

## Exact Permutation Algorithm for Paired Observations: The Challenge of R. A. Fisher

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

---

**Abstract:** The major handicap of permutation test is the logical and computational requirement necessary to develop and implement the exact permutation scheme. This study provides an algorithm that systematically enumerates all the distinct permutations of the paired observations in an experiment without the possibility of repeating any of the permutations. The permutation algorithm presented completely breaks down the permutation problem for ease of implementation and analysis. The algorithm was illustratively implemented in Intel Visual Fortran to recreate Fisher's manual compilation of 32,768 permutations of Charles Darwin's data on heights of cross-fertilized and self-fertilized plants. The algorithm provides exact p-values for any experiment involving paired observations and exposes the danger in using asymptotic or parametric distributions such as the t-test to analyze small data sets when the exact functional form of the distribution is not explicitly known. This becomes more obvious especially when the experiment leads to a p-value close to the threshold level of significance. The exact distribution and the graphical presentation provided in this study give credence to the use of the permutation test.

**Key words:** Algorithm, paired observations, permutation, p-value, t-test

---

### INTRODUCTION

In hypothesis testing, paired t-test is the procedure used when the independent variable is within subjects in nature and compares two levels of the independent variable assigned to the same group of subjects at different points in time. As is true for all t-tests, the standard error is not known and is estimated from sample data.

Fisher<sup>[1]</sup> proposed that randomization should be the basis for experimental design and statistical inference. The premise behind experiments involving paired observations is that a sample of experimental units, however acquired, is divided randomly into two or more groups. These are then exposed to different treatments. The null hypothesis is that the treatments have no differential effects on the groups with respect to a selected statistic. If there is no requirement that the test statistic should conform to a mathematically definable frequency distribution, then the exact sampling distribution of the test statistic can be compiled by permutation. Scheffe<sup>[2]</sup> demonstrates that for a general class of problems, the permutation approach is the only possible method of constructing exact tests of significance.

Theoretical frequency distribution of a test statistic for many nonparametric tests is estimated by either

using the small sample test statistic or using the large sample asymptotic distribution of the test statistic. For small sample size, the exact probability of obtaining the calculated value of the test statistic or any less likely value has to be determined. The use of the asymptotic test with small sample sizes may yield an incorrect p-value and therefore leads to a false acceptance or rejection of the null hypothesis. In an attempt to ensure that the probability of a type I error is exactly  $\alpha$ , an algorithm for obtaining exact permutation distribution of paired observations is presented in this work.

The permutation approach is computationally demanding viz; space and time complexities. Several approaches, which are computationally less demanding as the permutation approach, have been suggested as alternatives. Efron and Tibshirani<sup>[3]</sup>, Hall and Tajvidi<sup>[4]</sup>, Opdyke<sup>[5]</sup> presented Monte Carlo approaches. Other approaches like the Bayesian and the likelihood have also been found useful in obtaining approximate permutation distribution, see Bayarri and Berger<sup>[6]</sup>, Spiegelhalter<sup>[7]</sup>.

StatXact is nonparametric software that provides the distribution of various test statistics by adopting a network approach but chiefly addresses rank based statistics. Opdyke<sup>[5]</sup> observes that available permutation procedures can sample from the permutation sample space rather than carrying out complete enumeration of

all possible distinct rearrangements. These available 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 study was to provide a simple but systematic way of obtaining exact permutation distribution of paired observations by ensuring that a complete enumeration of all the distinct permutations of the actual observations is achieved. This leads to the recreation of the Fisher's manual compilation of the 32,768 permutation of Charles Darwin's data on heights of cross-fertilized and self-fertilized plants (Darwin<sup>[8]</sup>), see Ludbrook and Dudley<sup>[9]</sup>.

**Permutation test:** Pairing involves repeated measurements on the same set of experimental units, while paired permutation involves the compilation of all possible interchanges of the values attached to each experimental unit. The null hypothesis is that there is no differential effect of experimental conditions on the statistic of interest. The probability attached to the null hypothesis is therefore

$$P_{n_0} = \frac{\text{Number of the same or more extreme outcomes as that observed}}{\text{Total number of possible outcomes}}$$

Ludbrook and Dudley<sup>[9]</sup>. Clear descriptions of how permutation tests are conducted for categorical data can be found in Siegel and Castellan<sup>[10]</sup>. One of the practical difficulties with exact permutation tests is realizing all the possible distinct arrangement of the variates in an experiment.

For  $k$  repeated measurements on the same group of size  $n$ , the general formula is  $k^n$ . For example, for two ( $k = 2$ ) repeated measurements made on a group of size  $n = 15$ , as in the case of Charles Darwin's data on heights of cross-fertilized and self-fertilized plants, the number of possible distinct permutations is  $2^{15} = 32,768$ . Even with a fast microcomputer, compilation of the exact permutation distribution may take too long.

Computational time for a permutation test is highly prohibitive. 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 as reported in Ludbrook and Dudley<sup>[9]</sup>. It is believed that the challenge inherent in the task of manually compiling permutations is what possibly discouraged Fisher from further research into exact permutation tests. It is also believed that the nonparametric test whose theoretical basis is the permutation of differences between mean ranks invented by Wilcoxon in 1945 is an attempt to overcome the computational difficulties of permuting differences between means, see Ludbrook and Dudley<sup>[9]</sup>.

## MATERIALS AND METHODS

In this study, consideration is given to the permutation distribution of paired observations, see Agresti<sup>[11]</sup> for conditional permutation; Odiase and Ogbonmwan<sup>[12]</sup> for an algorithm for generating unconditional exact permutation distribution for a two-sample (independent samples) experiment. In Good<sup>[13]</sup>, a consideration was given to the tails of permutation distribution in order to arrive at p-values. This approach has no precise model for the tail of the distribution from which data are drawn, see Hall and Weissman<sup>[4]</sup>.

The difficulty in permutation test lies in obtaining all the distinct arrangements of the variates in a given experiment. For example, a paired two-sample experiment with 20 variates in each sample requires  $2^{20}$  or 1,048,576 permutations. A frequency distribution is subsequently arrived at for all the distinct occurrences of the test statistic from which the probability distribution of the test statistic is computed. A paired two-sample experiment with  $n$  pairs has  $2^n$  possible permutations of the variates with each permutation occurring with probability  $2^{-n}$ .

Given two paired samples  $X = (x_1, x_2, \dots, x_n)$  and  $Y = (y_1, y_2, \dots, y_n)$ , 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. Let

$$H_0 : F_X = F_Y \quad \text{against} \quad H_1 : F_X \neq F_Y \quad \text{or} \\ H_1 : F_X < F_Y \quad \text{or} \quad H_1 : F_X > F_Y$$

For  $k$  distinct values of the test statistic  $T$ , the probability distribution of the test statistic  $T = (T_1, T_2, \dots, T_k)$  under the null hypothesis  $H_0 : F_X = F_Y$  is given by

$$P(T_j = t_0 | H_0) = \sum_{i=1}^{f_j} \binom{2^{-n}}{i} = f_j \binom{2^{-n}}{1}$$

where  $f_j$  is the number of occurrences of  $T_j$ . For specified value of  $n$  and the level of significance  $\alpha$ , the critical value  $c$  corresponds to a level closest to  $\alpha$ . 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

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

and for the right tail,

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

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

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

Let  $\pi_1, \pi_2, \dots, \pi_n$  be a set of all distinct permutations of the data set such that  $\pi_i$  is the  $i$ th permutation. The permutation test procedure is defined as follows:

1. Compute the Test Statistic  $T_i$  for the original arrangement  $\pi_1$
2. Obtain a distinct permutation  $\pi_i$
3. Compute the Test Statistic for permutation  $\pi_i, T_i = T(\pi_i)$
4. Repeat Steps 2 and 3 for  $i = 2, 3, \dots, 2^n, n = \text{sample size}$
5. Construct an empirical cumulative distribution

$$P_0 = \hat{p}(T \leq T_i) = \frac{1}{2^n} \sum_{i=1}^{2^n} \theta(T_i - T_i)$$

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

6. Under the empirical distribution  $\hat{p}$  if  $p_0 \leq \alpha$ , reject the null hypothesis.

The procedure computes the cumulative distribution of the Test Statistic,  $T$ , under the null hypothesis.

**Implementation:** A discussion of a systematic way of obtaining all the possible permutations of the  $N$  variates now follows. Given a balanced two-sample layout as

$$\begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{pmatrix}, \text{ where } x_i \in X \text{ and } y_j \in Y. \text{ Let } n = 4, \text{ an}$$

illustrative implementation of paired permutation algorithm now follows:

Original arrangement of paired observations

$$\begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \\ x_4 & y_4 \end{pmatrix} \text{ yields } \binom{4}{0} \text{ permutation.}$$

Exchange of one pair of observations

$x_i \leftrightarrow y_i$  yields  $\binom{n}{1}$  permutations, that is:

$$\begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \\ y_4 & x_4 \end{pmatrix}, \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ y_3 & x_3 \\ x_4 & y_4 \end{pmatrix}, \begin{pmatrix} x_1 & y_1 \\ y_2 & x_2 \\ x_3 & y_3 \\ x_4 & y_4 \end{pmatrix}, \begin{pmatrix} y_1 & x_1 \\ x_2 & y_2 \\ x_3 & y_3 \\ x_4 & y_4 \end{pmatrix}$$

or  $\binom{4}{1}$  permutations.

Exchange of two pairs of observations

$\begin{pmatrix} x_i \\ x_j \end{pmatrix} \leftrightarrow \begin{pmatrix} y_i \\ y_j \end{pmatrix}; i \neq j$  yields  $\binom{n}{2}$  permutations, that

is:

$$\begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ y_3 & x_3 \\ y_4 & x_4 \end{pmatrix}, \begin{pmatrix} x_1 & y_1 \\ y_2 & x_2 \\ x_3 & y_3 \\ y_4 & x_4 \end{pmatrix}, \begin{pmatrix} y_1 & x_1 \\ x_2 & y_2 \\ x_3 & y_3 \\ y_4 & x_4 \end{pmatrix}, \begin{pmatrix} x_1 & y_1 \\ y_2 & x_2 \\ y_3 & x_3 \\ x_4 & y_4 \end{pmatrix},$$

$$\begin{pmatrix} y_1 & x_1 \\ x_2 & y_2 \\ y_3 & x_3 \\ x_4 & y_4 \end{pmatrix}, \begin{pmatrix} y_1 & x_1 \\ y_2 & x_2 \\ x_3 & y_3 \\ x_4 & y_4 \end{pmatrix} \text{ or } \binom{4}{2} \text{ permutations.}$$

Exchange of three pairs of observations

$\begin{pmatrix} x_i \\ x_j \\ x_k \end{pmatrix} \leftrightarrow \begin{pmatrix} y_i \\ y_j \\ y_k \end{pmatrix}; i \neq j \neq k$  yields  $\binom{n}{3}$  permutations,

that is:

$$\begin{pmatrix} x_1 & y_1 \\ y_2 & x_2 \\ y_3 & x_3 \\ y_4 & x_4 \end{pmatrix}, \begin{pmatrix} y_1 & x_1 \\ x_2 & y_2 \\ y_3 & x_3 \\ y_4 & x_4 \end{pmatrix}, \begin{pmatrix} y_1 & x_1 \\ y_2 & x_2 \\ x_3 & y_3 \\ y_4 & x_4 \end{pmatrix}, \begin{pmatrix} y_1 & x_1 \\ y_2 & x_2 \\ y_3 & x_3 \\ x_4 & y_4 \end{pmatrix}$$

or  $\binom{4}{3}$  Permutation.

Exchange of four pairs of observations

$$\begin{pmatrix} x_i \\ x_j \\ x_k \\ x_l \end{pmatrix} \leftrightarrow \begin{pmatrix} y_i \\ y_j \\ y_k \\ y_l \end{pmatrix}; i \neq j \neq k \neq l \text{ yields } \binom{n}{4} \text{ permutations.}$$

$$\begin{pmatrix} y_1 & x_1 \\ y_2 & x_2 \\ y_3 & x_3 \\ y_4 & x_4 \end{pmatrix} \text{ or } \binom{4}{4} \text{ permutation.}$$

This process is continued until the desired sample size is attained.

$$\text{Total (for 4-paired observations)} = \binom{4}{0} + \binom{4}{1} + \binom{4}{2} +$$

$$\binom{4}{3} + \binom{4}{4} = 16$$

Clearly, for a two-sample problem, the number of permutations is  $\sum_{i=0}^n \binom{n}{i}$ . This is the form required for a

complete enumeration of all the paired permutations of an experiment. Note that  $\sum_{i=0}^n \binom{n}{i} = 2^n$ .

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, bearing in mind that all the permutations are equally likely.

**Paired permutation algorithm:** The algorithm *Paired-Permutation* presented in this study can carry out a complete enumeration of all the possible distinct n-paired permutations by making the necessary adjustments to the algorithm to reflect the number of pairs. The algorithm is illustrated for  $\bar{v}$ e pairs of observations. Let  $x_{ij}$  represent the paired observations in an experiment, for  $i = 1(1)5; j = 1; 2$ .

**Algorithm** (Paired-Permutation)

- 1: **for**  $i_1 \leftarrow 1, 2$  **do**
- 2:  $T_1 \leftarrow X_{1,i_1}$
- 3: **if**  $i_1 \leftarrow 1$  **then**
- 4:  $S_1 \leftarrow X_{1,2}$
- 5: **else**

- 6:  $S_1 \leftarrow X_{1,1}$
- 7: **end if**
- 8: **for**  $i_2 \leftarrow 1, 2$  **do**
- 9:  $T_2 \leftarrow X_{2,i_2}$
- 10: **if**  $i_2 \leftarrow 1$  **then**
- 11:  $S_2 \leftarrow X_{2,2}$
- 12: **else**
- 13:  $S_2 \leftarrow X_{2,1}$
- 14: **end if**
- 15: **for**  $i_3 \leftarrow 1, 2$  **do**
- 16:  $T_3 \leftarrow X_{3,i_3}$
- 17: **if**  $i_3 \leftarrow 1$  **then**
- 18:  $S_3 \leftarrow X_{3,2}$
- 19: **else**
- 20:  $S_3 \leftarrow X_{3,1}$
- 21: **end if**
- 22: **for**  $i_4 \leftarrow 1, 2$  **do**
- 23:  $T_4 \leftarrow X_{4,i_4}$
- 24: **if**  $i_4 \leftarrow 1$  **then**
- 25:  $S_4 \leftarrow X_{4,2}$
- 26: **else**
- 27:  $S_4 \leftarrow X_{4,1}$
- 28: **end if**
- 29: **for**  $i_5 \leftarrow 1, 2$  **do**
- 30:  $T_5 \leftarrow X_{5,i_5}$
- 31: **if**  $i_5 \leftarrow 1$  **then**
- 32:  $S_5 \leftarrow X_{5,2}$
- 33: **else**
- 34:  $S_5 \leftarrow X_{5,1}$
- 35: **end if**
- 36: ...
- 37: **for**  $i_{15} \leftarrow 1, 2$  **do**
- 38:  $T_{15} \leftarrow X_{15,i_{15}}$
- 39: **if**  $i_{15} \leftarrow 1$  **then**
- 40:  $S_{15} \leftarrow X_{15,2}$
- 41: **else**

```

42:           S15 ← X15,1
43:           end if
44:           Compute Test Statistic
45:         end for
46:       end for
47:     end for
48:   end for
49: end for
50: ...
51: end for
    
```

**DISCUSSION**

R. A. Fisher compiled by hand 32,768 permutations of Charles Darwin’s data. It is believed that the enormity of this task possibly discouraged Fisher from further research into exact permutation tests, see Ludbrook and Dudley<sup>[9]</sup>. Clearly, our results agree with the results Fisher manually obtained as the number of permutations that were equal or more extreme than the observed difference of means. Observe that at 5% level of significance, the null hypothesis of no significant difference is rejected under the exact permutation distribution while the null hypothesis is accepted under the paired t-distribution. Permutation test can therefore serve as an independent check on the classical methods in common use, which agrees with the observation of Fisher<sup>[1]</sup>.

The problem with permutation tests has been high computational demands, viz; space and time complexities. Available permutation procedures can sample from the permutation sample space rather than carrying out complete enumeration of all possible distinct permutations. These available procedures cannot avoid the possibility of drawing the same sample more than once, thereby reducing the power of the permutation test, see Opdyke<sup>[5]</sup>.

This study formulates and implements a sure way of obtaining exact permutation distribution of paired observations by ensuring that a complete enumeration of all the distinct permutations is achieved. This produces exact p-values and ensures that the probability of a type I error is exactly  $\alpha$ . 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.

**RESULTS**

The Algorithm (Paired-Permutation) was implemented in Intel Visual FORTRAN to obtain the permutation distribution of the 15 pairs of values on heights of cross-fertilized and self-fertilized plants. Looking through the Appendix, (863 + 863) permutations are equal or more extreme than the observed difference of means (2.61667). The one-sided p-value is simply  $863/32768 = 0.02634$  and the two-sided p-value is  $(863 + 863)/32768 = 0.05267$ . The following results were obtained (Table 1).

The exact permutation distribution of the difference of means for paired observations along with their p-values is in the Appendix, while the graphical display is shown in Fig. 1.

Table 1: P-values for height of cross and self fertilized plants

	Test Statistic	P(T ≤ t) one-tailed p-value	P(T ≤ t) two-tailed p-value
t-test	2.14799	0.02485	0.04970
Paired Permutation	2.61667*	0.02634	0.05267

\* Actual difference of means

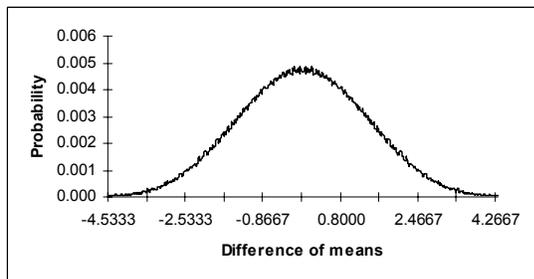


Fig. 1: Exact permutation distribution of height of cross and self fertilized plants

**CONCLUSION**

The p-value obtained through the permutation approach is exact, see Agresti<sup>[11]</sup> and Good<sup>[13]</sup>. Obtaining exact p-values through exact permutation has remained difficult because it is computationally intensive. For small sample sizes, the exact permutation distribution of a test statistic and its asymptotic equivalent can be quite discrepant.

A straight forward approach has been adopted in obtaining exact distribution of the paired observations

in an experiment through a careful enumeration of distinct permutations of the observations. The permutation algorithm presented in this study beats the limitations and difficulties inherent in the exact permutation approach that probably led to the introduction of other approximate methods, which do not truly provide the exact distribution of a test statistic.

#### REFERENCES

1. Fisher, R. A. 1936. The coefficient of racial likeness and the future of Craniometry. *J. Royal Anthropological Soc.*, 66: 57-63.
2. Scheffe, H., 1943. Statistical inference in the nonparametric case. *The Annals of Math. Stat.* 14: 305-332.
3. Efron, B. and R.J. Tibshirani, 1993. An introduction to the bootstrap. Chapman and Hall, New York.
4. Hall, P. and I. Weissman, 1997. On the estimation of extreme tail probabilities. *The Annals of Statistics*, 25: 1311-1326.
5. Opdyke, J. D. 2003. Fast permutation tests that maximize power under conventional Monte Carlo sampling for pairwise and multiple comparisons. *J. Modern Appl. Stat. Methods*, 2: 27-49.
6. Bayarri, M. J. and J.O. Berger, 2004. The interplay of Bayesian and frequentist analysis. *Stat.Sci.*,19: 58-80.
7. Spiegelhalter, D. J., 2004. Incorporating Bayesian ideas into health-care evaluation. *Stat.Sci.*, 19: 156-174.
8. Darwin, C., 1878. *The Effects of Cross and self-fertilization in the Vegetable Kingdom*, 2nd Edn., London: John Murray.
9. 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.
10. Siegel, S. and N.J. Castellan, 1988. *Nonparametric statistics for the behavioral sciences* (2nd Edn.). McGraw-Hill, New York.
11. Agresti, A., 1992. A survey of exact inference for contingency tables, *Stat. Sci.*, 7: 131-177.
12. Odiase, J. I. and S.M. Ogbonmwan, 2005. An algorithm for generating unconditional exact permutation distribution for a two-sample experiment. *J. Modern Appl. Stat. Methods*, 4: 319-332.
13. Good, P., 2000. *Permutation tests: a practical guide to resampling methods for testing hypotheses* (2nd Edn.). Springer Verlag, New York.