


5-1-2005

An Algorithm For Generating Unconditional Exact Permutation Distribution For A Two-Sample Experiment

Justice I. Odiase
University of Benin, Nigeria

Sunday M. Ogbonmwan
University of Benin, Nigeria

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Odiase, Justice I. and Ogbonmwan, Sunday M. (2005) "An Algorithm For Generating Unconditional Exact Permutation Distribution For A Two-Sample Experiment," *Journal of Modern Applied Statistical Methods*: Vol. 4 : Iss. 1 , Article 29.

DOI: 10.22237/jmasm/1114907340

Available at: <http://digitalcommons.wayne.edu/jmasm/vol4/iss1/29>

This Algorithms and Code is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.

An Algorithm For Generating Unconditional Exact Permutation Distribution For A Two-Sample Experiment

J. I. Odiase S. M. Ogbonmwan
Department of Mathematics
University of Benin, Nigeria

An Algorithm that generates the unconditional exact permutation distribution of a $2 \times n$ experiment is presented. The algorithm is able to handle ranks as well as actual observations. It makes it possible to obtain exact p-values for several statistics, especially when sample sizes are small and the application of large sample approximation is unreliable. An illustrative implementation is achieved and leads to the computation of exact p-values for the Mood test when the sample size is small.

Key words: permutation test, Monte Carlo test, p-value, rank order statistic, Mood test

Introduction

An important part of Statistical Inference is the representation of observed data in terms of a p-value. In fact, the p-value plays a major role in determining whether to accept or reject the null hypothesis. The p-value assists in establishing whether the observed data are statistically significant and so, any statistical approach that will guarantee its proper computation should be developed and employed in inferential statistics so that the probability of making a type I error is exactly α .

In practice, data are usually collected under varied conditions with some distributional assumptions such as that the data came from a normal distribution. It is advisable to avoid as much as possible making so many distributional

assumptions because data are usually never collected under ideal or perfect conditions, that is, do not conform perfectly to an assumed distribution or model being employed in its analysis. The p-value obtained through the permutation approach turns out to be the most reliable because it is exact, see Agresti (1992) and Good (2000).

If the experiment to be analyzed is made up of small or sparse data, large sample procedures for statistical inference are not appropriate (Senchaudhuri et al., 1995; Siegel & Castellan, 1988). In this article, consideration is given to the special case of $2 \times n$ tables with row and column totals allowed to vary with each permutation – this seems more natural than fixing the row and column totals. This is the unconditional exact permutation approach which is all-inclusive rather than the constrained or conditional exact permutation approach of fixing row and column totals. This later approach mainly addresses contingency tables (Agresti, 1992).

Several approaches have been suggested as alternatives to the computationally intensive unconditional exact permutation see Fisher (1935) and Agresti (1992) for a discussion on exact conditional permutation distribution. Also see Efron (1979), Hall and Tajvidi (2002), Efron and Tibshirani (1993), Opdyke (2003) for Monte Carlo approaches. Other approaches like the Bayesian and the likelihood have also been found useful in obtaining exact permutation

J. I. Odiase (justiceodiase@yahoo.com) is a Lecturer in the Department of Mathematics, University of Benin, Nigeria. His research interests include statistical computing and nonparametric statistics. S. M. Ogbonmwan (ogbonmwasmaltra@yahoo.co.uk) is an Associate Professor of Statistics, Department of Mathematics, University of Benin, Nigeria. His research interests include statistical computing and nonparametric statistics.

distribution (Bayarri & Berger, 2004; Spiegelhalter, 2004).

Large sample approximations are commonly adopted in several nonparametric tests as alternatives to tabulated exact critical values. The basic assumption required for such approximations to be reliable alternatives is that the sample size should be sufficiently large. However, there is no generally agreed upon definition of what constitutes a large sample size (Fahoome, 2002).

Available software for exact inference is expensive, with varied restrictions in the implementation of exact permutation procedures in the software. Computational time is highly prohibitive even with very fast processor speed of available personal computers. R. A. Fisher compiled by hand 32,768 permutations of Charles Darwin's data on the height of cross-fertilized and self-fertilized zea mays plants. The enormity of this task possibly discouraged Fisher from probing further into exact permutation tests (Ludbrook & Dudley, 1998).

Permutation tests provide exact results, especially when complete enumeration is feasible. A comprehensive documentation of the properties of permutation tests can be found in Pesarin (2001). The problem with permutation tests has been high computational demands, viz space and time complexities. Sampling from the permutation sample space rather than carrying out complete enumeration of all possible distinct rearrangements is what most of the available permutation procedures do, see Opdyke (2003) for a detailed listing of widely available permutation sampling procedures.

Opdyke (2003) however observed that most of the existing procedures can perform Monte Carlo sampling without replacement within a sample, but none can avoid the possibility of drawing the same sample more than once, thereby reducing the power of the permutation test.

The purpose of this article is to fashion out a sure and efficient way of obtaining unconditional exact permutation distribution by ensuring that a complete enumeration of all the distinct permutations of any 2-sample experiment is achieved. This will produce exact p-values and therefore ensure that the probability of making a type I error is exactly α .

This article also provides computer algorithms for achieving complete enumeration.

Methodology

Good (2000) considered the tails of permutation distribution in order to arrive at p-values, though he never carried out complete enumeration required for a permutation test. This approach has no precise model for the tail of the distribution from which data are drawn, (Hall & Weissman, 1997). The five steps for a permutation test presented in Good (2000) can be summarized thus:

1. Analyze the problem.
2. Choose a statistic and establish a rejection rule that will distinguish the hypothesis from the alternative.
3. Compute the test statistic for the original observations.
4. Rearrange the observations, compute the test statistic for every new arrangement and repeat this process until all permutations are obtained.
5. Construct the distribution for the test statistic based on Step 4.

Step 4 is where the difficulty in permutation test lies because a complete enumeration of all distinct permutations of the experiment is required. A 2-sample experiment with 15 variates in each sample requires 15,117,520 permutations. Clearly, the enumeration cannot be done manually, even if the computer produces 1000 permutations in a second, over 43 hours will be required for a complete enumeration. When this is achieved, p-values can be computed. Good (2000) identified the sufficient condition for a permutation test to be exact and unbiased against shifts in the direction of higher values as the exchangeability of the observations in the combined sample.

Let $X_i = (x_{i1}, x_{i2}, \dots, x_{in_i})^T$, $i = 1, 2$ and n_i is the i^{th} sample size. Also, let $X_N = (X_1, X_2)$, where $N = n_1 + n_2$. X_N is composed of N independent and identically distributed random variables. We have $\frac{(n_1 + n_2)!}{n_1! n_2!} = \frac{N!}{n_1! n_2!}$

possible permutations of the N variates of the 2 samples of size n, i = 1, 2 which are equally likely, each having the probability

$$\left(\frac{N!}{n_1!n_2!} \right)^{-1}$$

For equal sample sizes, n = n₁ = n₂, the number of permutations = $\frac{(2n)!}{(n!)^2}$ or $\frac{N!}{(n!)^2}$ and

the probability of each permutation = $\frac{(n!)^2}{N!}$.

For all possible permutations of the N variates, systematically develop a pattern necessary for the algorithm required for the generation of all the distinct permutations. The presentation of the systematic generation of all the possible permutations of the N variates now follows.

Examine an experiment of two samples (treatments), each with two variates, i.e., $\begin{pmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \end{pmatrix}$, where x₁₁, x₁₂, x₂₁ and x₂₂ represent sample values. Number of distinct arrangements = $\frac{4!}{2!2!} = 6$ (permutations) as listed in Table 1.

Table 1: Permutations of a 2 x 2 Experiment.

1	2	3	4	5	6
x ₁₁ x ₂₁	x ₂₁ x ₁₁	x ₂₂ x ₂₁	x ₁₁ x ₁₂	x ₁₁ x ₂₁	x ₂₁ x ₁₁
x ₁₂ x ₂₂	x ₁₂ x ₂₂	x ₁₂ x ₁₁	x ₂₁ x ₂₂	x ₂₂ x ₁₂	x ₂₂ x ₁₂

Numbers 1 – 6 on top of the permutations represent the permutation numbers

The actual process of permuting the variates of the experiment reveals the following.

$$\begin{pmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \end{pmatrix} \text{ original arrangement of the experiment } 1 \text{ permutation}$$

$$x_{11} \leftarrow x_{2i}, \quad i = 1, 2 \quad 2 \text{ permutations}$$

$$x_{12} \leftarrow x_{2i}, \quad i = 1, 2 \quad 2 \text{ permutations}$$

$$\begin{pmatrix} x_{21} & x_{11} \\ x_{22} & x_{12} \end{pmatrix} \text{ exchange the samples (columns) } 1 \text{ permutation}$$

In an attempt to offer a mathematical explanation for the method of exchanges of variates leading to the algorithm, observe that

$$\begin{pmatrix} 2 \\ 0 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \end{pmatrix} = 1 \times 1 = 1 \text{ Permutation (original arrangement of the experiment)}$$

$$\begin{pmatrix} 2 \\ 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 2 \times 2 = 4 \text{ Permutations (using one variate from first sample)}$$

$$\begin{pmatrix} 2 \\ 2 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \end{pmatrix} = 1 \times 1 = 1 \text{ Permutation (exchange the samples, i.e., 2 variates)}$$

$$\text{Total} = 1 + 4 + 1 = 6$$

Observe that permutation (1) is the original arrangement, permutations (2) to (5) are obtained by using the elements of the first column to interchange the elements of the second column, one at a time. Permutation (6) is obtained by interchanging the columns of the original arrangement of the experiment, making use of the two elements in the first column.

Examine a 2-sample experiment, where

each sample has 3 variates, i.e.
$$\begin{pmatrix} X_{11} & X_{21} \\ X_{12} & X_{22} \\ X_{13} & X_{23} \end{pmatrix}.$$

The expectation is to have $\frac{6!}{3!3!} = 20$

permutations, which are given in Table 2.

The process of permuting the variates reveal the following:

$$\begin{pmatrix} X_{11} & X_{21} \\ X_{12} & X_{22} \\ X_{13} & X_{23} \end{pmatrix} \quad \text{original}$$

arrangement of the experiment 1 permutation

$$\begin{array}{lll} X_{11} \leftarrow X_{2i}, & i = 1, 2, 3 & 3 \text{ permutations} \\ X_{12} \leftarrow X_{2i}, & i = 1, 2, 3 & 3 \text{ permutations} \\ X_{13} \leftarrow X_{2i}, & i = 1, 2, 3 & 3 \text{ permutations} \end{array}$$

$$\begin{pmatrix} x_{1s} \\ x_{1t} \end{pmatrix} \leftarrow \begin{pmatrix} x_{2i} \\ x_{2j} \end{pmatrix};$$

$s \neq t, i \neq j$ (3 x 3) 9 permutations

$$\begin{pmatrix} X_{21} & X_{11} \\ X_{22} & X_{12} \\ X_{23} & X_{13} \end{pmatrix}$$

exchange the samples (columns) 1 permutation

Again, observe that

$$\begin{pmatrix} 3 \\ 0 \end{pmatrix} \begin{pmatrix} 3 \\ 0 \end{pmatrix} = 1 \times 1 = 1 \text{ Permutation}$$

(original arrangement of the experiment)

$$\begin{pmatrix} 3 \\ 1 \end{pmatrix} \begin{pmatrix} 3 \\ 1 \end{pmatrix} = 3 \times 3 = 9 \text{ Permutations}$$

(using one variate from first sample)

$$\begin{pmatrix} 3 \\ 2 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} = 3 \times 3 = 9 \text{ Permutations}$$

(using two variates from first sample)

$$\begin{pmatrix} 3 \\ 3 \end{pmatrix} \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 1 \times 1 = 1 \text{ Permutation}$$

(exchange samples, i.e., three variates)

$$\text{Total} = 1 + 9 + 9 + 1 = 20$$

Similarly, observe that permutation (1) is the original matrix, permutations (2) to (10) are obtained by using the elements of the first column to interchange the elements of the second column, one at a time. Permutations (11) to (19) are obtained by using 2 elements of the first column to interchange the elements of the second column, and permutation (20) is obtained by interchanging the columns of the original arrangement of the experiment.

Continuing in the above fashion, clearly, the number of permutations for any 2-sample experiment can be written as

$$\begin{aligned} \sum_{i=0}^n \binom{n}{i} \binom{n}{i} &= \sum_{i=0}^n \binom{n}{i}^2 \\ &= \binom{2n}{n} \end{aligned}$$

for equal sample sizes. An adjustment for unequal sample sizes yields $\sum_{i=0}^{\min(n_1, n_2)} \binom{n_1}{i} \binom{n_2}{i}$

permutations, because $\binom{a}{b} = 0$ for $b > a$.

After obtaining all the distinct permutations from a complete enumeration, the statistic of interest is computed for each permutation. Each value of the statistic obtained from a complete enumeration occurs with

probability $\left(\frac{N!}{n_1! n_2!} \right)^{-1}$ for sample sizes, n_1

and n_2 , $N = n_1 + n_2$, this translates to $\frac{(n!)^2}{N!}$ for $n = n_1 = n_2$. The distribution of the statistic is

Table 2: Permutations of a 2 x 3 Experiment.

1 X ₁₁ X ₂₁ X ₁₂ X ₂₂ X ₁₃ X ₂₃	2 X ₂₁ X ₁₁ X ₁₂ X ₂₂ X ₁₃ X ₂₃	3 X ₂₂ X ₂₁ X ₁₂ X ₁₁ X ₁₃ X ₂₃	4 X ₂₃ X ₂₁ X ₁₂ X ₂₂ X ₁₃ X ₁₁	5 X ₁₁ X ₁₂ X ₂₁ X ₂₂ X ₁₃ X ₂₃
6 X ₁₁ X ₂₁ X ₂₂ X ₁₂ X ₁₃ X ₂₃	7 X ₁₁ X ₂₁ X ₂₃ X ₂₂ X ₁₃ X ₁₂	8 X ₁₁ X ₁₃ X ₁₂ X ₂₂ X ₂₁ X ₂₃	9 X ₁₁ X ₂₁ X ₁₂ X ₁₃ X ₂₂ X ₂₃	10 X ₁₁ X ₂₁ X ₁₂ X ₂₂ X ₂₃ X ₁₃
11 X ₂₁ X ₁₁ X ₂₂ X ₁₂ X ₁₃ X ₂₃	12 X ₂₁ X ₁₁ X ₂₃ X ₂₂ X ₁₃ X ₁₂	13 X ₂₂ X ₂₁ X ₂₃ X ₁₁ X ₁₃ X ₁₂	14 X ₂₁ X ₁₁ X ₁₂ X ₁₃ X ₂₂ X ₂₃	15 X ₂₁ X ₁₁ X ₁₂ X ₂₂ X ₂₃ X ₁₃
16 X ₂₂ X ₂₁ X ₁₂ X ₁₁ X ₂₃ X ₁₃	17 X ₁₁ X ₁₂ X ₂₁ X ₁₃ X ₂₂ X ₂₃	18 X ₁₁ X ₁₂ X ₂₁ X ₂₂ X ₂₃ X ₁₃	19 X ₁₁ X ₂₁ X ₂₂ X ₁₂ X ₂₃ X ₁₃	20 X ₂₁ X ₁₁ X ₂₂ X ₁₂ X ₂₃ X ₁₃

Numbers 1 – 20 on top of the permutations represent the permutation numbers.

thereafter obtained by simply tabulating the distinct values of the statistic against their probabilities of occurrence in the complete enumeration.

This method of obtaining unconditional exact permutation distribution also suffices when ranks of observations of an experiment are used instead of the actual observations. In handling ranks with this approach, tied observations do not pose any problems because the permutation process will be implemented as if the tied observations or ranks are distinct.

Given an n x p experiment,

$$X_N = \begin{pmatrix} x_{11} & \cdots & x_{p1} \\ \vdots & \vdots & \vdots \\ x_{1n} & \cdots & x_{pn} \end{pmatrix}, N = np$$

with x_{ij} as actual observations, $i = 1, 2, \dots, p, j = 1, 2, \dots, n$ for some rank order statistic, replace these observations with ranks. In order to achieve this, do a combined ranking from the smallest to the largest observation. For equal sample sizes, this yields an $n \times p$ matrix of ranks represented as follows:

$$R_N = \begin{pmatrix} R_1^{(1)} & \cdots & R_1^{(p)} \\ \vdots & \vdots & \vdots \\ R_n^{(1)} & \cdots & R_n^{(p)} \end{pmatrix},$$

$N = np$ and $R_i^{(j)}$ is the i th rank for sample j , see Sen and Puri (1967) for an expository discussion of rank order statistics. At this stage, the method can now be applied to this matrix of ranks. Note that any rearrangement or permutation of this matrix of ranks can be used in generating all the other distinct permutations.

The first step in developing the algorithm is to formulate the matrix of ranks, by adopting the trivial permutation, because it does not matter what rearrangement of the actual matrix of ranks is used in initiating the process of permutation, that is,

$$\begin{pmatrix} 1 & n_1 + 1 \\ 2 & n_1 + 2 \\ 3 & n_1 + 3 \\ \vdots & \vdots \\ n_1 & n_1 + n_2 \end{pmatrix}$$

and in the case of $n_1 = n_2 = n$,

$$\begin{pmatrix} 1 & n + 1 \\ 2 & n + 2 \\ 3 & n + 3 \\ \vdots & \vdots \\ n & 2n \end{pmatrix}.$$

For the above matrix of ranks, ensure that ties are taken care of, by replacing ranks of tied observations with the mean of their ranks.

In designing the computer algorithm for the method of complete enumeration via permutation described so far, it is intended that all statements should be read like sentences or as a sequence of commands. We write Set $T \leftarrow 1$, where Set is part of the statement language and T is a variable. Words that form the statement language required for this work include: do, od, else, for, if, fi, set, then, through, to, as used in Goodman and Hedetniemi (1977). To distinguish variable names from words in the statement language, variable names appear in full capital letters.

As a way of illustration, in formulating the computer algorithm for unconditional exact permutation distribution, a consideration is given to rank order statistic. The computer algorithm for the generation of the “trivial” matrix of ranks is presented in the next session for equal sample sizes.

Results

Algorithm (RANK) Generation of the trivial matrix of ranks

- Step 1. Set $P \leftarrow$ number of treatments;
 $K \leftarrow$ Number of variates
- Step 2. For $I \leftarrow 1$ to P do through Step 4
- Step 3. For $J \leftarrow 1$ to K do through Step 4
- Step 4. [X is the matrix of ranks]
 Set $X(J, I) \leftarrow (I - 1)K + J$ od

For all possible permutations of the N samples of p subsets of size n, the model of the number of permutations required for the computer algorithm for an experiment of two samples is:

$$\sum_{i=0}^n \binom{n}{i} \binom{n}{i} \text{ permutations}$$

where n is the number of variates in each sample (column) i.e., the balanced case. The computer algorithm now follows.

Algorithm (PERMUTATION)

Step 1 For $J1 \leftarrow 1$ to K do through Step 5
 Step 2 Set $TEMP \leftarrow X(J1, P - 1)$, $I1 \leftarrow P$
 Step 3 For $J2 \leftarrow 1$ to K do Step 5
 Step 4 Set $X(J1, P - 1) \leftarrow X(J2, I1)$, $X(J2, I1) \leftarrow TEMP$
 Step 5 [Compute statistic and restore original values of X] od
 Step 6 For $I \leftarrow 1$ to $K - 1$ do through Step 16
 Step 7 Set $TEMP1 \leftarrow X(I, P - 1)$
 Step 8 For $J \leftarrow I + 1$ to K do through Step 16
 Step 9 Set $TEMP2 \leftarrow X(J, P - 1)$
 Step 10 For $L \leftarrow P$ to P do through Step 16
 Step 11 For $I1 \leftarrow 1$ to K do through Step 16
 Step 12 For $L1 \leftarrow L$ to P do through Step 16
 Step 13 If $L \leftarrow L1$ then Set $T \leftarrow I1 + 1$
 else Set $T \leftarrow 1$ fi
 Step 14 For $J1 \leftarrow T$ to K do Step 16
 Step 15 Set $X(I, P - 1) \leftarrow X(I1, L)$, $X(I1, L) \leftarrow TEMP1$,
 $X(J, P - 1) \leftarrow X(J1, L1)$, $X(J1, L1) \leftarrow TEMP2$
 Step 16 [Compute statistic and restore original values of X] od
 Step 17 For $I \leftarrow 1$ to $K - 2$ do through Step 32
 Step 18 Set $TEMP1 \leftarrow X(I, P - 1)$
 Step 19 For $J \leftarrow I + 1$ to $K - 1$ do through Step 32
 Step 20 Set $TEMP2 \leftarrow X(J, P - 1)$
 Step 21 For $M \leftarrow J + 1$ to K do through Step 32
 Step 22 Set $TEMP3 \leftarrow X(M, P - 1)$
 Step 23 For $L \leftarrow P$ to P do through Step 32
 Step 24 For $I1 \leftarrow 1$ to K do through Step 32
 Step 25 For $L1 \leftarrow L$ to P do through Step 32
 Step 26 If $L \leftarrow L1$ then Set $T \leftarrow I1 + 1$
 else Set $T \leftarrow 1$ fi
 Step 27 For $J1 \leftarrow T$ to K do through Step 32
 Step 28 For $L2 \leftarrow L1$ to P do through Step 32
 Step 29 If $L1 \leftarrow L2$ then Set $T1 \leftarrow J1 + 1$
 else Set $T1 \leftarrow 1$ fi
 Step 30 For $J2 \leftarrow T1$ to K do Step 32
 Step 31 Set $X(I, P - 1) \leftarrow X(I1, L)$, $X(I1, L) \leftarrow TEMP1$,
 $X(J, P - 1) \leftarrow X(J1, L1)$, $X(J1, L1) \leftarrow TEMP2$,
 $X(M, P - 1) \leftarrow X(J2, L2)$, $X(J2, L2) \leftarrow TEMP3$
 Step 32 [Compute statistic and restore original values of X] od
 Step 33 For $I \leftarrow 1$ to $K - 3$ do through Step 53
 Step 34 Set $TEMP1 \leftarrow X(I, P - 1)$
 Step 35 For $J \leftarrow I + 1$ to $K - 2$ do through Step 53
 Step 36 Set $TEMP2 \leftarrow X(J, P - 1)$
 Step 37 For $M \leftarrow J + 1$ to $K - 1$ do through Step 53
 Step 38 Set $TEMP3 \leftarrow X(M, P - 1)$
 Step 39 For $N \leftarrow M + 1$ to K do through Step 53
 Step 40 Set $TEMP4 \leftarrow X(N, P - 1)$
 Step 41 For $L \leftarrow P$ to P do through Step 53
 Step 42 For $I1 \leftarrow 1$ to K do through Step 53


```

Step 43   For L1 ← L to P do through Step 53
Step 44   If L ← L1 then Set T ← I1 + 1
           else Set T ← 1 fi
Step 45   For J1 ← T to K do through Step 53
Step 46   For L2 ← L1 to P do through Step 53
Step 47   If L1 ← L2 then Set T1 ← J1 + 1
           else Set T1 ← 1 fi
Step 48   For J2 ← T1 to K do through Step 53
Step 49   For L3 ← L2 to P do through Step 53
Step 50   If L2 ← L3 then Set T2 ← J2 + 1
           else Set T2 ← 1 fi
Step 51   For J3 ← T2 to K do Step 53
Step 52   Set X(I, P - 1) ← X(I1, L), X(I1, L) ← TEMP1,
           X(J, P - 1) ← X(J1, L1), X(J1, L1) ← TEMP2,
           X(M, P - 1) ← X(J2, L2), X(J2, L2) ← TEMP3,
           X(N, P - 1) ← X(J3, L3), X(J3, L3) ← TEMP4 od
Step 53   [Compute statistic and restore original values of X] od
Step 54   [Interchange samples and compute statistic]

```

The PERMUTATION algorithm was translated to FORTRAN codes and implemented in Intel Visual FORTRAN for a 2 x 5 experiment. The 252 distinct permutations generated are presented in the Appendix. The algorithm can be extended to any sample size, depending on the processor speed and memory space of the computer being used to implement the algorithm. For an optimal management of computer memory (space complexity), the permutations are not stored, they are discarded immediately the statistic of interest is computed.

By way of illustration, generate the p-values for a 2 x 5 experiment for the Mood test. Fahoome (2002) noted that when $\alpha = 0.05$, sample size should exceed 5 for the large sample approximation to be adopted for the Mood test. The unconditional permutation approach makes it possible to obtain exact p-values even for fairly large sample sizes. Given two samples,

$y_{11}, y_{12}, \dots, y_{1n}$ and $y_{21}, y_{22}, \dots, y_{2n}$, the test statistic for the Mood test is

$$M = \sum_{i=1}^n \left(R_{li} - \frac{2n+1}{2} \right)^2$$

for equal sample sizes.

R_{li} is the rank of y_{li} , $i = 1, 2, \dots, n$ obtained after carrying out a combined ranking for the two samples. The large sample approximation for equal samples is

$$z = \frac{M - \frac{n(N^2 - 1)}{12}}{\sqrt{\frac{n^2(N+1)(N^2 - 4)}{180}}}$$

where $N = 2n$ and M the Mood test statistic.

The p-values obtained are presented in Table 3 and the distribution of the test statistic is represented graphically in Figure 1.

Table 3. p-values for Mood Statistic.

M	p(M)	p-value	M	p(M)	p-value
11.25	0.0079	0.0079	43.25	0.0397	0.6032
15.25	0.0079	0.0159	45.25	0.0476	0.6508
17.25	0.0159	0.0317	47.25	0.0714	0.7222
21.25	0.0317	0.0635	49.25	0.0397	0.7619
23.25	0.0317	0.0952	51.25	0.0397	0.8016
25.25	0.0159	0.1111	53.25	0.0476	0.8492
27.25	0.0397	0.1508	55.25	0.0397	0.8889
29.25	0.0476	0.1984	57.25	0.0159	0.9048
31.25	0.0397	0.2381	59.25	0.0317	0.9365
33.25	0.0397	0.2778	61.25	0.0317	0.9682
35.25	0.0714	0.3492	65.25	0.0159	0.9841
37.25	0.0476	0.3968	67.25	0.0079	0.9921
39.25	0.0397	0.4365	71.25	0.0079	1.0000
41.25	0.1270	0.5635			

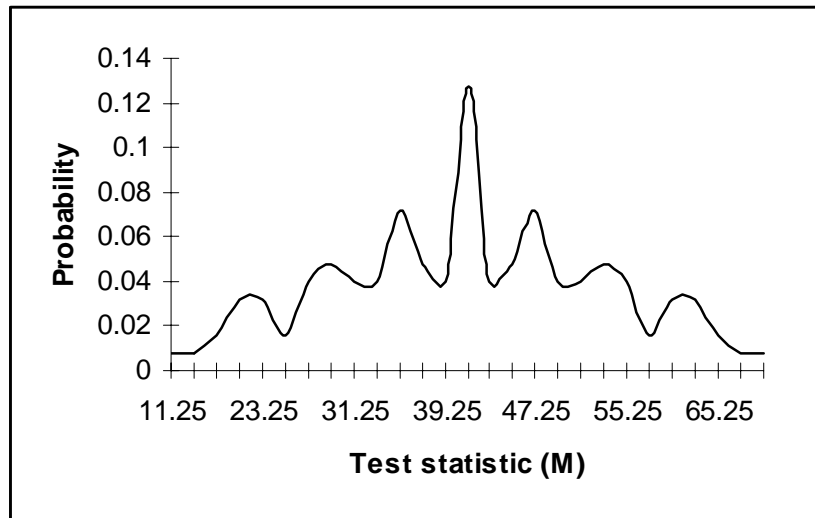


Figure 1. Exact distribution of Mood test statistic for a 2 x 5 experiment.

Clearly, results obtained from using Normal distribution, which is the large sample asymptotic distribution for the Mood test, will certainly not be exactly the same as using the exact permutation distribution, especially for small sample sizes. The permutation approach produces the exact p-values.

Example

Consider the following example on page 278 of Freund (1979) on difference of means.

Table 2: Heat-producing capacity of coal in millions of calories per tonne

Mine1	Mine2
8400	7510
8230	7690
8380	7720
7860	8070
7930	7660

Subjecting the data in Table 2 to Mood test, the test statistic (M) is 39.25 and from Table 3 containing unconditional exact permutation distribution of Mood test statistic, the corresponding p-value is 0.4365 which exceeds $\alpha = 0.05$, suggesting that we cannot reject the null hypothesis of no difference between the heat-producing capacity of coal from the two mines. Adopting the large sample Normal approximation for Mood test, z calculated is -0.17 which gives a p-value of 0.4325 and this exceeds $\alpha/2 = 0.025$, meaning that the observed data are compatible with the null hypothesis of no difference as earlier obtained from the exact permutation test.

Conclusion

Several authors have attempted to obtain exact p-values for different statistics using the permutation approach. Two things have made their attempts an uphill task. First is the speed of computer required to perform a permutation test. Until recently, the speed of available computers has been grossly inadequate to handle complete enumeration for even small sample sizes. Recent advances in computer design has drawn researchers in this area closer to the realization of complete enumeration even for fairly large

sample sizes. Secondly, the intensive looping in computer programming required for complete enumeration for unconditional exact permutation test demands a good programming skill.

In this article, a straight forward but computer intensive approach has been adopted in creating an algorithm that can carryout a systematic enumeration of distinct permutations of a 2-sample experiment. With this algorithm, the p-values for statistics involving two samples can be accurately generated, thereby ensuring that the probability of making a type I error is exactly α

References

- Agresti, A. (1992). A survey of exact inference for contingency tables, *Statistical Science*, 7, 131-177.
- Bayarri, M. J. & Berger, J. O. (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science*, 19, 58-80.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7, 1-26.
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. NY: Chapman and Hall.
- Fahoome, G. (2002). Twenty nonparametric statistics and their large sample approximations. *Journal of Modern Applied Statistical Methods*, 1, 248-268.
- Fisher, R. A. (1935). *Design of experiments*. Oliver and Boyd, Edinburgh.
- Freund, J. E. (1979). *Modern elementary statistics* (5th edition). Englewood Cliff, N. J.: Prentice-Hall.
- Good, P. (2000). *Permutation tests: a practical guide to resampling methods for testing hypotheses* (2nd edition). NY: Springer Verlag.
- Goodman, S. E., & Hedetniemi, S. T. (1977). *Introduction to the design and analysis of algorithms*. London: McGraw-Hill Book Company.
- Hall, P., & Tajvidi, N. (2002). Permutation tests for equality of distributions in high dimensional settings. *Biometrika*, 89, 359-374.
- Hall, P., & Weissman, I. (1997). On the estimation of extreme tail probabilities. *The Annals of Statistics*, 25, 1311-1326.

Ludbrook, J., & Dudley, H. (1998). Why permutation tests are superior to t and F tests in biomedical research. *The American Statistician*, 52, 127-132.

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.

Pesarin, F. (2001). *Multivariate permutation tests*. NY: Wiley.

Sen, P. K., & Puri, M. L. (1967). On the theory of rank order tests for location in the multivariate one sample problem. *The Annals of Mathematical Statistics*, 38, 1216-1228.

Senchaudhuri, P., Mehta, C. R., & Patel, N. T. (1995). Estimating exact p-values by the method of control variates or Monte Carlo rescue. *Journal of the American Statistical Association*, 90, 640-648.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd edition). NY: McGraw-Hill.

Spiegelhalter, D. J. (2004). Incorporating Bayesian ideas into health-care evaluation. *Statistical Science*, 19, 156-174

Appendix: Permutations of a 2 x 5 Experiment.

1	2	3	4	5	6	7	8	9	10	11	12
1 6	6 1	7 6	8 6	9 6	10 6	1 2	1 6	1 6	1 6	1 6	1 3
2 7	2 7	2 1	2 7	2 7	2 7	6 7	7 2	8 7	9 7	10 7	2 7
3 8	3 8	3 8	3 1	3 8	3 8	3 8	3 8	3 2	3 8	3 8	6 8
4 9	4 9	4 9	4 9	4 1	4 9	4 9	4 9	4 9	4 2	4 9	4 9
5 10	5 10	5 10	5 10	5 10	5 1	5 10	5 10	5 10	5 10	5 2	5 10
13	14	15	16	17	18	19	20	21	22	23	24
1 6	1 6	1 6	1 6	1 4	1 6	1 6	1 6	1 6	1 5	1 6	1 6
2 3	2 7	2 7	2 7	2 7	2 4	2 7	2 7	2 7	2 7	2 5	2 7
7 8	8 3	9 8	10 8	3 8	3 8	3 4	3 8	3 8	3 8	3 8	3 5
4 9	4 9	4 3	4 9	6 9	7 9	8 9	9 4	10 9	4 9	4 9	4 9
5 10	5 10	5 10	5 3	5 10	5 10	5 10	5 10	5 4	6 10	7 10	8 10
25	26	27	28	29	30	31	32	33	34	35	36
1 6	1 6	6 1	6 1	6 1	6 1	7 6	7 6	7 6	8 6	8 6	9 6
2 7	2 7	7 2	8 7	9 7	10 7	8 1	9 1	10 1	9 7	10 7	10 7
3 8	3 8	3 8	3 2	3 8	3 8	3 2	3 8	3 8	3 1	3 1	3 8
4 5	4 9	4 9	4 9	4 2	4 9	4 9	4 2	4 9	4 2	4 9	4 1
9 10	10 5	5 10	5 10	5 10	5 2	5 10	5 10	5 2	5 10	5 2	5 2
37	38	39	40	41	42	43	44	45	46	47	48
6 1	6 1	6 1	6 1	7 6	7 6	7 6	8 6	8 6	9 6	6 1	6 1
2 3	2 7	2 7	2 7	2 1	2 1	2 1	2 7	2 7	2 7	2 4	2 7
7 8	8 3	9 8	10 8	8 3	9 8	10 8	9 1	10 1	10 8	3 8	3 4
4 9	4 9	4 3	4 9	4 9	4 3	4 9	4 3	4 9	4 1	7 9	8 9
5 10	5 10	5 10	5 3	5 10	5 10	5 3	5 10	5 3	5 3	5 10	5 10

Appendix Continued:

49	50	51	52	53	54	55	56	57	58	59	60
6 1	6 1	7 6	7 6	7 6	8 6	8 6	9 6	6 1	6 1	6 1	6 1
2 7	2 7	2 1	2 1	2 1	2 7	2 7	2 7	2 5	2 7	2 7	2 7
3 8	3 8	3 4	3 8	3 8	3 1	3 1	3 8	3 8	3 5	3 8	3 8
9 4	10 9	8 9	9 4	10 9	9 4	10 9	10 1	4 9	4 9	4 5	4 9
5 10	5 4	5 10	5 10	5 4	5 10	5 4	5 4	7 10	8 10	9 10	10 5
61	62	63	64	65	66	67	68	69	70	71	72
7 6	7 6	7 6	8 6	8 6	9 6	1 2	1 2	1 2	1 2	1 6	1 6
2 1	2 1	2 1	2 7	2 7	2 7	6 3	6 7	6 7	6 7	7 2	7 2
3 5	3 8	3 8	3 1	3 1	3 8	7 8	8 3	9 8	10 8	8 3	9 8
4 9	4 5	4 9	4 5	4 9	4 1	4 9	4 9	4 3	4 9	4 9	4 3
8 10	9 10	10 5	9 10	10 5	10 5	5 10	5 10	5 10	5 3	5 10	5 10
73	74	75	76	77	78	79	80	81	82	83	84
1 6	1 6	1 6	1 6	1 2	1 2	1 2	1 2	1 6	1 6	1 6	1 6
7 2	8 7	8 7	9 7	6 4	6 7	6 7	6 7	7 2	7 2	7 2	8 7
10 8	9 2	10 2	10 8	3 8	3 4	3 8	3 8	3 4	3 8	3 8	3 2
4 9	4 3	4 9	4 2	7 9	8 9	9 4	10 9	8 9	9 4	10 9	9 4
5 3	5 10	5 3	5 3	5 10	5 10	5 10	5 4	5 10	5 10	5 4	5 10
85	86	87	88	89	90	91	92	93	94	95	96
1 6	1 6	1 2	1 2	1 2	1 2	1 6	1 6	1 6	1 6	1 6	1 6
8 7	9 7	6 5	6 7	6 7	6 7	7 2	7 2	7 2	8 7	8 7	9 7
3 2	3 8	3 8	3 5	3 8	3 8	3 5	3 8	3 8	3 2	3 2	3 8
10 9	10 2	4 9	4 9	4 5	4 9	4 9	4 5	4 9	4 5	4 9	4 2
5 4	5 4	7 10	8 10	9 10	10 5	8 10	9 10	10 5	9 10	10 5	10 5
97	98	99	100	101	102	103	104	105	106	107	108
1 3	1 3	1 3	1 3	1 6	1 6	1 6	1 6	1 6	1 6	1 3	1 3
2 4	2 7	2 7	2 7	2 3	2 3	2 3	2 7	2 7	2 7	2 5	2 7
6 8	6 4	6 8	6 8	7 4	7 8	7 8	8 3	8 3	9 8	6 8	6 5
7 9	8 9	9 4	10 9	8 9	9 4	10 9	9 4	10 9	10 3	4 9	4 9
5 10	5 10	5 10	5 4	5 10	5 10	5 4	5 10	5 4	5 4	7 10	8 10
109	110	111	112	113	114	115	116	117	118	119	120
1 3	1 3	1 6	1 6	1 6	1 6	1 6	1 6	1 4	1 4	1 4	1 4
2 7	2 7	2 3	2 3	2 3	2 7	2 7	2 7	2 5	2 7	2 7	2 7
6 8	6 8	7 5	7 8	7 8	8 3	8 3	9 8	3 8	3 5	3 8	3 8
4 5	4 9	4 9	4 5	4 9	4 5	4 9	4 3	6 9	6 9	6 5	6 9
9 10	10 5	8 10	9 10	10 5	9 10	10 5	10 5	7 10	8 10	9 10	10 5
121	122	123	124	125	126	127	128	129	130	131	132
1 6	1 6	1 6	1 6	1 6	1 6	6 1	6 1	6 1	6 1	6 1	6 1
2 4	2 4	2 4	2 7	2 7	2 7	7 2	7 2	7 2	8 7	8 7	9 7
3 5	3 8	3 8	3 4	3 4	3 8	8 3	9 8	10 8	9 2	10 2	10 8
7 9	7 5	7 9	8 5	8 9	9 4	4 9	4 3	4 9	4 3	4 9	4 2
8 10	9 10	10 5	9 10	10 5	10 5	5 10	5 10	5 3	5 10	5 3	5 3

Appendix Continued:

133	134	135	136	137	138	139	140	141	142	143	144
7 6	7 6	7 6	8 6	6 1	6 1	6 1	6 1	6 1	6 1	7 6	7 6
8 1	8 1	9 1	9 7	7 2	7 2	7 2	8 7	8 7	9 7	8 1	8 1
9 2	10 2	10 8	10 1	3 4	3 8	3 8	3 2	3 2	3 8	3 2	3 2
4 3	4 9	4 2	4 2	8 9	9 4	10 9	9 4	10 9	10 2	9 4	10 9
5 10	5 3	5 3	5 3	5 10	5 10	5 4	5 10	5 4	5 4	5 10	5 4
145	146	147	148	149	150	151	152	153	154	155	156
7 6	8 6	6 1	6 1	6 1	6 1	6 1	6 1	7 6	7 6	7 6	8 6
9 1	9 7	7 2	7 2	7 2	8 7	8 7	9 7	8 1	8 1	9 1	9 7
3 8	3 1	3 5	3 8	3 8	3 2	3 2	3 8	3 2	3 2	3 8	3 1
10 2	10 2	4 9	4 5	4 9	4 5	4 9	4 2	4 5	4 9	4 2	4 2
5 4	5 4	8 10	9 10	10 5	9 10	10 5	10 5	9 10	10 5	10 5	10 5
157	158	159	160	161	162	163	164	165	166	167	168
6 1	6 1	6 1	6 1	6 1	6 1	7 6	7 6	7 6	8 6	6 1	6 1
2 3	2 3	2 3	2 7	2 7	2 7	2 1	2 1	2 1	2 7	2 3	2 3
7 4	7 8	7 8	8 3	8 3	9 8	8 3	8 3	9 8	9 1	7 5	7 8
8 9	9 4	10 9	9 4	10 9	10 3	9 4	10 9	10 3	10 3	4 9	4 5
5 10	5 10	5 4	5 10	5 4	5 4	5 10	5 4	5 4	5 4	8 10	9 10
169	170	171	172	173	174	175	176	177	178	179	180
6 1	6 1	6 1	6 1	7 6	7 6	7 6	8 6	6 1	6 1	6 1	6 1
2 3	2 7	2 7	2 7	2 1	2 1	2 1	2 7	2 4	2 4	2 4	2 7
7 8	8 3	8 3	9 8	8 3	8 3	9 8	9 1	3 5	3 8	3 8	3 4
4 9	4 5	4 9	4 3	4 5	4 9	4 3	4 3	7 9	7 5	7 9	8 5
10 5	9 10	10 5	10 5	9 10	10 5	10 5	10 5	8 10	9 10	10 5	9 10
181	182	183	184	185	186	187	188	189	190	191	192
6 1	6 1	7 6	7 6	7 6	8 6	1 2	1 2	1 2	1 2	1 2	1 2
2 7	2 7	2 1	2 1	2 1	2 7	6 3	6 3	6 3	6 7	6 7	6 7
3 4	3 8	3 4	3 4	3 8	3 1	7 4	7 8	7 8	8 3	8 3	9 8
8 9	9 4	8 5	8 9	9 4	9 4	8 9	9 4	10 9	9 4	10 9	10 3
10 5	10 5	9 10	10 5	10 5	10 5	5 10	5 10	5 4	5 10	5 4	5 4
193	194	195	196	197	198	199	200	201	202	203	204
1 6	1 6	1 6	1 6	1 2	1 2	1 2	1 2	1 2	1 2	1 6	1 6
7 2	7 2	7 2	8 7	6 3	6 3	6 3	6 7	6 7	6 7	7 2	7 2
8 3	8 3	9 8	9 2	7 5	7 8	7 8	8 3	8 3	9 8	8 3	8 3
9 4	10 9	10 3	10 3	4 9	4 5	4 9	4 5	4 9	4 3	4 5	4 9
5 10	5 4	5 4	5 4	8 10	9 10	10 5	9 10	10 5	10 5	9 10	10 5
205	206	207	208	209	210	211	212	213	214	215	216
1 6	1 6	1 2	1 2	1 2	1 2	1 2	1 2	1 6	1 6	1 6	1 6
7 2	8 7	6 4	6 4	6 4	6 7	6 7	6 7	7 2	7 2	7 2	8 7
9 8	9 2	3 5	3 8	3 8	3 4	3 4	3 8	3 4	3 4	3 8	3 2
4 3	4 3	7 9	7 5	7 9	8 5	8 9	9 4	8 5	8 9	9 4	9 4
10 5	10 5	8 10	9 10	10 5	9 10	10 5	10 5	9 10	10 5	10 5	10 5

Appendix Continued:

217	218	219	220	221	222	223	224	225	226	227	228
1 3	1 3	1 3	1 3	1 3	1 3	1 6	1 6	1 6	1 6	6 1	6 1
2 4	2 4	2 4	2 7	2 7	2 7	2 3	2 3	2 3	2 7	7 2	7 2
6 5	6 8	6 8	6 4	6 4	6 8	7 4	7 4	7 8	8 3	8 3	8 3
7 9	7 5	7 9	8 5	8 9	9 4	8 5	8 9	9 4	9 4	9 4	10 9
8 10	9 10	10 5	9 10	10 5	10 5	9 10	10 5	10 5	10 5	5 10	5 4
229	230	231	232	233	234	235	236	237	238	239	240
6 1	6 1	7 6	6 1	6 1	6 1	6 1	7 6	6 1	6 1	6 1	6 1
7 2	8 7	8 1	7 2	7 2	7 2	8 7	8 1	7 2	7 2	7 2	8 7
9 8	9 2	9 2	8 3	8 3	9 8	9 2	9 2	3 4	3 4	3 8	3 2
10 3	10 3	10 3	4 5	4 9	4 3	4 3	4 3	8 5	8 9	9 4	9 4
5 4	5 4	5 4	9 10	10 5	10 5	10 5	10 5	9 10	10 5	10 5	10 5
241	242	243	244	245	246	247	248	249	250	251	252
7 6	6 1	6 1	6 1	6 1	7 6	1 2	1 2	1 2	1 2	1 6	6 1
8 1	2 3	2 3	2 3	2 7	2 1	6 3	6 3	6 3	6 7	7 2	7 2
3 2	7 4	7 4	7 8	8 3	8 3	7 4	7 4	7 8	8 3	8 3	8 3
9 4	8 5	8 9	9 4	9 4	9 4	8 5	8 9	9 4	9 4	9 4	9 4
10 5	9 10	10 5	10 5	10 5	10 5	9 10	10 5	10 5	10 5	10 5	10 5

Numbers 1 – 252 on top of the permutations represent the permutation numbers