=1}^k n'_i (\bar{X}_{i(1)} + \bar{X}_{i(2)})}{\sum_{i=1}^k n'_i}\right)^2 / (k - 1)}{\left(\sum_{i=1}^k \sum_{l=1}^2 (n_{il} - 1) S_{i(l)}^2\right) / (N - 2k)},$$

has a Fisher distribution with $k - 1$ and $N - 2k$ degrees of freedom. □

Remark 3.1. It follows, from Theorem 3.1, that for testing the null hypothesis in (1), the test statistic F_m is of level α . The corresponding null statistic is given by

$$F_s^{null}(\mathbf{Z}) = \frac{\sum_{i=1}^k n'_i \left(\bar{X}_{i(1)} - \bar{X}_{i(2)} - \frac{\sum_{i=1}^k n'_i (\bar{X}_{i(1)} - \bar{X}_{i(2)})}{\sum_{i=1}^k n'_i}\right)^2 / (k - 1)}{\left(\sum_{i=1}^k \sum_{l=1}^2 (n_{il} - 1) S_{i(l)}^2\right) / (N - 2k)}. \tag{10}$$

The critical value, called c , is determined such that $Pr(F_s(\mathbf{Z}) \geq c | H_0 = \alpha)$.

Theorem 3.2. F_s and F_s^{null} statistics satisfy the following condition:

$$P(F_s(\mathbf{Z}) > f_s | H_0) = P(F_s^{null}(\mathbf{Z}) > f_s), f_s \in I R.$$

Proof. Since the denominator of $F_s(\mathbf{Z})$ and $F_s^{null}(\mathbf{Z})$ is the same, it is enough to show that the distribution of $T_1 = \bar{X}_{i(1)} + \bar{X}_{i(2)} - \frac{\sum_{i=1}^k n'_i (\bar{X}_{i(1)} + \bar{X}_{i(2)})}{\sum_{i=1}^k n'_i}$ under the null hypothesis is the same as for distribution of $T_2 = \bar{X}_{i(1)} - \bar{X}_{i(2)} - \frac{\sum_{i=1}^k n'_i (\bar{X}_{i(1)} - \bar{X}_{i(2)})}{\sum_{i=1}^k n'_i}$. It is evident that the distributions of $\bar{X}_{i(1)} + \bar{X}_{i(2)}$ and $\frac{\sum_{i=1}^k n'_i (\bar{X}_{i(1)} + \bar{X}_{i(2)})}{\sum_{i=1}^k n'_i}$ follow normal distributions with means $2\mu_i$, $2 \sum_{i=1}^k \frac{n'_i \mu_i}{N'}$ and variances $\frac{\sigma^2}{n'_i}$ and $\frac{\sigma^2}{\sum_{j=1}^k n'_j}$, respectively. Therefore, under the null hypothesis, T_1 follows a zero-mean normal distribution with variance $\sigma^2 \left(\frac{1}{n'_i} + \frac{1}{\sum_{j=1}^k n'_j} \right)$. Similarly, $\bar{X}_{i(1)} - \bar{X}_{i(2)}$ and $\frac{\sum_{i=1}^k n'_i (\bar{X}_{i(1)} - \bar{X}_{i(2)})}{\sum_{i=1}^k n'_i}$ have zero-mean normal distributions with variances $\frac{\sigma^2}{n'_i}$ and $\frac{\sigma^2}{\sum_{j=1}^k n'_j}$, respectively. This implies that both T_1 and T_2 have the same distribution. \square

Remark 3.2. To estimate the p -value in each test, each group is divided into two partitions as (9) and computed the observed value of (8), f_s . Then, the samples in all groups are permuted in all tests separately and randomly; and computed (10) for all of them. The p -value can be estimated by (6).

4. RESULTS

This section is organized into two parts. In the first subsection, the robustness of the proposed test statistic, $F_s(\mathbf{Z})$ and accuracy of the estimated p -value are inspected through some simulation studies. In the second subsection, we apply the proposed method and the usual permutation method to two real data sets.

4.1. Simulated Data

Robustness of the new test statistic: The condition (5) is convinced under the normality condition by Theorem 3.2. Considering some candidates for the underlying distribution of the random sample, we investigate if the proposed test statistic, $F_s(\mathbf{Z})$ and the corresponding null statistic, $F_s^{null}(\mathbf{Z})$ are robust with respect to the departure from normality. In fact, the condition (5) is examined under the violation of the normality assumption.

Without loss of generality, take $k = 3$ and assume three candidates for the underlying distribution: t -Student, Laplace, and Logistic. Two cases are considered for the sample size for each group, $(n_1, n_2, n_3) = (4, 4, 4)$ and $(30, 30, 30)$. We then generate the random samples from the candidate distributions and obtain the observed values of $F_s(\mathbf{Z})$ and $F_s^{null}(\mathbf{Z})$ for their corresponding parameters. The number of iterations is taken to be 10 000. The outcomes are analyzed using the Anderson—Darling (AD) test. Table 1 presents the results. For the first candidate, we suppose that the distribution of the random variables in each group is t -Student with the different degrees of freedoms, 4, 10, 15, 25. The obtained p -values indicate that both in small and large sample size, the condition (5) is satisfied. For the second candidate, it is assumed that the underlying variables follow a Logistic distribution. The location and scale parameters are taken to be as $\mu = 0$ and $\sigma = 1$, respectively. Since the proposed test statistic is invariant in location, we only consider $\mu = 0$. The obtained p -value verifies that the condition (5) is met. The same conclusion is drawn for the final candidate with the Laplace distribution for the underlying variables.

P-value behavior: Since we want to implement the multiple testing, the random samples in each group are simulated for 1000 tests in two categories. In the first category, 90 percent of them are simulated under the null hypothesis, and in the second category, the remaining

Table 1 | The estimated p -value by AD-test.

Candidate Distribution	Family Parameter	Sample Sizes	
		$(n_1, n_2, n_3) = (4, 4, 4)$	$(n_1, n_2, n_3) = (30, 30, 30)$
t_ν	4	0.723	0.952
	10	0.281	0.569
	15	0.992	0.872
	25	0.792	0.958
Logistic (μ, σ)	(0, 1)	0.204	0.569
Laplace (μ, σ)	(0, 1)	0.339	0.640

AD, Anderson—Darling; The null hypothesis is that two samples come from the common distribution.

10 percent are simulated under the alternative hypothesis. For this purpose, the random samples are generated from $N(0, 1)$ in the first category, and generated from $N(\mu_1, 1)$, $N(\mu_2, 1)$ and $N(\mu_3, 1)$ in the second category, such that $\mu_i, i = 1, 2, 3$ are generated from a normal distribution with zero-mean and variance 4, respectively. In each category, we consider three sample sizes: $(n_1, n_2, n_3) = (6, 6, 6)$, $(8, 8, 8)$, and $(10, 10, 10)$.

Let f_s^* and f^* be the observed values of $F_s^{null}(\mathbf{Z}^*)$ and $F(\mathbf{Z}^*)$, respectively when we permute \mathbf{Z} . We draw 1000, 2000, 4000 permuted samples from $\mathbf{Z} = (X_1, X_2, X_3)$ for the different sample sizes $(n_1, n_2, n_3) = (6, 6, 6)$, $(8, 8, 8)$, and $(10, 10, 10)$, respectively. Then; we compute f_s^* and f^* for all permuted samples. In the Usual Permutation Method, called UPM, the p -values are computed through the method described by [22]. In the Proposed Permutation Method, called PPM, they are computed through $\hat{p}(F_s^{null}(\mathbf{Z}^*); f_\alpha)$.

Remark 4.1. When we consider no effect in the groups, the labels assigned subject to the groups are interchangeable. Therefore, for each permuted sample, the homogeneity of variance must be tested.

Considering the above remark, when we permute a sample, randomly, and divide it into two partitions, some of the obtained permutation samples will display substantive heterogeneity of variances. These so-called extraneous permutation samples (EPSs) should be removed before computing p -values. Figure 1 shows a rate of about 4 to 10 percent of EPS in 1000 comparisons under different values of α and (n_1, n_2, n_3) . We investigate three cases for the nominal value $\alpha = 0.05, 0.1, 0.2$. Figure 2 depicts boxplots of the estimated p -values. In each subfigure, the left and right boxplots present the distribution of the estimated p -values in PPM and UPM, respectively.

We can see that the median of the estimated p -values through PPM is closer to the nominal significance levels α than UPM. Moreover, Figure 2 clearly shows that UPM has a considerably higher dispersion than PPM. For more clarity, Table 2 shows the computed interquartile range (IQR) for each subfigure of Figure 1. We can see that the computed IQR for PPM, both with EPS and without EPS, is smaller than for UPM. This simulation study also considered the case in which the distribution is skew-normal with the location, scale, and shape parameters zero, one, and two, respectively. The results are depicted in Figure 3 in this situation.

Heterogeneity of the variance effect on the condition (5): Here, we address an important goal, namely the investigation of the effect of heterogeneity of variance on condition (5). We assume normal distributions for the underlying variables with zero means and the different values in the variances, $\sigma = 0.5, 1, \text{ and } 3$. The results are shown in Table 3. We can see that the obtained p -values through the AD-test are sufficiently large to satisfy the condition (5).

4.2. Real Data Sets

We have applied our method to two types of microarray data. The first study [13] is aimed to investigate the effectiveness of the processes of brain aging in increasing of Alzheimer's disease. Gene expression values were measured for 30 male rats equally allocated to three groups: aged, mid-aged, and young. For technical reasons, one chip in the young group was lost. We used the updated database since 2014, comprising the measured values of 8799 probe sets on each of the 29 rats and employed the RMAexpress software to calculate the normalized data. Table 4 shows the results. The number of genes identified as differentially expressed is 1081 under a one-way ANOVA, compared with 1098 and 948 obtained using UPM and PPM, respectively.

The second example, aimed to identify the reasons for preconceptional endometrial deregulations within in vitro fertilization (IVF) and intracytoplasmic sperm injection (ICSI) [23]. The study investigated whether there is any relation between preconceptional endometrial deregulations and the two events, implantation failures (IFs) and recurrent miscarriages (MSs), when IVF/ICSI is utilized. Three groups are thus considered: a fertile control group comprising women who have had at least one birth, and two test groups comprising women with IF and IM, respectively. Each group contained five subjects. In total, the gene expression values of 22 482 genes related to 54 675 probe sets were measured. The data were normalized by RMA software. The results of UPM and PPM analysis of this data set are presented in Table 5. Based on a nominal value $\alpha = 0.01$, the RAM-seq software detected 4894 genes with differential expression, compared with 1183 and 321 genes detected by UPM and PPM, respectively.

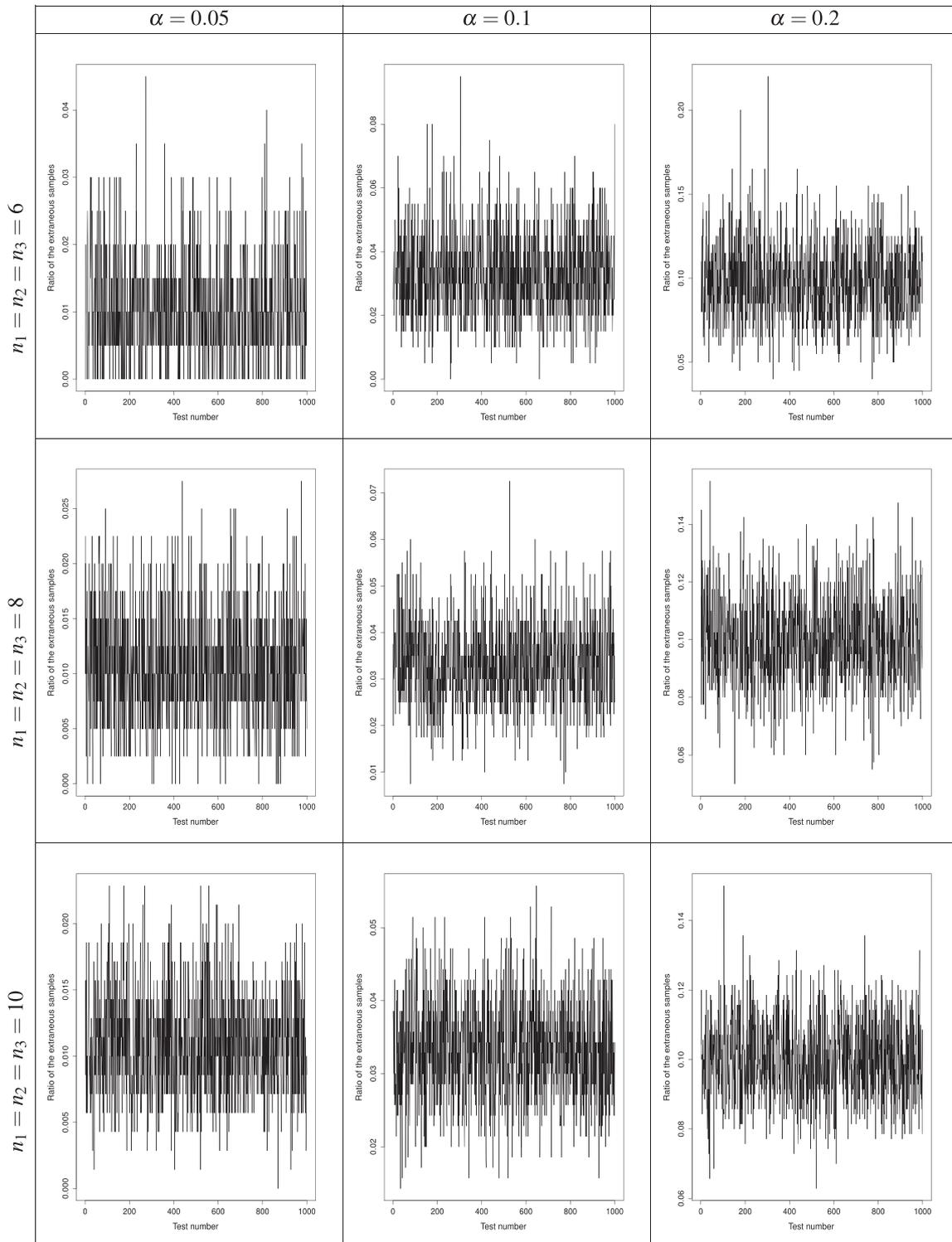


Figure 1 | The ratio of extraneous permutation sample (EPS) from the total number of permuted samples in 1000 multiple testing. Each subfigure shows this ratio for the different values of and the sample sizes.

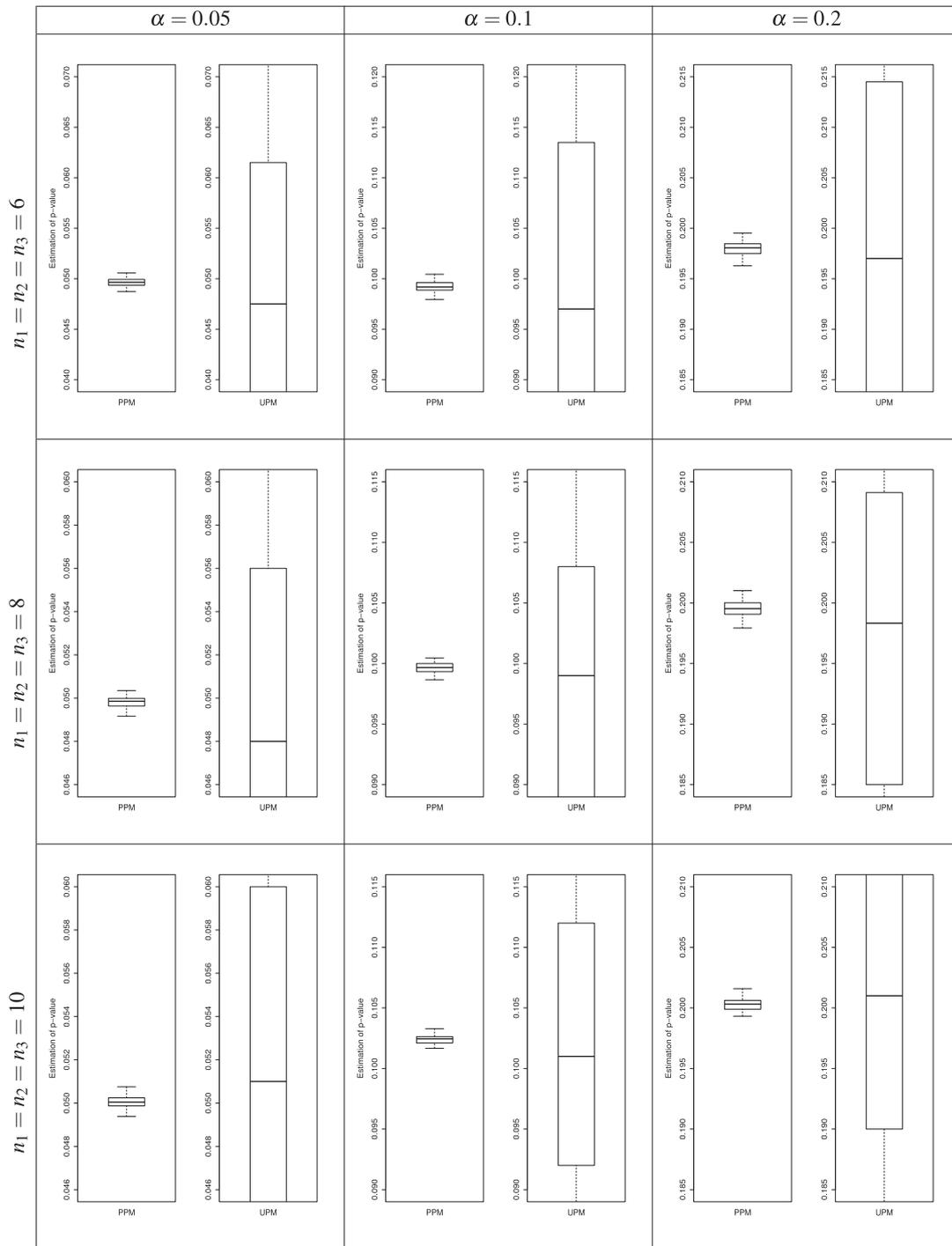


Figure 2 | Boxplots of the estimated p -values by UPM and PPM in three cases $(n_1; n_2; n_3)$: (6,6,6), (8,8,8), (10,10,10). The random samples are generated from a normal distribution. UPM and PPM in each sub grid correspond to the usual permutation method and $\hat{p}(F_s^{null}(Z^*); f_\alpha)$.

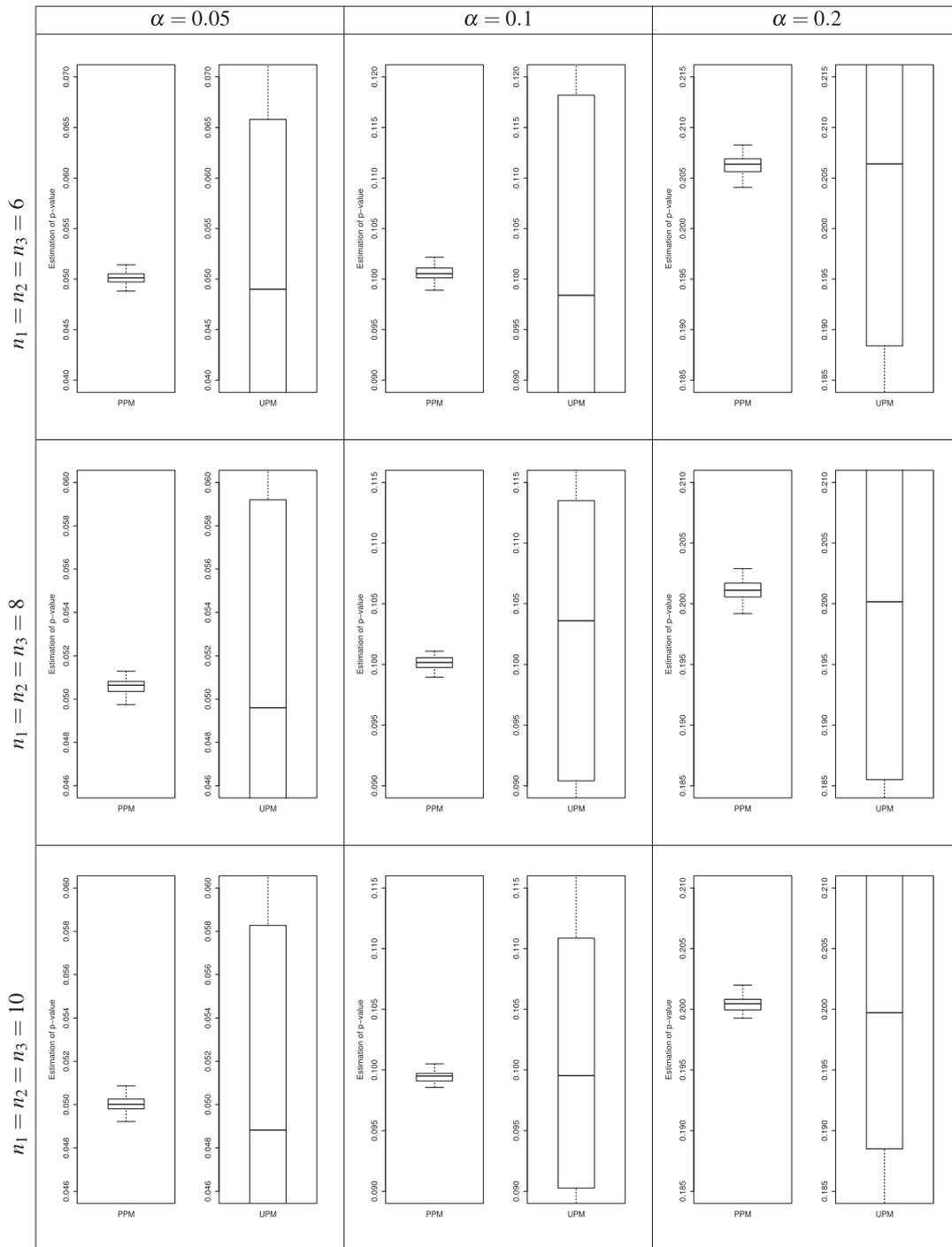


Figure 3 | Boxplots of the estimated p-values by PPM and UPM in three cases $(n_1; n_2; n_3)$: (6,6,6), (8,8,8), (10,10,10). The random samples are generated from the skew-normal distribution. PPM and UPM in each subfigure correspond to the usual permutation method and $\hat{p}(F_s^{null}(Z^*); f_\alpha)$.

Table 2 | Comparing the computed IQR in UPM and PPM (with and without EPS).

Method	Sample Sizes (n_1, n_2, n_3)								
	(6,6,6)			(8,8,8)			(10,10,10)		
	α			α			α		
	0.05	0.1	0.2	0.05	0.1	0.2	0.05	0.1	0.2
UPM	0.0420	0.063	0.0917	0.0350	0.0535	0.0730	0.0302	0.0490	0.0700
PPM with EPS	0.0006	0.0007	0.0011	0.0003	0.0010	0.0007	0.0004	0.0005	0.0005
PPM without EPS	0.0005	0.0007	0.0009	0.0003	0.0006	0.0009	0.0003	0.0005	0.0007

IQR, interquartile range; UPM, Usual Permutation Method; PPM, Proposed Permutation Method; EPS, extraneous permutation sample.

Table 3 | The estimated p -value by AD-test, under heterogeneity of variances.

Sample Sizes (n_1, n_2, n_3)	$N(\mu, \sigma)$	p -value
(4, 4, 4)	(0, 0.5), (0, 1), (0, 3)	0.694
(30, 30, 30)	(0, 0.5), (0, 1), (0, 3)	0.549

AD, Anderson–Darling; The null hypothesis is that two samples come from the common distribution.

Table 4 | The number of detected genes as differential gene expression through three methods: one-way ANOVA, UPM, and PPM for the male Fischer rats.

Method	The Number of Detected Genes
One–way ANOVA	1081
UPM	1098
PPM	948

ANOVA, analysis of variance; UPM, Usual Permutation Method; PPM, Proposed Permutation Method.

Table 5 | The number of detected genes as differential gene expression through three methods: one-way ANOVA, UPM, and PPM for the preconceptional endometrial deregulations.

Method	The Number of Detected Genes
One–way ANOVA	4894
UPM	1183
PPM	321

ANOVA, analysis of variance; UPM, Usual Permutation Method; PPM, Proposed Permutation Method.

5. CONCLUSIONS

In this paper, we first introduced a new permutation F -test to test the null hypothesis of equal means, for multiple testing with three or more groups. We then modified this test statistic to a null statistic which estimates the corresponding p -values more accurately. We presented several scenarios to evaluate the performance of our procedure for various α and the sample sizes. In the case studies, we confined our attention to three independent groups, although the simulation results clearly illustrate that our methodology performs well under a much wider set of scenarios. The new method is shown to produce p -values with the median much closer to the true α and with a smaller IQR, compared with the usual permutation method.

CONFLICT OF INTEREST

The authors declare they have no conflicts of interest.

ACKNOWLEDGMENTS

The comments of an associate editor and reviewers improved the manuscript significantly. We wish to express our appreciation to them. Thanks are due to our colleague Professor M. Ebrahimi, who assisted in extracting the real data sets. We are also thankful to Dr. S. K. Ghoreishi for his comments.

REFERENCES

1. N. Bansal, G. Hamedani, M. Maadooliat, *Biometrics*. 72 (2016), 494–502.
2. B. Efron, R. Tibshirani, J. Storey, *J. Am. Stat. Assoc.* 96 (2001), 1151–1160.
3. X. Gao, *Bioinformatics*. 22 (2006), 1486–1494.
4. W. Pan, *Bioinformatics*. 19 (2003), 1333–1340.
5. W. Pan, J. Lin, C. Le, *Funct. Integr. Genomics*. 3 (2003), 117–124.
6. S. Scheid, R. Spang, in: A. Apostolico, C. Guerra, S. Istrail, P.A. Pevzner, M. Waterman (Eds.), *Research in Computational Molecular Biology*, Springer, Berlin, 2006, pp. 338–347.
7. L. Southworth, S. Kim, A. Owen, *J. Comput. Biol.* 16 (2009), 625–638.
8. J. Storey, R. Tibshirani, *Proc. Natl. Acad. Sci.* 100 (2003), 9440–9445.
9. V. Tusher, R. Tibshirani, G. Chu, *Proc. Natl. Acad. Sci.* 98 (2001), 5116–5121.
10. J. Xu, X. Cui, *Bioinformatics*. 24 (2008), 1056–1062.
11. Y. Zhao, W. Pan, *Bioinformatics*. 19 (2003), 1046–1054.
12. H. Fujisawa, T. Sakaguchi, *TEST Official J. Span. Soc. Stat. Oper. Res.* 21 (2012), 280–300.
13. E. Blalock, K. Chen, K. Sharrow, J. Herman, N. Porter, T. Foster, P. Landfield, *J. Neurosci.* 23 (2003), 3807–3819.
14. E. Sweeney, C. Crainiceanu, J. Gertheiss, *Stat. Appl. Genet. Mol. Biol.* 15 (2016), 213–235.
15. L. Tan, C. Tan, Z. Rahman, H. Goh, I. Ismail, Z. Zainal, *Data Brief*. 14 (2017), 267–271.
16. H. Dai, R. Charnigo, *J. Appl. Stat.* 35 (2008), 31–47.
17. S. Zhao, W. Fung-Leung, A. Bittner, K. Ngo, X. Liu, *PLoS One*. 9 (2014), e78644.
18. B. Boldstad, R. Irizarry, M. Astrand, T. Speed, *Bioinformatics*. 19 (2003), 185–193.
19. R. Irizarry, B. Bolstad, F. Collin, L. Cope, B. Hobbs, T. Speed, *Nucleic Acids Res.* 31 (2003), e15.
20. R. Irizarry, B. Hobbs, F. Collin, Y. Beazer-Barclay, K. Antonellis, U. Scherf, T. Speed, *Biostatistics*. 4 (2003), 249–264.
21. E. Lehmann, J. Romano, *Testing Statistical Hypotheses*, Springer Science & Business Media, New York, 2006.
22. F. Pesarin, L. Salmaso, *Permutation Tests for Complex Data: Theory, Applications and Software*, John Wiley & Sons, New York, 2010.
23. N. Lédée, C. Munaut, J. Aubert, V. Sérazin, M. Rahmati, G. Chaouat, O. Sandra, *J. Pathol.* 225 (2011), 554–564.