Empirical permutation paradigm for two-way repeated measures ANOVA

Justice Ighodaro Odiase

**Department of Mathematics, University of Benin,
Benin City, Nigeria**

Abstract

Exact test of significance can only be guaranteed by the use of the distribution of a test statistic resulting from an exhaustive enumeration of all the distinct permutations of the observations in an experiment, especially when threshold p-values are involved. The permutation paradigm requires no distributional assumptions and works well with values that are normal, almost normal and non-normally distributed. Under the truth of the null hypothesis, permutation test only requires exchangeability of observations either within pairs, between samples, or within blocks. This paper examines empirically the permutation distribution of the observations in an experiment in the context of exchangeability within blocks as applicable to two-way repeated measures analysis of variance. The methodology developed in this paper enumerates all the distinct permutations of the observations or ranks of observations in an experiment and is illustratively applied to the Friedman test statistic. The Friedman test is used to detect differences in treatments across multiple test attempts and only requires exchangeability (or, if variances differ, compound symmetry) and the ability to rank the data. The exact distribution of the Friedman test statistic for small samples is therefore generated empirically, leading to the production of exact critical values at different levels of significance.

Keywords: Permutation test, Friedman test, p-value, two-way ANOVA, exchangeability
AMS Classification: 62E15, 62G10

1.0   Introduction

A major problem of statistical inference is how to obtain an exact test of significance when the form of the underlying probability distribution is unknown. The idea of a general method of dealing with this problem of obtaining an exact test of significance originated from Fisher [1]. Descriptions of the permutation test for the difference between means for a two-sample problem were first presented by Fisher [1]. The essential feature of the permutation method is that all the distinct permutations of the observations are considered with each permutation equally likely under the hypothesis to be tested. An exact test at the level of significance α is constructed by choosing a proportion, α, of the permutation space as the critical region. Scheffe [2] demonstrates that the permutation approach is the only possible method of truly constructing exact tests of significance.

Permutation tests have received attention under the guise of bootstrap [3], Bayesian and likelihood [4, 5], etc. All these approaches for obtaining the probability distribution of a test statistic have gained prominence due to the fact that they are computationally less demanding than the permutation approach. This was a major concern of Fisher [6] that *"the statistician does not carry out this very simple and very tedious process, but his conclusions have no justification beyond the fact that they agree with those which could have been arrived at by this elementary method"*. Ernst [7] believes that today, with fast computers, there is little reason for the statistician not to carry out this *"very tedious process"*. In the last three decades, statisticians have tried to evolve the underlying theory of permutation tests due to their wide range of application and elegance, simply put by Good [8] thus: *"Permutation tests permit us to choose the test statistic best suited to the task at hand. This freedom of choice opens up a thousand practical applications, including many which are beyond the reach of conventional parametric statistics. Flexible, robust in the face of missing data and violations of assumptions, the*

Corresponding author: E-mail: justice.odiase@uniben.eduyahoo.com

permutation test is among the most powerful of statistical procedures. Through sample size reduction, permutation tests can reduce the costs of experiments and surveys".

In this paper, consideration is given to the exhaustive permutation of the observations or ranks of the observations in a multi-sample experiment as relevant to two-way repeated measures ANOVA and illustratively applied to the Friedman test statistic. The Friedman test is used to detect differences in treatments across multiple test attempts and only requires exchangeability (or, if variances differ, compound symmetry) and the ability to rank the data.

The motivation for this paper stems from the problems highlighted in the literature [1, 8, 9 – 17], viz.: computational difficulties of permutation test, non-availability of a methodology for an exhaustive permutation enumeration, non-availability of exact statistical tables for some test statistics, and the fact that the probability of a type I error is not exactly α for several entries in statistical tables in common use.

Fisher analyzed Charles Darwin's data of 1878 [18] on the height of cross and self-fertilized Zea mays plants [19], was tested using the paired t-test and the exact permutation test. Interestingly, at 5% level of significance, the null hypothesis of no significant difference is accepted under the exact permutation distribution (p = 0.05267) while the null hypothesis is rejected under the paired t-distribution (p = 0.04970). This observation was also made in [20] and it buttresses the fact that the use of exact p-values is important, especially when threshold p-values are involved. Permutation test can therefore serve as an independent check on the classical methods in common use, as observed by Fisher [6]. It is only by deciding to accept or reject the null hypothesis on the basis of an exact p-value that one is guaranteed to be protected from type I error at the desired significance level. As a practical matter, of course it is not possible to use exact p-values all the time. It is difficult to quantify just how large a data set can be solved by the exact algorithms, because that depends on so many factors other than just the sample size. The type of exact test, the degree of imbalance in the allocation of subjects to treatments, the number of rows and columns in a cross-tabulation, and a variety of other factors interact in complicated ways in determining if a particular data set is amenable to exact inference.

2.0 Exhaustive Permutation Procedures

The process of obtaining the permutations starts by choosing the test statistic T and the acceptable significance level α. Let π_1, π_2, …, π_N be a set of all distinct permutations of the observations or ranks of the observations in the experiment.

Compute the test statistic T_i for permutation π_i, that is, $T_i = T(\pi_i)$. Construct an empirical cumulative distribution for T.

$$p_0 = p(T \leq T_i) = \frac{1}{N}\sum_{i=1}^{N}\psi(t_0 - T_i) \text{, where } \psi(\cdot) = \begin{cases} 1, & \text{if } t_0 \geq T_i \\ 0, & \text{if } t_0 < T_i \end{cases}.$$

Under the empirical distribution, if $p_0 \leq \alpha$, reject the null hypothesis.

2.1 Paired Permutation

Given two paired samples $X = (x_1, x_2, …, x_n)$ and $Y = (y_1, y_2, …, y_n)$, suppose a sample of n units from the population distribution F_X is paired with a sample of n units from the population distribution F_Y and are simultaneously tested in an experiment with T as the test statistic. For k distinct values of the test statistic T, the probability distribution of the test statistic $T = (T_1, T_2, …, T_k)$ under the null hypothesis $H_0 : F_X = F_Y$ is given by

$$P(T_j = t_0 / H_0) = \sum_{i=1}^{f_j}\left(2^{-n}\right) = f_j\left(2^{-n}\right),$$

where f_j is the number of occurrences of T_j. For specified value of n and the level of significance α, the critical value c corresponds to a level closest to α. Ordering all the distinct occurrences of T in ascending order of magnitude, and if g is the position of the observed value of T, then the following significance level for the left tail of the distribution of the test statistic is

$$\alpha = P(T_g \leq c \mid H_0) = \sum_{j=1}^{g}\sum_{i=1}^{f_j}\left(2^{-n}\right) = \left(2^{-n}\right)\sum_{j=1}^{g}f_j$$

and for the right tail,

$$\alpha = P(T_g \geq c \mid H_0) = \left(2^{-n}\right)\sum_{j=g}^{k}f_j.$$

For a two-tailed test, the left and right tails are summed up. Clearly, when the distribution of the test statistic is symmetric,

$$\sum_{j=1}^{g} f_j = \sum_{j=k-g+1}^{k} f_j$$
.

The distribution of the test statistic is obtained by tabulating the distinct values of the statistic against their probabilities of occurrence in the complete enumeration. A detail description of the implementation of the paired permutation algorithm is found in [20]. Further, Odiase and Ogbonmwan [20] illustrated the implementation of paired permutation algorithm for a balanced two-sample layout of the form $\begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{pmatrix}^T$, where $x_i \in X$ and $y_j \in Y$. In this setting, the number of permutations is $\sum_{i=0}^{n} \binom{n}{i} = 2^n$.

The test statistic is computed for each permutation in the complete enumeration of the distinct permutations. The distribution of the test statistic is obtained by tabulating the distinct values of the statistic against their probabilities of occurrence in the complete enumeration, where all the permutations are equally likely.

Considering consecutive number of pairs for a given experiment, the growth rate of the permutations from $n-1$ pairs to n pairs in a two-sample paired permutation experiment is $\dfrac{2^n}{2^{n-1}} = 2$, which means it doubles each time a single pair of observations is added.

2.2   Independent samples experiment

Given a multisample experiment with $X_i = \left(X_{i1}, X_{i2}, ..., X_{in_i} \right)^T$, $i = 1(1)p$ and $X_N = \left(X_1, X_2, ..., X_p \right)$. Under the null hypothesis, X_N is composed of $N = \sum_{i=1}^{p} n_i$ independent and identically distributed random variables. An exhaustive permutation of the observations yields $\dfrac{N!}{\prod_{i=1}^{P} \left[(n_i)! \right]}$ permutations of the N variates of p subsets of size n_i, $i = 1(1)p$ which are equally likely, each having the conditional probability $\left(\dfrac{N!}{\prod_{i=1}^{P} \left[(n_i)! \right]} \right)^{-1}$. The number of permutations for a two-sample experiment is $\sum_{i=0}^{n} \binom{n_1}{i}\binom{n_2}{i}$, $n = \min(n_1, n_2)$ [16]. After obtaining the permutations of a two sample experiment, find the number of ways to permute any n_3 elements of the combined $(n_1 + n_2 + n_3)$ variates of the three treatments. This yields

$$\binom{n_1 + n_2 + n_3}{n_3} \sum_{i=0}^{n} \binom{n_1}{i}\binom{n_2}{i} = \binom{\sum_{r=1}^{3} n_r}{n_3} \sum_{i=0}^{n} \binom{n_1}{i}\binom{n_2}{i}$$

By following the same procedure as done for the case of three treatments, a complete enumeration of the distinct permutations for a four-treatment experiment yields

$$\binom{\sum_{r=1}^{4} n_r}{n_4}\binom{\sum_{r=1}^{3} n_r}{n_3} \sum_{i=0}^{n} \binom{n_1}{i}\binom{n_2}{i} = \prod_{j=3}^{4} \binom{\sum_{r=1}^{j} n_r}{n_j} \sum_{i=0}^{n} \binom{n_1}{i}\binom{n_2}{i}$$

Continuing in this manner, for p ≥ 3 treatments, the distinct permutations are enumerated through the expression

$$\prod_{j=3}^{p}\begin{pmatrix}\sum_{r=1}^{j} n_r \\ n_j\end{pmatrix}\sum_{i=0}^{n}\begin{pmatrix}n_1\\i\end{pmatrix}\begin{pmatrix}n_2\\i\end{pmatrix}, \text{ but } \sum_{i=0}^{n}\begin{pmatrix}n_1\\i\end{pmatrix}\begin{pmatrix}n_2\\i\end{pmatrix}=\begin{pmatrix}\sum_{r=1}^{2}n_r\\n_2\end{pmatrix}.$$ Therefore, in a p-sample problem, the number of distinct

permutations is $\prod_{j=1}^{p}\begin{pmatrix}\sum_{r=1}^{j}n_r\\n_j\end{pmatrix}$. An illustrative implementation of distinct permutation is found in [17]. Observe that for the

balanced case, the number of distinct permutations is $\prod_{j=3}^{p}\begin{pmatrix}jn\\n\end{pmatrix}\sum_{i=0}^{n}\begin{pmatrix}n\\i\end{pmatrix}^2=\prod_{j=1}^{p}\begin{pmatrix}jn\\n\end{pmatrix}$. Again, considering

consecutive number of treatments for a given experiment, the growth rate of the permutations from p–1 treatments to p

treatments is $\prod_{j=1}^{p}\begin{pmatrix}jn\\n\end{pmatrix}\bigg/\prod_{j=1}^{p-1}\begin{pmatrix}jn\\n\end{pmatrix}=\begin{pmatrix}np\\n\end{pmatrix}.$

3.0   The repeated measures (block) permutation

The repeated measures ANOVA approach tests the equality of means and is used when all members of a random sample are measured under a number of different conditions. This approach is used for several reasons. First, some research hypotheses require repeated measures as in longitudinal research. Second, in cases where there is a great deal of variation between sample members. Third, when sample members are difficult to recruit. In the repeated measures design, each trial represents the measurement of the same characteristic under a different condition. For the multivariate design, each trial represents the measurement of a different characteristic.

Given the layout of a multi-sample (n x k) experiment as $\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$, where x_{ij} is an observation in the *j*th

treatment and the *i*th block and the total number of observations in the experiment is *nk*. Rank the observations for each row from 1 (smallest x_{ij} on row *i*) to *k* (largest x_{ij} on row *i*). The permutation test procedure presented in this paper enumerates all the distinct permutations for a repeated measures experiment and therefore allows us to compute an empirical estimate of the cumulative distribution of the Friedman test statistic under the null hypothesis. Let the layout of the ranks (r_{ij}) of the

observations x_{ij} be $L_R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1k} \\ r_{21} & r_{22} & \cdots & r_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nk} \end{pmatrix}$. The data are arranged in *k* columns (treatments) and *n* rows (blocks),

where each block contains *k* repeated observations. Obviously, there are *k*! possible arrangements or permutations of each block and due to multiplication of choices, the entire layout of the *n* x *k* experiment requires $(k!)^n$ permutations of the observations to yield the exact distribution of a test statistic, with the permutations equally likely and each having the conditional probability $(k!)^{-n}$. Observe that the paired permutation model is a special case of the repeated measures ANOVA, i.e., for k = 2, $(k!)^n = 2^n$, and this explains the shapes of the graphs for experiments involving two blocks.

The first step in developing permutation algorithm is to formulate an initial configuration of the ranks of the variates of an experiment by taking the "trivial" configuration given below since any configuration of the ranks can engender all the distinct permutations.

	*Tr.*1	*Tr.*2	*Tr.*3	$\cdots$	*Tr.k*
Block 1	1	2	3	$\cdots$	k
Block 2	1	2	3	$\cdots$	k
Block 3	1	2	3	$\cdots$	k
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
Block n	1	2	3	$\cdots$	k

The test statistic is computed for each permutation in the complete enumeration of all the distinct permutations. The distribution of the test statistic is obtained by tabulating the distinct values of the test statistic against their probabilities of occurrence in the complete enumeration.

3.1 Methodology

The permutation methodology for repeated measures ANOVA is first built within a block for the number of treatments before implementing it recursively for all the blocks in the experiment.

Given the first block in the experiment, assuming there are two treatments, the 2! = 2 permutations of the variates are obtained by simply switching the variates: (α, β) and (β, α).

Assuming there are three treatments, the 3! = 6 permutations of the variates are obtained by first allowing the first variate "α" to occupy all the three possible positions in the layout: (α, β, χ), (β, α, χ) and (χ, β, α). The last two variates of these arrangements are subjected to the earlier permutation methodology of two variates to yield 3 x (2) = 6 permutations.

In a four treatment experiment, (α, β, χ, δ), the 4! = 24 permutations of the variates are obtained by first allowing the first variate "α" to occupy all the four possible positions in the layout: (α, β, χ, δ), (β, α, χ, δ), (χ, β, α, δ) and (δ, β, χ, α). The next three variates are subjected to the previous methodology for permuting three variates to yield 4 x (3 x 2) = 24 permutations.

The methodology described above is extended to accommodate the *k* treatments in the block to arrived at *k*! permutations, this therefore ensures the exhaustive enumeration of the permutations of the variates in the block. This methodology is recursively applied to all the *n* blocks in the experiment to arrive at $(k!)^n$ permutations, due to multiplication of choices, for an exhaustive enumeration.

The major problem with permutation test is that it requires a very large number of configurations to achieve an exhaustive enumeration. Considering two consecutive number of blocks for a given experiment, the growth rate of the permutations from *n*–1 blocks to *n* blocks is $\dfrac{(k!)^n}{(k!)^{n-1}} = k!$ and the growth rate of the permutations from *k*–1 treatments to

k treatments is $\dfrac{(k!)^n}{\left[(k-1)!\right]^n} = k^n$ and clearly, *k*! grows faster than k^n for a fixed *n*, and for a fixed *k*, *k*! is constant while k^n

explodes as *n* increases. Therefore, the growth rate of the permutations is higher for a unit increase in blocks than a unit increase in treatments for a fixed number of treatments and the reverse is the case when it is the number of blocks that is fixed.

As an illustration, an *n* x *k* experiment with *k* = 3 treatments (A, B, C) and *n* = 2 blocks (1, 2) could have the "trivial" configuration or permutation of ranks represented as:

	A	B	C
Block 1 $\rightarrow$	1	2	3
Block 2 $\rightarrow$	1	2	3

, each row gives rise to 3! = 6 permutations, that is, $(1\ \ 2\ \ 3)$, $(2\ \ 1\ \ 3)$, $(3\ \ 2\ \ 1)$, $(1\ \ 3\ \ 2)$, $(2\ \ 3\ \ 1)$ and $(3\ \ 1\ \ 2)$. Again, each of the 3! = 6 permutations of the first row (Block 1) combines with each of the 3! = 6 permutations of the second row (Block 2), leading to $(3!)^2 = 36$ permutations for an exhaustive enumeration as compiled in Table 1.

Table 1: Exhaustive permutation enumeration for *n*=2; *k*=3

1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
1	2	3	2	1	3	3	2	1	1	3	2	2	3	1	3	1	2
2	1	3	2	1	3	2	1	3	2	1	3	2	1	3	2	1	3
1	2	3	2	1	3	3	2	1	1	3	2	2	3	1	3	1	2
3	2	1	3	2	1	3	2	1	3	2	1	3	2	1	3	2	1
1	2	3	2	1	3	3	2	1	1	3	2	2	3	1	3	1	2
1	3	2	1	3	2	1	3	2	1	3	2	1	3	2	1	3	2
1	2	3	2	1	3	3	2	1	1	3	2	2	3	1	3	1	2
2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1
1	2	3	2	1	3	3	2	1	1	3	2	2	3	1	3	1	2
3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2
1	2	3	2	1	3	3	2	1	1	3	2	2	3	1	3	1	2

3.2   Block permutation test procedure

Let π_1, π_2, …, π_n be a set of all distinct permutations of the data set such that π_i is the *i*th permutation of the observations or ranks of the observations. Let $N = (k!)^n$, the permutation test procedure is defined as follows:

1.   Rank or use the observations in each Block independently $j = 1(1)k$ for *k* treatments

2.   Compute the Test Statistic T_1 for the original arrangement of ranks or observations π_1

3.   Obtain a distinct permutation π_p ; $p = 1(1)N$, of the experiment by permuting each block independently

4.   Compute the Test Statistic for permutation π_p , $T_p = T(\pi_p)$,

5.   Repeat Steps 3 and 4 for $p = 2, 3, …, N$

6.   Construct an empirical cumulative probability distribution

$$p_0 = \hat{p}(T_1 \le T_i) = \frac{1}{N}\sum_{i=1}^{N}\theta(T_1 - T_i)$$

     where θ is a step-function, that is, $\theta = 1$, if $T_1 \ge T_i$,   and $\theta = 0$ otherwise.

7.   Under the empirical distribution $\hat{p}$ if $p_0 \le \alpha$, reject the null hypothesis.
     The procedure computes the cumulative distribution of the Test Statistic, *T*, under the null hypothesis. This algorithm ensures that the p-values are accurately generated, thereby ensuring that the probability of making a type I error is exactly α.
     The problem with permutation tests has been high computational demands as in Steps 3-5. Available permutation procedures can sample from the permutation sample space rather than carrying out complete enumeration of all possible distinct permutations. In a survey of existing permutation procedures [14], it was discovered that none of these procedures can avoid the possibility of drawing the same sample more than once, thereby reducing the power of the permutation test.

3.3   Block permutation computer algorithm

The computer algorithm below implements the permutation of observations or rank of observations in a block with at most eight treatments. It can be extended accordingly by repeating sections of the codes. For a full implementation of the algorithm leading to the realisation of the repeated measures ANOVA, the algorithm should CALL itself *n* times for an *n*-block experiment.

Permutation Algorithm for Repeated Measures

1: Input number of repeated measures (k)
2: Rank each block from 1 to k
3: $Count \leftarrow 0$
4: for $j \leftarrow 1, k$ do
5:     $X_j \leftarrow Z_j \leftarrow j$
6: end for
7: for $j \leftarrow 1, k$ do
8:     $TP \leftarrow Z_1; X_1 \leftarrow Z_j; X_j \leftarrow TP$
9:     for $j1 \leftarrow 1, k$ do
10:       $X1_{j1} \leftarrow X_{j1}$
11:     end for
12:     for $j1 \leftarrow 2, k$ do
13:       $TP \leftarrow X_2; X1_2 \leftarrow X_{j1}; X1_{j1} \leftarrow TP$
14:       for $j2 \leftarrow 1, k$ do
15:         $X2_{j2} \leftarrow X1_{j2}$
16:       end for
17:       for $j2 \leftarrow 3, k$ do
18:         $TP \leftarrow X1_3; X2_3 \leftarrow X1_{j2}; X2_{j2} \leftarrow TP$
19:         for $j3 \leftarrow 1, k$ do
20:           $X3_{j3} \leftarrow X2_{j3}$

21:         end for
22:         for $j3 \leftarrow 4, k$ do
23:           $TP \leftarrow X2_4; X3_4 \leftarrow X2_{j3}; X3_{j3} \leftarrow TP$
24:           for $J4 \leftarrow 1, k$ do
25:             $X4_{j4} \leftarrow X3_{j4}$
26:           end for
27:           for $j4 \leftarrow 5, k$ do
28:             $TP \leftarrow X3_5; X4_5 \leftarrow X3_{j4}; X4_{j4} \leftarrow TP$
29:             for $j5 \leftarrow 1, k$ do
30:               $X5_{j5} \leftarrow X4_{j5}$
31:             end for
32:             for $j5 \leftarrow 6, k$ do
33:               $TP \leftarrow X4_6; X5_6 \leftarrow X4_{j5}; X5_{j5} \leftarrow TP$
34:               for $J6 \leftarrow 1, k$ do
35:                 $X6_{j6} \leftarrow X5_{j6}$
36:               end for
37:               for $j6 \leftarrow 7, k$ do
38:                 $TP \leftarrow X5_7; X6_7 \leftarrow X5_{j6}; X6_{j6} \leftarrow TP$

```
39:                              Count ← Count + 1
40:                              Print Count, X6_t; t ← 1(1)7
41:                              Compute and update pdf of test statistic
42:                              for jj ← 1, k do
43:                                  X6_jj ← X5_jj
44:                              end for
45:                          end for
46:                          for jj ← 1, k do
47:                              X5_jj ← X4_jj
48:                          end for
49:                      end for
50:                      for jj ← 1, k do
51:                          X4_jj ← X3_jj
52:                      end for
53:                  end for
54:                  for jj ← 1, k do
55:                      X3_jj ← X2_jj
56:                  end for
57:              end for
58:              for jj ← 1, k do
59:                  X2_jj ← X1_jj
60:              end for
61:          end for
62:          for jj ← 1, k do
63:              X1_jj ← X_jj
64:          end for
65:      end for
66:      for jj ← 1, k do
67:          X_jj ← Z_jj
68:      end for
69: end for
```

3.4.   Friedman test for randomized block design

The Friedman test is a non-parametric statistical test used for two-way repeated measures ANOVA by ranks similar to the parametric repeated measures ANOVA. It can detect differences in treatments across multiple test attempts. The procedure involves ranking each row (or block) together, then considering the values of ranks by columns. The hypotheses of interest are:

H_0: the k samples come from the same population.
H_1: at least one of the samples comes from a different population.

Under the truth of the null hypothesis, this test only requires exchangeability (or, if variances differ, compound symmetry) and the ability to rank the data. Given data $\{x_{ij}\}_{n \times k}$, that is, a tableau with n rows (the blocks), k columns (the treatments) and a single observation at the intersection of each block and treatment, calculate the ranks within each block. If there are tied values, assign to each tied value the average of the ranks that would have been assigned without ties. Replace the data with a new tableau $\{r_{ij}\}_{n \times k}$ where the entry r_{ij} is the rank of x_{ij} within block i.

Finally, when n or k is large (i.e. $n > 15$ or $k > 4$), the probability distribution of M can be approximated by that of a chi-square distribution. In this case the p-value is given by $P(\chi^2_{k-1} \geq M)$. If n or k is small, the approximation to chi-square becomes poor and the exact p-value of M for the Friedman test should be used when available. If the p-value is significant, appropriate post-hoc multiple comparisons tests would be performed.

$$M = \frac{12}{nk(k+1)} \sum_{j=1}^{k} R_j^2 - 3n(k+1)$$

where n is the number of blocks, k is the number of treatments, and R_j is the rank sum for the jth treatment, $j = 1, 2, 3, ..., k$. For large sample sizes, the critical values can be approximated by χ^2 with k-1 degrees of freedom [21 - 23].

4.0   Results

The algorithm (Permutation Algorithm for Repeated Measures) is implemented in Intel Visual Fortran. Figures 1-5 show the small sample probability distribution of the Friedman test statistic for different number of treatments and blocks. The resulting tables of exact critical values as obtained from the exact permutation distribution of the Friedman test statistic are presented in Tables 2-6, exact p-values in parenthesis.
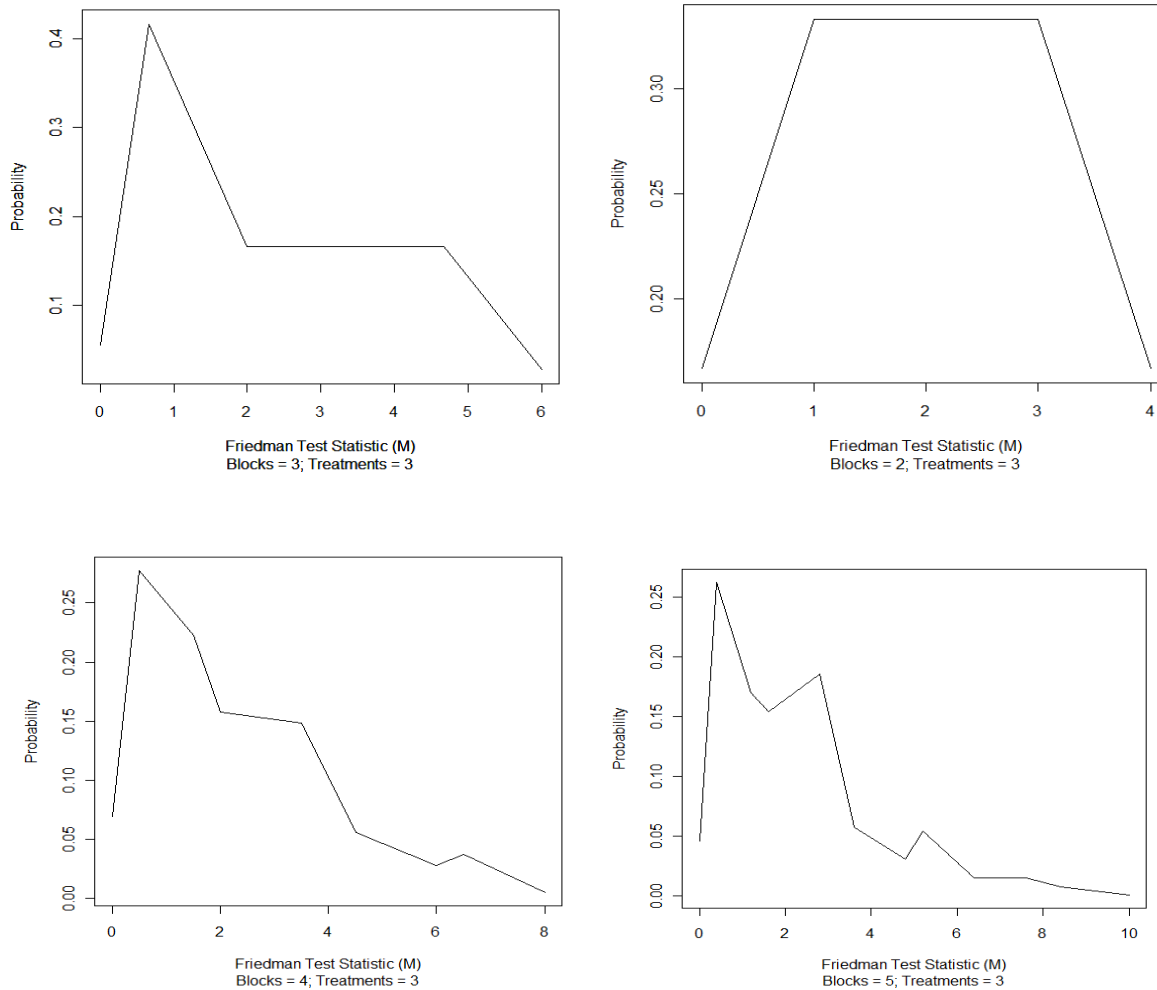
Figure 1: Distribution of Friedman test statistic for three treatments

Table 2: Upper Critical values for Friedman test statistic (3 Treatments)

| No of Blocks	$H_{0.9000}$	$H_{0.9500}$	$H_{0.9750}$	$H_{0.9900}$	$H_{0.9950}$	$H_{0.9975}$	$H_{0.9990}$
2	4.0 (0.1667)						
3	4.666667 (0.194444)	6.0 (0.027778)	6.0 (0.027778)				
4	4.5 (0.1250)	6.5 (0.0417)	8.0 (0.0046)	8.0 (0.0046)			
5	5.2 (0.0934)	6.4 (0.0394)	7.6 (0.0239)	8.4 (0.0085)	8.4 (0.0085)	10 (0.0008)	10 (0.0008)
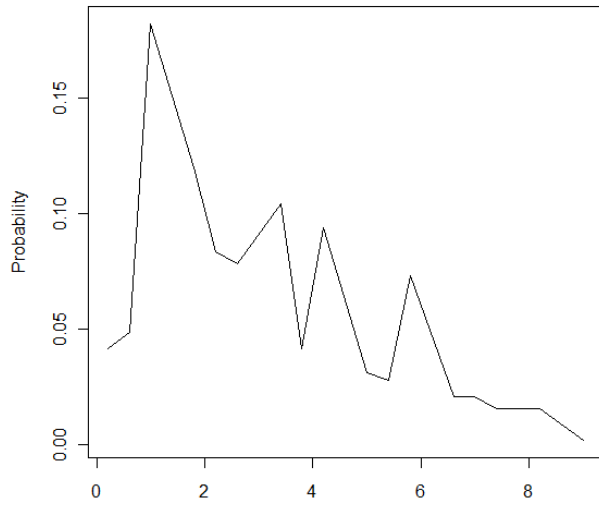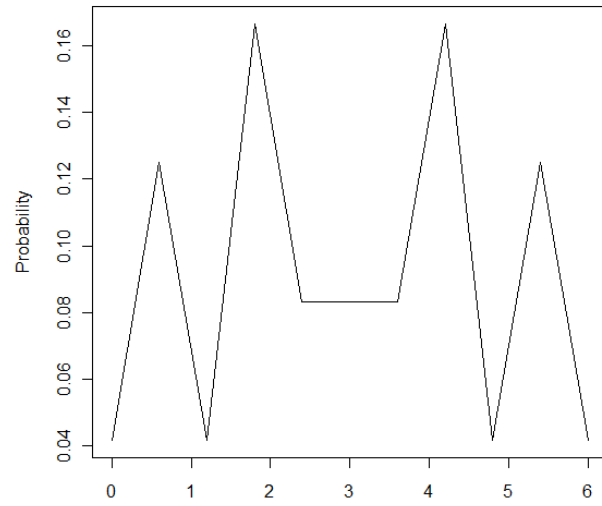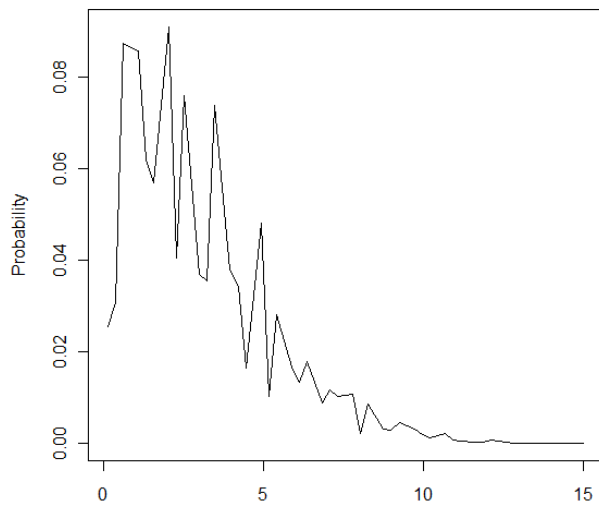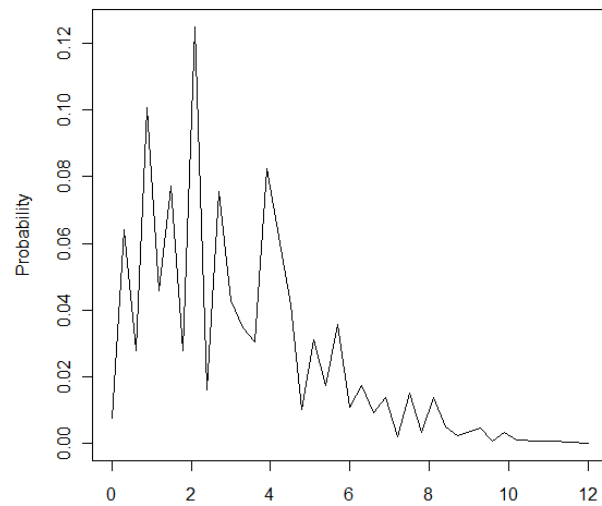
Figure 2: Distribution of Friedman test statistic for four treatments

Table 3: Upper Critical values for Friedman test statistic (4 Treatments)

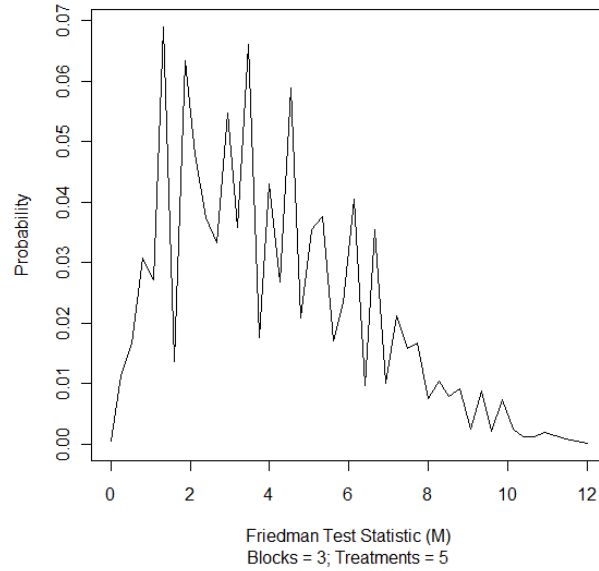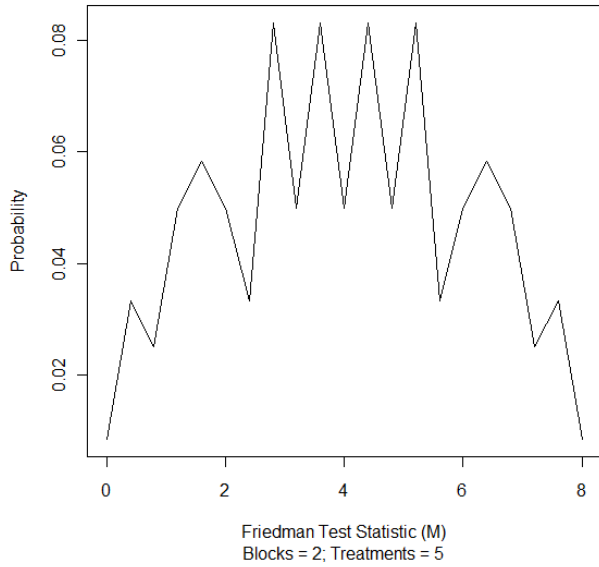| No of Blocks	$M_{0.9000}$	$M_{0.9500}$	$M_{0.9750}$	$M_{0.9900}$	$M_{0.9950}$	$M_{0.9975}$	$M_{0.9990}$
2	5.4 (0.1667)	6.0 (0.0417)					
3	6.6 (0.074653)	7.0 (0.053819)	8.2 (0.017361)	9.0 (0.001736)			
4	6.0 (0.1053)	7.5 (0.0517)	8.4 (0.0190)	9.3 (0.0115)	9.9 (0.0062)	10.2 (0.0027)	11.1 (0.0009)
5	6.36 (0.0933)	7.32 (0.0548)	8.76 (0.0226)	9.72 (0.0120)	10.68 (0.0055)	11.16 (0.0023)	12.12 (0.0014)
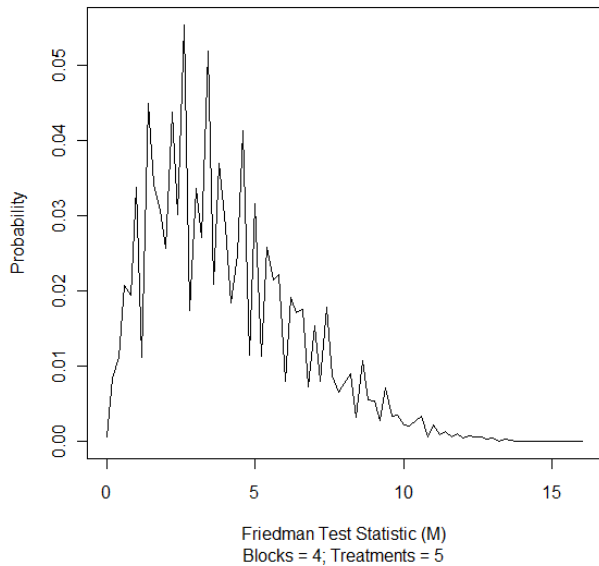
Figure 3: Distribution of Friedman test statistic for five treatments

Table 4: Upper Critical values for Friedman test statistic (5 Treatments)

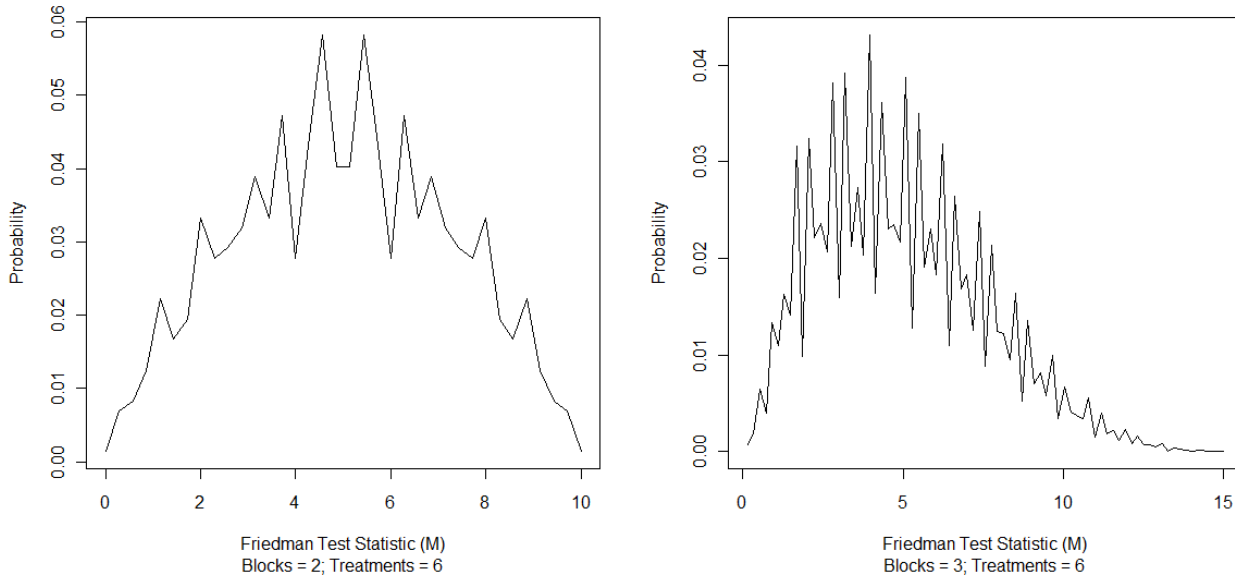| No of Blocks	$M_{0.9000}$	$M_{0.9500}$	$M_{0.9750}$	$M_{0.9900}$	$M_{0.9950}$	$M_{0.9975}$	$M_{0.9990}$
2	6.8 (0.1167)	7.6 (0.0417)	7.6 (0.0417)	8.0 (0.0083)			
3	7.466667 (0.0959)	8.533334 (0.0455)	9.333333 (0.0259)	10.13333 (0.0078)	10.4 (0.0053)	10.93333 (0.0028)	11.46667 (0.0009)
4	7.6 (0.0950)	8.8 (0.0489)	9.8 (0.0246)	11.0 (0.0102)	11.8 (0.0051)	12.6 (0.0022)	13.2 (0.0009)

Figure 4: Distribution of Friedman test statistic for six treatments

Table 5: Upper Critical values for Friedman test statistic (6 Treatments)

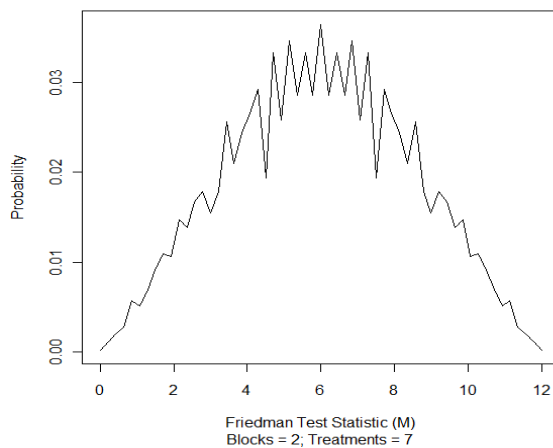| No of Blocks	$M_{0.9000}$	$M_{0.9500}$	$M_{0.9750}$	$M_{0.9900}$	$M_{0.9950}$	$M_{0.9975}$	$M_{0.9990}$
2	8.285714 (0.0875)	8.857142 (0.0514)	9.142858 (0.0292)	9.714286 (0.0083)	9.714286 (0.0083)	10.0 (0.0014)	10.0 (0.0014)
3	8.714286 (0.0957)	9.666667 (0.0560)	10.80952 (0.0247)	11.7619 (0.0095)	12.33333 (0.0052)	12.90476 (0.0022)	13.28571 (0.0010)

Figure 5: Distribution of Friedman test statistic for seven treatments

Table 6: Upper Critical values for Friedman test statistic (7 Treatments)

No of Blocks	$M_{0.9000}$	$M_{0.9500}$	$M_{0.9750}$	$M_{0.9900}$	$M_{0.9950}$	$M_{0.9975}$	$M_{0.9990}$
2	9.428572 (0.1000)	10.07143 (0.0548)	10.71429 (0.0240)	11.14286 (0.0119)	11.35714 (0.0062)	11.57143 (0.0034)	11.78571 (0.0014)

5.0   Conclusion

The difficulty of implementing the permutation test lies in obtaining all the distinct permutations of the observations in a given experiment. For example, a six-sample experiment with six blocks requires $(6!)^6$ = 139,314,069,504,000,000 (about 1.4×10^{17}) distinct permutations. The frequency distribution is constructed for all the distinct occurrences of the test statistic from which the probability distribution of the test statistic is computed.

       Figures 1-5 reveal the fact that the chi squared distribution, which is the large sample approximation of the Friedman test statistic, will poorly approximate the exact distribution of the Friedman test statistic for very small sample sizes. As sample and block sizes increase, the shape of the chi squared distribution begins to emerge. The critical values for the Friedman test statistic are determined by cutting off the most extreme $100\alpha\%$ of the theoretical frequency distribution of the test statistic, where α is the level of significance [24]. The critical values of the Friedman test statistic contained in Tables 2-6 are obtained from the enumeration of all the distinct permutations of the ranks of the variates in a repeated measures experiment. These critical values are exact and therefore ensure that the probability of a type I error in decisions arising from the use of the Friedman test is exactly α.

References

[1]   Fisher, R. A. (1935). *The design of experiments.* Oliver and Boyd, Edinburgh.
[2]   Scheffe, H. (1943) Statistical inference in the nonparametric case. *The Annals of Mathematical Statistics, 14,* 305-332.
[3]   Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans*. SIAM, Philadelphia, PA.
[4]   Bayarri, M. J. and Berger, J. O. (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science*, 19: 58-80.
[5]   Spiegelhalter, D. J. (2004). Incorporating Bayesian ideas into health-care evaluation. *Statistical Science*, 19, 156-174.
[6]   Fisher, R. A. (1936). The coefficient of racial likeness and the future of Craniometry. *J. Royal Anthropological Soc.*, 66: 57-63.
[7]   Ernst M. D. (2004). Permutation methods: A basis for exact inference. *Statistical Science*, 19: 676-685.
[8]   Good, P. (2000). *Permutation tests: a practical guide to resampling methods for testing hypotheses* (2nd edition). Springer Verlag, New York.
[9]   Wald, A and Wolfowitz, J. (1944). Statistical tests based on permutation of the observations. *The Annals of Mathematical Statistics*, 15, 358-372.
[10]   Agresti, A. (1992). A survey of exact inference for contingency tables, *Statistical Science, 7,* 131-177.
[11]   Hall, P. and I. Weissman (1997). On the estimation of extreme tail probabilities. *The Annals of Statistics*, 25: 1311-1326.
[12]   Ludbrook, J. and H. Dudley (1998). Why permutation tests are superior to t and F tests in biomedical research. *The Am. Stat.*, 52: 127-132.
[13]   Fahoome, G. (2002) Twenty nonparametric statistics and their large sample approximations. *Journal of Modern Applied Statistical Methods*, 1: 248-268.
[14]   Opdyke, J. D. (2003). Fast permutation tests that maximize power under conventional Monte Carlo sampling for pairwise and multiple comparisons. *Journal of Modern Applied Statistical Methods*, 2: 27-49.
[15]   Headrick, T. C. (2003). An algorithm for generating exact critical values for the Kruskal-Wallis one-way ANOVA. *Journal of Modern Applied Statistical Methods, 2,* 268-271.
[16]   Odiase, J. I. & Ogbonmwan, S. M. (2005a) An algorithm for generating unconditional exact permutation distribution for a two-sample experiment. *Journal of Modern Applied Statistical Methods, 4,* 319-332.
[17]   Odiase, J. I. and Ogbonmwan, S. M. (2005b). Exact permutation critical values for the kruskal-wallis one-way ANOVA. *Journal of Modern Applied Statistical Methods,* 4(2): 609-620.
[18]   Darwin, C. (1878). *The Effects of Cross and self-fertilization in the Vegetable Kingdom*, 2nd Ed., London: John Murray.
[19]   Fisher, R.A. (1971), *The design of experiments* (8th ed.), New York: Hafner Publishing.
[20]   Odiase, J. I. and Ogbonmwan, S. M. (2007). Exact permutation of paired observations: The challenge of R. A. Fisher. *Journal of Mathematics and Statistics,* 3(3): 116-121.
[21]   Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association* 32 (200): 675–701.
[22]   Friedman, M. (1939). A correction: The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 34 (205): 109.
[23]   Friedman, M. (1940). A comparison of alternative tests of significance for the problem of m rankings. *The Annals of Mathematical Statistics* 11 (1): 86–92.
[24]   Siegel, S. and Castellan, N. J. (1989). *Nonparametric statistics for the behavioural sciences* (3rd edition). McGraw-Hill, New York.