THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

# ALWAYS USABLE LIMITED-LENGTH SEQUENTIAL PERMUTATION TESTS FOR TWO-WAY ANOVA

by

John E. Walsh

Technical Report No. 84
Department of Statistics THEMIS Contract

October 1, 1970

Research sponsored by the Office of Naval Research
Contract N00014-68-A-0515
Project NR 042-260



DEPARTMENT OF STATISTICS
Southern Methodist University

# ALWAYS USABLE LIMITED-LENGTH SEQUENTIAL PERMUTATION TESTS FOR TWO-WAY ANOVA

John E. Walsh*

Southern Methodist University**

ABSTRACT

The data are independent observations from two or more sources. Under the null hypothesis, observations from the same source have a common unknown distribution (can be arbitrary). Observations are obtained in successive groups each containing a set of specified size from each source, with a stated maximum number of groups available. An overall test consists of a succession of subtests and is significant when at least one subtest is significant. Exact null probabilities are obtainable, through use of appropriate permutation models and special kinds of subtest statistics. The subtests are such that the significance level of each new subtest is independent of preceding subtest results. The overall test terminates when a significant subtest occurs (with saving in time and expense). Subtests are always included wherein the second and each following group is compared with the totality of previous groups to investigate whether observations from the same source continue to be from the same population. Ordinarily, a single comprehensive subtest occurs for each comparison. However, a separate subtest could occur for each source. Also, an additional subtest (or separate subtests for sources) could be made to investigate whether the first group satisfies the null hypothesis. Unconditional tests can be obtained when ranks are used and in some other cases. Some possible uses in quality control are outlined.

*Based on work performed at the Quality Evaluation Laboratory, U. S. Naval Torpedo Station, Keyport, Washington.

**Research partially supported by ONR Contract N00014-68-A-0515 and by Mobil Research and Development Corporation. Also associated with NASA Grant NGR 44-007-028.

## INTRODUCTION AND DISCUSSION

Sequential significance tests are considered for a kind of two-way analysis of variance (ANOVA) situation. The data are independent univariate observations that are taken from two or more sources. These observations are obtained in successive groups, where each group contains a set of specified size from each source. The number of groups that can be available is limited. Under the null hypothesis, observations obtained from the same source have the same unknown distribution, which can be arbitrary. The null distribution for a source does not necessarily bear any relation to that for any other source.

The overall test consists of a succession of subtests. A subtest occurs for each new group of observations (after the first group, and maybe also for this group). The null hypothesis for each of these subtests asserts that, for every source, the new observations from a source and the previous observations from that source are all from the same unknown distribution. In addition, if desired, an initial subtest can be included to investigate whether, for every source, the observations in the first group from that source constitute a random sample. An overall test is significant if and only if at least one of the subtests is significant. The obtaining of groups can be stopped at the first subtest that is significant (with a saving in time and/or expense). Significance fails to occur for an overall test if and only if the maximum number of groups is obtained without significance for any subtest.

The development of subtests that use all data through the sequential step under consideration seems very desirable. However, use of the data

from all previous groups can complicate determination of the significance level for the overall test, and for the second and following subtests. That is, the conditional effect of results for preceding subtests needs to be taken into consideration. These significance levels are most readily evaluated when the development is such that the significance level for each subtest is independent of the outcomes for preceding subtests. This can be accomplished by use of a suitable kind of subtest statistic and use of appropriate permutation models for the null case.

This permutation approach has the extra advantage of yielding overall tests that are always applicable. In addition, the suitable subtest statistics include types that are appropriate for investigating whether, for each source, observations of a new group continue to be from the same population as that yielding observations from this source for the previous groups. The suitable subtests can have broad ranges of significance levels and can emphasize many kinds of alternative hypotheses.

The data for a subtest are the new group and all previous groups (if any). For the null case, and each source, the possible values for the observations from that source are conditionally fixed at the values that occurred. Probability considerations arise only in random association of the possible values with the observations. These associations can be expressed as permutations of the possible values from a source in positions of a sequence (number of positions equals the number of observations). For definiteness, and convenience, the sequence positions can represent the order in which the observations from the source are obtained. A unique sequence of positions is established, with an arbitrary choice of sequence

However, for the other data fixed, the subtest statistic has the same value for every permutation of this totality of previous observations. The outcome for a subtest corresponds directly to the permutation that randomly occurs. Thus, the significance level for a subtest with this kind of statistic is not influenced by the results for previous subtests.

In general, the subtests are conditional since, for each source, the possible values of observations are fixed at those which occurred. However, the observation values considered to be possible need not be conditionally fixed for subtests that are rank tests. Such subtests are unconditional if the data are continuous or independent randomization is used to break ties (approximately unconditional if some other method, such as midranks, is used to break ties). Also, unconditional subtests that apply when the permutation model holds are obtained by constructing independent statistics whose distributions are symmetrical about zero under the null hypothesis. A subtest is based on order statistics of these statistics and often involves extreme values.

The sequential tests considered are useful for situations where a change in distribution may occur for one or more sources and rapid identification of when this has happened is desired. They are especially useful when also very little is known about the distributions involved.

Quality control represents an application area. The limited number of steps for an overall test does not prohibit its use for quality control. Many sequential tests for quality control actually have a limited number of steps. That is, the tests for different steps are independent (nonoverlapping data) and have the same significance level. Overall, a

## OUTLINE OF TESTS

The observations are univariate and independent. They are obtained in consecutive groups whose compositions can be different. The notation used is

$G$ = maximum number of groups that are obtainable

$i$ = designation index for i-th group that is obtained $(i=1,\ldots,G)$

$M$ = number of sources $(M \geq 2)$

$j$ = designation index for j-th source $(j=1,\ldots,M)$

$n_{ij}$ = number of observations from source j that are in the i-th group. The value of $n_{ij}$ is nonzero and ordinarily at least 3 or 4

$N_{ij} = \sum_{k=1}^{i-1} n_{kj}$ = total number of observations in groups $k=1,\ldots,i-1$ that are from the j-th source $(i \geq 2)$, with $N_{1j} = 0$

$S_i$ = statistic for the subtest where the i-th group of observations is first used.

$T_i$ = subtest that is based on $S_i$

$\alpha_i$ = significance level for $T_i$, $(0 < \alpha_i < 1)$

In $S_i$, for $i \geq 2$, the $N_{ij}$ observations from source j occur symmetrically. That is, for the other data fixed, the value of $S_i$ is the same for all possible assignments of identities (sequence positions) to these $N_{ij}$ observations.

Let us consider the permutation models that are used for the various values of i and j. For given i, the same kind of model is used for each value of j. Thus, consideration of permutation models that occur for an

arbitrary but fixed value of j, and all values of i, is sufficient.

For given j, consider the available data when the i-th group is the new group. These data are the $N_{ij}$ previous observations and the $n_{ij}$ new observations. The possible values for this totality of $N_{ij} + n_{ij}$ observations are conditionally fixed at the observed values. Probabilistic aspects enter only in the random assignment of identities (as observations) to these $N_{ij} + n_{ij}$ values, which is equivalent to randomly assigning these values to positions in a sequence of $N_{ij} + n_{ij}$ positions. Under the null hypothesis, all $(N_{ij} + n_{ij})!$ ways of making such an assignment are equally likely.

For each i, the overall permutation model consists of a combination of separate and independent use of the permutation models for j=1,...,M. That is, there are

$$W_i = \prod_{j=1}^{M} (N_{ij} + n_{ij})!$$

possible ways of assigning the totality of observations to sequence positions. A value of $S_i$, not necessarily unique, occurs for each of the $W_i$ ways.

Now consider general development of exact subtests by use of $S_i$ and this permutation model. Let the $W_i$ values for $S_i$ be ordered according to increasing value (arbitrary ordering within a set of tied values) and consider the location in this ordering of the value actually observed for $S_i$. The subtest $T_i$ is one-sided upper-tail when significance occurs if and only if the observed $S_i$ equals or is less than at most $\alpha_i W_i$ of the values in this ordering. $T_i$ is one-sided lower-tail when significance

$$W_i' = \prod_{j=1}^{M} (n_{ij} + N_{ij})!/n_{ij}!N_{ij}!$$

is the possible number of divisions (over all sources). For this case, $S_i$ is such that it has the same value for all permutations that yield the same division. The method of constructing subtests is the same as when $W_i$ ways are considered. That is, the $W_i'$ divisions yield $W_i'$ values for $S_i$ (not necessarily unique) and these are ordered according to increasing value, etc.

These exact subtests are, in general, of a conditional nature. This is due to the conditional fixing, for each source, of the totality of observations at the values which occurred. However, this permutation procedure yields unconditional exact subtests for cases where $S_i$ can be based exclusively on ranks and ties among ranks occur with zero probability (for example, the data are continuous or ties are eliminated by randomization). It should be noted that some statistics which may not appear to be based entirely on ranks can be stated in that form. Unconditional exact subtests can also occur under other circumstances, and a case of this nature (involving construction of statistics whose null distributions are symmetrical about zero) is considered in the next section.

Application of an exact subtest can involve an excessive amount of effort. Exceptions are use of suitably tabulated rank tests, use of $S_i$ based on statistics whose null distributions are symmetrical about zero, and cases where the number of possible permutations, or divisions, is not extremely large. Also, breaking of ties in ranks by randomization can require moderate extra effort and sometimes may be considered an undesirable

procedure. Consequently, subtests with approximately determined significance levels are often used. A subtest is always approximate when an approximate method is used to evaluate its significance level. For example, a significance level may be based on the first few terms of an expansion and only usable when the $n_{ij}$ and/or $N_{ij}$ are sufficiently large (depending on the size of the significance level). Also, a subtest based on ranks is approximate when ties among ranks are broken by an averaging process, such as use of midranks, but the significance level is determined by assuming that ties do not occur.

A subtest that directly uses the observations (without conversion to ranks, etc.) can be simultaneously approximate in two ways. That is, it is an approximate permutation test and also an approximately unconditional test. The terminology "robust" is used for such tests (by Box and Andersen in ref. 1).

Another way of conducting subtest $T_i$ is to do a separate testing for each source. Here, $T_i$ is significant if and only if significance occurs for at least one source. A statistic $S_i$ can be developed for $T_i$ but separate consideration of each testing for a source is more convenient. Thus, at each sequential step, the two-way ANOVA situation is converted into M separate and independent one-way ANOVA situations. Effectively, the choice of a test statistic for each of these one-way ANOVA situations is the same as if an overall one-way ANOVA were being conducted using the data from the source considered. Methods of selecting statistics for limited-length sequential permutation tests in one-way ANOVA are given in ref. 2.

The major new consideration is that the testing for each source must have a very small significance level if $\alpha_i$ is to be reasonably small. The principal considerations in conducting $T_i$ by a separate testing for each source are, effectively, covered by the material of ref. 2. Consequently, not much is given concerning properties of this kind of subtest.

For one type of overall test, significance occurs if and only if at least one of $T_2,\ldots,T_G$ is significant. The overall test has

$$\alpha = 1 - \prod_{i=2}^{G} (1 - \alpha_i)$$

for its significance level, since the properties of $S_i$ imply that a subtest is independent of the outcomes for all previous subtests. If some or all of the subtests are approximate, the value of $\alpha$ is approximate.

For the other type, where the random sample hypotheses (for each source) is investigated for the first group, significance occurs for the overall test if and only if at least one of $T_1,\ldots,T_G$ is significant. This overall test has significance level

$$\alpha = 1 - \prod_{i=1}^{G} (1 - \alpha_i),$$

which is approximate if at least one of $\alpha_1,\ldots,\alpha_G$ is approximate.

## SELECTION OF STATISTICS AND SUBTESTS

The selection and use of the $S_i$ is influenced by many considerations besides the requirement imposed on use of the observations from previous groups. One important consideration is the alternative hypotheses that

are emphasized. Also, restrictions on the sizes of G and the $n_{ij}$ can be important when a small size is desired for $\alpha$, and/or nearly equal values are desired for the $\alpha_i$, and/or approximate subtests occur. Moreover, subtests that are not conditional can be desired. In addition, when several forms of a statistic yield an equivalent subtest, the least comlicated form is ordinarily used for exact subtests but the form with the most convenient type of null distribution is used for approximate subtests.

The case where a separate testing occurs for each source is examined first. Let $\alpha_{ij}$ be the significance level for the testing of the j-th source in the i-th group. Since the observations are independent.

$$\alpha_i = 1 - \prod_{j=1}^{M} (1 - \alpha_{ij}) .$$

If a small value is desired for $\alpha_i$, the $\alpha_{ij}$ must all be very small. However, for i = 1, the value of $\alpha_{1j}$ is at least $1/n_{1j}!$ for a one-sided testing and at least $2/n_{1j}!$ for a two-sided testing. For $i \geq 2$, the value of $\alpha_{ij}$ is at least $n_{ij}!N_{ij}!/(n_{ij} + N_{ij})!$ for a one-sided testing and at least double this value for a two-sided testing. These lower bounds are all sharp. This implies that the $n_{1j}$, $n_{2j}$, and perhaps the $n_{3j}$ are of at least moderate size, especially if M is not small. Other considerations for the case of a separate testing for each source are discussed in ref. 2.

The remainder of this section is concerned with subtests where separate testings do not occur for the sources. For some such subtests, a value of $\alpha_i$ as small as $1/W_i$ can occur for one-sided subtests, and as small as $2/W_i$ for two-sided subtests. Often, however, the smallest possible value of

$\alpha_i$ for $i \geq 2$ is at least $1/W_i'$ for one-sided tests and at least $2/W_i'$ for two-sided tests.

When M is at least 3 or 4 and $i \geq 2$, the smallest possible values for the $\alpha_i$ are frequently small enough for applications, even when the $n_{ij}$ are as small as 4 or 5. At least moderate sized values may be needed for the $n_{ij}$ when $i = 1$, if $\alpha_1$ is to be sufficiently small.

For given $\alpha$, the allowable sizes of the $\alpha_i$ tend to decrease as G increases. Alternately, for given $\alpha$ and desired sizes for the $\alpha_i$, the allowable values for G have an upper limit. However, G can be very large when the $\alpha_i$ are all very small.

Suppose that the $n_{ij}$ for fixed j are required to be equal, the $n_{ij}$ are all required to be small, and nearly equal small values are desired for the $\alpha_i$. Then, a compromise may be needed in which $\alpha_1$, and perhaps $\alpha_2$, are larger than the other $\alpha_i$ (which are very nearly equal). A similar situation is that where the $\alpha_i$ are required to be small and very nearly equal, small values are desired for the $n_{ij}$, and, for fixed j, equal values are desired for the $n_{ij}$. Here, a compromise may be needed in which, for fixed j, the value of $n_{1j}$ is much larger than the values of the $n_{ij}$ for $i \geq 2$ (which can be equal or nearly equal).

Next, consider the alternative hypotheses emphasized. Separate consideration of each source can be helpful. Often, whether the expected values of the observations from a source tend to be nondecreasing from group to group, with increase for the new group, is of interest. Also, whether the expected values tend to be nonincreasing, with decrease for the new group, can be of interest.

Most of the kinds of subtests considered emphasize alternative hypotheses where the groups can be arranged so that, simultaneously for every source, the expected values follow a trend in the same direction (with a change in expected value for at least one group of one source). Consequently, coordination in data use should occur with respect to the direction of the trends considered for the sources. As an example, suppose that nondecreasing expected values are of interest for some sources while nonincreasing expected values are of interest for the other sources. Modification of the data for these other sources through multiplication by -1 yield a data situation where nondecreasing expected values are of interest for all sources.

As another example, suppose that two types of alternatives are of interest. For one type, the combination of nondecreasing expected values for the sources of a given set and nonincreasing expected values for the other sources is of interest. For the other type, the combination of nonincreasing expected values for the given set and nondecreasing expected values for the other sources is of interest. Then, multiplication of the observations for the given set by -1 yields data where, for the two alternatives of interest, the groups can be arranged so that the expected values follow a trend in the same direction for all sources.

Now, consider choice of $S_i$. Interest is in use of results that are already developed. A correspondence to two-way ANOVA results that are expressed in terms of rows and columns can be obtained by using the sources to represent "rows" and the groups to represent "columns." Then, under the

null hypothesis, all the observations for a row have the same distribution (no column effects occur). Alternately, consider use of results for randomized blocks with a single kind of treatment. The sources can be represented as the "blocks" factor and the groups as the "treatment" factor. The treatment levels are equivalent under the null hypothesis. That is, within each block, the null distribution of the observations is the same for all the levels of the treatment.

Some $S_i$ require all the observations to be expressed in the same unit. When this restriction is not satisfied, these $S_i$ are eliminated from possible use. Cases often occur where the type of unit for one source is fundamentally different from the type for another source (distance and time, etc.).

A large number of statistics that could be $S_i$, with various properties and uses, can be developed. In particular, statistics using extreme observations can be obtained. The permutation basis, in principle, allows the development of tests that could use almost any kind of statistic.

Actually, for the case where the sources are not investigated separately, the kinds of $S_i$ for which practically usable results have been developed are somewhat limited. Also, some of these $S_i$ require that all the observations be expressed in the same unit. Moreover, some $S_i$ yield unconditional subtests while others do not. In addition, some $S_i$ directly allow for replication (such as occurs within each group for every source) while others do not. Also, some $S_i$ impose conditions on the $n_{ij}$, such as $n_{ij}$ has the same value for all j (possibly different values of $n_{ij}$ for different i).

A statistic can be useful for a subtest even though it does not directly allow for the replication that occurs within each group or the replication that, in most cases, is considered to occur for the previous data. Here, for fixed i, the $n_{ij}$ are the same for all j. The procedure is to coordinate the observations from the different sources, say, accord- to the sequence order in which they are obtained. This can be considered to divide the observations into small sets ("groups") such that each set contains one observation from every source. That is, the first set consists of the first observations from the sources, the second set consists of the second observations, etc. These sets, each of size M, are then used in the same manner as if they were the groups considered. For cases where $S_i$ is expressed in terms of rows and columns, the sources represent rows and the sets represent columns. For cases where $S_i$ is expressed in terms of blocks and a treatment, the sources represent blocks and the sets represent levels of the treatment.

## SPECIFIC STATISTICS AND SUBTESTS

Several kinds of useful subtests are identified in this section. Most of these subtests are approximate (at least, in the way they are applied). Additional conditions are imposed on the $n_{ij}$ for the approximate subtests.

Subtests based on some robust results for the randomized block design with one treatment are considered first. These results are approximate and require all the observations to be expressed in the same unit. Also, for fixed i, the $n_{ij}$ are required to be equal. Subtests which do not directly allow for replication can be obtained from the material of Box

and Andersen in ref. 1. These results, along with conditions on the $n_{ij}$ for their use, are also given on pages 324-325 of ref. 3. Subtests that do directly allow for replication, but use some specialized conditions, can be obtained from the material of Wilk in ref. 4. These results, including conditions on the $n_{ij}$, are also stated on pages 325-326 of ref. 3. Both of the kinds of subtests considered here are approximately unconditional.

Subtests based on some unconditional results for the randomized block design with one treatment are considered next. These results are applicable under the permutation model but are of an unconditional nature. For meaningful interpretation, most of these subtests require all the observations to be expressed in the same unit. However, some results do not require the same unit for observations from any different sources. Subtests can be obtained from the results of ref. 5 for the case where all observations used are involved in the null hypothesis. Subtests can also be obtained from the material on pages 383-385 of ref. 3, for Case (I) and the situation where all the $e'''$ are zero ; also, for fixed i, the $n_{ij}$ are equal. Most of these subtests are of an exact nature. The basis for a subtest is the suitable formation of a function of the observations separately for each source. These functions are independent and, under the null hypothesis, have distributions that are symmetrical about zero. Use of an appropriate test for symmetry about zero yields a subtest. Ordinarily the number of sources should be at least four.

Specifically, order statistics of the functions of observations are used for the subtests. All observations must be expressed in the same

unit for the case where one or more sums of two different order statistics occur in the subtest statistic. No requirement about the same unit for different sources arises for the other case, where a one-sided subtest based on a single order statistic is considered, or where a two-sided test based on two order statistics (that are not summed) is considered. Use of values of the functions as if they were all expressed in the same unit yields a valid and meaningful subtest for this case, even though the function values may occur in two or more different units. This elimination of conditions on units is possible because the outcome for a subtest of this nature is completely determined by the signs of the values for the functions of observations. That is, use of single order statistic for a one-sided test and of two order statistics for a two-sided test is equivalent to use of the signs of the functions yielding the order statistics.

Now, consider subtests that are directly based on ranks. These results are of an unconditional nature. The observations from any two different sources are not required to be expressed in the same unit. Exact and approximate results both occur. The exact subtests considered do not directly allow for replication and require the $n_{ij}$ to have the same value for fixed i. Approximate subtests occur that directly allow for replication and that do not require equal values of the $n_{ij}$ for fixed i. Chapter 11 of ref. 3 contains a number of rank tests, plus discussion. Only subtests with fixed factors are of interest.

The Friedman-Kendall-Smith rank tests, which are exact or approximate and do not directly allow for replication, are stated on pages 412-416 of ref. 3. The Ehrenberg-Terpstra-Kendall-Smith rank tests are given on

pages 416-419. One type of approximate test does not directly allow for replication but the other type does (moreover, for fixed i, the $n_{ij}$ are not required to be equal). The most general results are those of Benard and van Elteren (see, for example, pages 427-428 of ref. 3). Some conditions on the $n_{ij}$ for use of the various rank tests are given in the statements of results in Chapter 11 of ref. 3.

Finally, consider some subtests that have median ANOVA as a basis. These results are of an unconditional nature, since they can be obtained using ranks, and observations from two different sources need not be expressed in the same unit. As applied, these subtests are approximate. Direct allowance is made for replication and, for fixed i, the $n_{ij}$ are required to be equal. These subtests are given on pages 553-555 of ref. 3, with interest in the case of both factors fixed. Included are conditions on the $n_{ij}$ for use of the approximations. Also, with suitable interpretation, other tests of ref. 3 that are of a categorical nature and with a permutation basis can often be used.

## USE FOR QUALITY CONTROL

Successive groups for quality control use, with observations from two or more sources, ordinarily have the properties: (1) All groups, except possibly the first group, have the same composition. (2) Within each group, the number of observations obtained from each source is the same. (3) The number of observations obtained from a source is small for

every group except possibly the first group. Now, consider the successive quality control tests. These tests have the property: (4) The significance levels are equal (or almost equal) and very small.

Suitable overall permutation tests with subtests that satisfy (4) and also apply to groups of this nature nearly always can be developed. Some development considerations that apply to the case where a separate testing occurs for each source are outlined in ref. 2. An additional consideration, which can be a difficulty for this case, is the way the significance level for a subtest depends on the significance levels for the separate testings. Except for remarks about group composition, the remainder of the discussion here is devoted to considerations for the other case (where subtests are not based on separate testings for the sources).

On the basis of properties (1) - (3), the $n_{1j}$ are equal and also, for $i \geq 2$, the $n_{ij}$ are all equal and their common value (denoted by, say, $n_{2j}$) is small. The $n_{1j}$ are not necessarily small but, usually, have the smallest value such that (4) is satisfied. For a given total number of observations from each source, substantial information can be lost by using too large a value for the $n_{1j}$. Ordinarily, subject to satisfying (1) - (4), the $n_{1j}$ should be as small as possible and $n_{2j}$ should be as large as possible. An exception is when suitable past data are available that could be used for the first group. Then, $n_{1j}$ should be as large as possible.

The permutation tests, which are generally applicable, are most appropriate when little is already known about the population yielding the observations. This is often the case for quality control situations.

Of course, as more groups are obtained, the amount of information increases. Thus, the subtests tend to become more efficient, although a plateau should be reached moderately soon.

Frequently, not all of the observations for a quality control situation are expressible in a common unit. Then, subtests that require a common unit are not usable. In some respects, the results directly based on ranks seem to be most suitable for quality control use, although other results might be preferable in special cases.

Restrictions on the $n_{ij}$ for use of approximate subtests sometimes are not in agreement with properties (1) - (4). However, this could happen for small i and not be the case for larger i. Moreover, appropriate subdivision of the previous observations, and perhaps also the observations of the new group, nearly always will result in usable subtests for cases where nothing suitable is available for direct use. This subdivision would be of the same nature as that outlined for use of results that do not directly allow for replication.

## REFERENCES


1. G. E. P. Box and S. L. Andersen, "Permutation theory in the derivation of robust criteria and study of departures from assumptions," Journal of the Royal Statistical Society, Series B, Vol. 17 (1955), pp. 1-34.

2. John E. Walsh, Generally Applicable Limited-Length Sequential Permutation Tests for One-Way ANOVA, Themis report 80, Statistics Department, Southern Methodist University, September 1970, 20 pp. Submitted for publication in Australian Journal of Statistics.

3. John E. Walsh, Handbook of Nonparametric Statistics, III: Analysis of Variance, D. Van Nostrand Co., Inc., Princeton, N. J., 1968, 771 pp.

4. M. B. Wilk, "The randomization analysis of a generalized randomized block design," Biometrika, Vol. 42 (1955), pp. 70-79.

5. John E. Walsh, "Exact nonparametric tests for randomized blocks," Annals of Mathematical Statistics, Vol. 30 (1959), pp. 1034-1040.