THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

GENERALLY USABLE SEQUENTIAL RANDOMIZATION TESTS FOR TWO-WAY

ANOVA THAT EMPHASIZE THE MORE RECENT DATA

by

John E. Walsh

Technical Report No. 85
Department of Statistics THEMIS Contract

October 1, 1970

DEPARTMENT OF STATISTICS
Southern Methodist University

GENERALLY USABLE SEQUENTIAL RANDOMIZATION TESTS FOR TWO-WAY ANOVA THAT EMPHASIZE THE MORE RECENT DATA

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 the same (arbitrary) distribution. Observations are obtained in successive groups each containing a set of specified size from each source (with a stated maximum total available from each source). An overall test is a succession of subtests with significance when at least one subtest is significant. All observations are re-used until a specified number of groups occur. Then, independently for each source, a stated number of the observations from a source are chosen by randomization (all possibilities equally likely). Data for re-use are now the group of observations chosen by the randomizations and newly obtained groups. Additional groups are taken until the number for re-use reaches a given value. Then, independently for each source, a stated number of the observations in these groups from a source are chosen by randomization. The resulting group is the data for re-use at this stage. Additional new groups are taken, etc. Exact null probabilities are obtainable, through use of appropriate randomization models and special kinds of subtest statistics. Subtests are such that significance levels of new subtests are independent of previous subtest results. The overall test ends when a significant subtest occurs (thus saving time and expense). Subtests are always included wherein the second and each following group is compared with the data for re-use, to investigate whether observations from the same source continue to be from the same population. Customarily, 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 null hypothesis holds for the first group. Some possible uses in quality control are outlined.

INTRODUCTION AND DISCUSSION

Significance tests of a sequential nature but with a limited number of steps are considered for two-way analysis of variance (ANOVA) situations of a special kind. Independent univariate observations are taken from two or more sources in successive groups. Each group contains a specified number of observations from each source and a stated maximum number of groups can be obtained. Observations from the same source have the same unknown distribution, which can be arbitrary, when the null hypothesis is satisfied. The null distributions for the sources do not necessarily have any relationship with each other.

An overall sequential test is conducted as a series of subtests, with significance if and only if at least one of the subtests is significant. A subtest occurs for testing each new group of observations (after the first, and perhaps for the first group, too). For every subtest, the null hypothesis asserts that new observations from a source continue to be from the same distribution as the previous observations from that source. If desired, an additional subtest, which would be the first subtest, can be included to investigate whether for every source the observations from that source are a random sample. The taking of groups can be terminated when the first significant subtest occurs (with a saving in time and cost). An overall test is not significant if and only if all the possible subtests are not significant.

The use of subtests that take into consideration the observations of all the previous groups seems very desirable. However, many situations

are such·that data from the more recent groups should receive more emphasis. This is the case when the pertinence of any specified group of previous observations decreases as the number of groups increases. The subtests that are developed emphasize the more recent data.

Determination of the significance levels for the subtests (and the overall test can be complicated by the consideration of the data used for preceding subtests. That is, allowance for the conditional effect of the results for preceding subtests is needed. Evaluation of these significance levels is perhaps easiest when the development has the property that re-sults for previous subtests impose no conditional effect on the signifi-cance level for a subtest. This property, and that of emphasizing the more recent data, can be attained by a combination of use of randomization for determining which prior data continue to be used, use of permutation models for the null case, and use of suitable types of subtest statistics.

This permutation-randomization method has the additional advantage of providing subtests that are always usable. Moreover, the suitable sub-test statistics include types that are appropriate for the kind of inves-tigation considered. The permissible subtests can emphasize many kinds of alternative hypotheses and can have a wide range of significance levels.

The observations from all prior groups are used in the subtests until a specified number of groups have been obtained and used in a subtest. Then, separately and independently for each source, a given number are chosen from the totality of observations from that source by randomization (all possible selections equally likely). Under the null hypothesis, each

of these·sets is a random sample from the distribution for the source considered. Also, since all observations are independent, these sets are mutually independent. For the next subtest, the previous data are the observations chosen by randomization. For the following subtest, the previous observations are those of the group obtained at the immediately preceding step and those chosen by randomization, etc. Separately for each source, the previous data (including that from new groups) constitute a random sample when the null hypothesis holds.

Following the randomization, new groups are taken until a specified number have been obtained and used in a subtest. Then, separately and independently for each source, a given number are chosen by randomization from the totality of available observations (those from the new groups and those obtained by the previous randomization). This procedure is repeated, with new randomizations, until the overall test terminates.

Usually, for every source, the number of observations selected by randomization increases with each randomization. However, this need not be the case.

The observations for a subtest are the new group and the available previous observations. For the null situation, and each source, the possible values for the available observations from that source are conditionally fixed at the observed values. Probability considerations enter only in random association of the possible values with the observations. For each source, such associations can be represented as permutations of the possible values in positions of a sequence, where the number of positions equals the number of observations.

4

To be definite, and for convenience, the order in which the observations are obtained can be used to establish the sequence positions. An arbitrary, but definite, sequence order is used for situations where some observations are obtained at the same time. When this method is used to establish sequence positions for a source, observations from the new group are represented by last positions in the sequence. All possible assignments of the values for a source to their sequence positions are equally likely under the null hypothesis. That is, separately and independently for each source, all possible permutations of the observed values in the sequence positions are equally likely when the null hypothesis holds.

Some subtests are such that every sequence position receives consideration for every source. At the other extreme, some subtests only consider a subdivision into positions that correspond to the new group and the set of the remaining positions. However, the type of permutation model used is applicable for any emphasis on sequence positions that may be adopted.

A way of obtaining a subtest statistic which is such that the subtest significance level is not influenced by the results for prior subtests is given next. The observations for a statistic are the new group and the available previous observations. All statistics having a special property are eligible for use. Specifically, separately for each source, the statistic is symmetrical in the totality of available previous observations. That is, for all the other observations fixed at any values, the statistic value is the same for all permutations of the sequence positions for the available previous observations from this source. A statistic

with this property for all sources is independent of the results for the preceding subtests (see the next section for proof).

The conditional fixing, for each source, of the possible values at those observed shows that, in general, the subtests are conditional. However, some statistics (and their use) are such that this conditional fixing of possible values is not required. This is the case when a statistic has the same distribution for all values that could occur for the available observations. This happens when a statistic is based exclusively on ranks of the observations and ties do not occur (randomization is used to break ties or the data are continuous). Other unconditional subtests, which apply under the permutation models, can be developed by constructing statistics that are independent and, under the null hypothesis, have distributions that are symmetrical about zero. A subtest is based on order statistics of these statistics and frequently uses extreme order statistics.

The sequential permutation tests are particularly useful for cases where very little is known about properties of the distributions yielding the observations. Their orientation is toward situations where a change of distribution may occur for one or more sources and rapid identification of this change is desired.

Quality control is an application area for these sequential tests, even though they have a limited number of steps. Actually, many sequential tests for quality control, such as quality control charts, have a limited number of steps. Otherwise, these sequential tests, which are

6

based on independent subtests with the same significance level, would have unit significance level. In practice, use of a very small significance level for the subtests tends to bypass this difficulty. The overall significance level for a large number of steps can then be of a magnitude that is ordinarily used for tests.

The same way of avoiding this difficulty can be used in quality control applications of the sequential permutation tests. More specifically, all subtests, except maybe the first one or two, have very small significance levels that are equal (or very nearly equal). The exception is due to the possibility that none of the possible significance levels for the first one or two subtests may have the desired small magnitude. That is, in quality control uses, the amount of data for a group ordinarily is small and all groups, except maybe the first, have the same composition.

Justification of the independence between an eligible statistic and the results for prior subtests is given in the next section. The following section contains a description of the sequential permutation tests (including the special case where, in a subtest, a separate testing occurs for each source and the subtest is significant if and only if at least one testing is significant). Considerations in the choice of statistics and subtests are discussed in the next following section. Some statistics and associated subtests which have occurred in the statistical literature are outlined in the next to last section. Extreme values are useful for some cases. The final section contains some comments about quality control uses.

INDEPENDENCE BETWEEN STATISTIC AND PRIOR SUBTESTS

An eligible statistic has the property that, separately for each source, it is symmetrical in the totality of available previous observations from that source. The justification of independence between such a statistic and the outcomes for the preceding subtests, when the permutation models are used, is divided into two cases that are comprehensive and mutually exclusive. Of course, available previous data refers to the observations occurring in previous groups that are actually used in the statistic.

The first case is that where none of the previous data was obtained by randomization selection (happens for the first groups). Then, for each source, a permutation that could occur for any prior subtest corresponds to subclass of the possible permutations for the values of the previous observations from this source. However, the statistic has the same value for all the permutations of all such subclasses. Hence, since the results for the prior subtests are determined by the observed permutations for these subtests, the statistic is independent of these results.

The other case is that where at least some observations of the available previous data were obtained by randomization selection. Then, the statistic is independent of outcomes for subtests that occurred before the last randomization used. This is verified by showing that, for each source, the available previous data are independent of the results for the subtests occurring before the last randomization used (since observations from different sources are independent and also the groups obtained after

the last randomization are independent of the groups obtained before then).
It is sufficient to verify the independence for subtests occurring before
the last randomization but after the next to last randomization (if any).
Induction, starting with the next to last randomization and working
backward, provides the remainder of the proof.

Now, for each source, consider verification that the available
previous data are independent of outcomes for the preceding subtests
that occurred before the last randomization but after the next to last
randomization. This follows from the fact that the composition of the
totality of observation values to which the last randomization is applied
does not change due to any permutations that can occur for these values
or any subsets of them.

Finally, let us consider any subtests that occur after the last
randomization. This situation is effectively the same as for the case
where none of the available previous data was obtained by randomization.
That is, for each source, a permutation that can occur for any of these
subtests corresponds to a subclass of the permutations for the values of
the available previous data. The statistic has the same value for all
the permutations of all such subclasses.

TEST DESCRIPTIONS

The observations, univariate and independent, are obtained in
consecutive groups whose compositions can be different. Let

G = maximum number of groups obtainable

9

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

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

$i(k)$ = group designation such that, after use of group $i(k) - 1$ but before use of group $i(k)$, k-th randomization selection is made from the available previous data ($k = 1,\ldots$; subject to $i(k) \leq M$), where $i(0) = 0$

$c_j(k)$ = number of observations from j-th source chosen at k-th randomization selection, with $c_j(k) \geq 1$, $c(0) = 0$, and $c_j(k)$ less than the number of past observations from the j-th source available for the selection.

N_{ij} = number of available previous observations from the j-th source when the i-th group of observations is obtained, with $N_{1j} = 0$.

$$= c_j(k-1) + \sum_{w=i(k-1)}^{i} n_{wj}, \qquad \text{for } i(k-1) \leq i < i(k)$$

S_i = statistic for the subtest where the i-th group of observations is first used (could represent two statistics when corresponding subtest is two-sided)

T_i = subtest that is based on S_i

α_i = significance level of T_i, $(0 < \alpha_i < 1)$.

In S_i, for $i \geq 2$, the N_{ij} available previous observations from source j occur symmetrically. That is, for the other data of S_i fixed, the value

10

S_i is the same for all possible permutations of the values for these N_{ij} observations.

The permutation models used for the various values of i and j are discussed next. For given i, the same kind of permutation model is used for all the values of j. Hence, consideration of an arbitrary but fixed value of j, and all values of i, is sufficient.

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

For each i, the overall permutation model is a combination of separate and independent use of the permutation models for the values of j. Thus, there are

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

possible ways of assigning the possible values to the corresponding observations. A value for S_i, not necessarily unique, occurs for each of these W_i ways.

General development of exact subtests by use of S_i and the overall

permutation model is discussed next. First, the W_i values for S_i are ordered according to increasing value (arbitrary but definite order for a set of tied values) and location in this ordering of the observed value for S_i is considered. Subtest T_i is one-sided upper-tail when significance occurs if an 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 occurs if and only if the observed S_i equals or exceeds at most $\alpha_i W_i$ values of the ordering. Now, consider two-sided subtests (which are not used very much for two-way ANOVA). Significance occurs if and only if either the observed S_i equals or is less than at most $\alpha_i' W_i$ of the values in the ordering, or the observed S_i equals or exceeds at most $(\alpha_i - \alpha_i')W_i$ of the values in the ordering, where $0 < \alpha_i' < \alpha_i$. This T_i has null probability α_i' for the upper tail and null probability $\alpha_i - \alpha_i'$ for the lower tail. The null values α_i, α_i', $\alpha_i - \alpha_i'$ must, of course, be attainable for these one-sided and two-sided tests. This implies that $\alpha_i W_i$ and $\alpha_i' W_i$ are integers.

Some two-sided subtests could use two statistics for S_i. One statistic furnishes a one-sided upper-tail test with significance level α_i' and the other a one-sided lower-tail subtest with significance level $\alpha_i - \alpha_i'$. The two-sided subtest is significant if and only if one of these one-sided tests is significant. A restriction on the one-sided tests is that they cannot both be significant.

Frequently, not all of the W_i ways of assigning the values to corresponding observation need to be considered separately. For example, the

12

interest is in subsets of these ways and S_i is such that its value is the same for all the ways in a subset. For $i \geq 2$, a commonly encountered situation of this nature is that where the only interest is in division of the $n_{ij} + N_{ij}$ values into a set of size n_{ij} and a set of size N_{ij}. This can be done in

$$(n_{ij} + N_{ij})!/n_{ij}!N_{ij}!$$

different ways, all of which are equally likely when the null permutation model is used. Then, for any i at least equal to 2,

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

is the combined possible number of divisions (over all sources). The method of developing exact tests is the same as that outlined for the null permutation model. That is, each of the W_i' divisions provides a value for S_i (not necessarily unique). These W_i' numbers are ordered according to increasing value, etc.

Use of an exact subtest can require an exhorbitant level of work. Exceptions occur in the use of rank tests with suitable tabulation, in cases where the number of possible permutations (or divisions) is not overly large, and in use of S_i that are based on statistics with null distributions that are symmetrical about zero. Hence, the subtests used ordinarily have significance levels whose values are only approximately evaluated.

An approximate subtest occurs when the significance level is evaluated by an approximate method. As an example, the significance level may

13

be determined from the first few terms of an expansion and usable only when the amount of data is large enough. This implies restrictions on the values of the n_{ij} and $c_j(k)$. As another example, a subtest based on ranks is approximate when the midrank method is used for ties but the significance level is evaluated under the assumption that ties cannot happen. Some subtests that use the observation values directly (no conversion to ranks, etc.) are simultaneously approximate in two respects. These subtests, called robust by Box and Andersen (ref. 1), are approximate permutation tests and also approximately unconditional.

Another approach to performing a subtest is to do a separate testing for each source. Then, T_i is significant if and only if significance occurs for at least one source. A statistic S_i could be developed, but separate consideration of the testing for each source is more convenient. Thus, for each value of i, the two-way ANOVA situation is converted to M separate and independent one-way ANOVA situations. Essentially, the selection of a test statistic for each of these one-way ANOVA situations is the same as if an overall one-way ANOVA (that emphasizes the more recent data) were being conducted on the basis of the observations from the source considered. Advice on choosing statistics for limited-length sequential permutation tests in one-way ANOVA, with emphasis on the more recent data, is given in ref. 2.

The principal new consideration is that the testings for the sources must all have very small significance levels if the subtest significance level is to be acceptably small. The other important aspects of conducting T_i by a separate testing for each source are covered by the material of ref. 2.

14

Hence, only a small amount of consideration is devoted to properties for subtests of this nature.

Significance occurs for one type of overall test if and only if at least one of $T_2, \ldots, T_G$ is significant. The significance level for this overall test is

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

The value of α is approximate when some or all of $T_2, \ldots, T_G$ are approximate.

Subtest T_1, where the random sample hypothesis is investigated for each source and the data are from the first group, is also included for the other type of overall test. Significance occurs if and only if at least one of $T_1, \ldots, T_G$ is significant and

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

is the significance level. This value is approximate if at least one of $\alpha_1, \ldots, \alpha_G$ is approximate.

CHOICE OF SUBTESTS AND THEIR STATISTICS

Choice of the S_i, and their use, includes many considerations besides symmetrical use of the available previous data. Moreover, many forms for a subtest statistic can result in equivalent subtests. Ordinarily, the least complicated form of statistic is adopted for exact subtests but

forms with approximately determined null distributions of a convenient nature are used for approximate subtests.

One consideration is provided by the alternative hypotheses that are to be emphasized. Another consideration is limitations on G, the n_{ij}, and the N_{ij}. These restrictions can be important when nearly equal values are desired for the α_i and/or a small size is desired for α and/or approximate subtests are used. Also, use of subtests that are uncondititional can be desirable.

Selection of the i(k) and c_j(k) establishes, for each source, the relative emphasis placed on data in the various past groups. For source j, consider the fraction of the observations from past group v that are still being used, on the average, in statistic S_i, where k is determined from i(k-1) $\leq$ i $<$ i(k). When the subtest using S_i occurs, this fraction is 1 for i(k-1) $\leq$ v $<$ i and is

$$\prod_{u=U}^{k-1} [c_j(u)/N_{i(u)j}]$$

for i(U-1) $\leq$ v $<$ i(U), where 1 $\leq$ U $\leq$ k-1. The values of these fractions, as a function of v, furnish a basis for choosing the i(k) and c_j(k) so that, for each source, the desired relative emphasis is placed on past data.

The i(k) and the c_j(k) could be selected to that the available past data remains about the same or even decreases. Ordinarily, for given j, steadily increasing values of c_j(k) are desirable. In fact, use of c_j(k) such that $c_j(k)/N_{i(k)j}$ is near one, say 9/10 or more, can be

16

desirable. Even then, for i large and on the average, the fraction of the observations (in S_i) from any of the first few groups can be small.

First, let us examine the case where a separate testing is made for each source. Use α_{ij} to denote the significance level for the testing of the j-th source in the i-th group. The subtest significance level for the i-th group is

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

since the observations are independent. If a small α_i is desired, all the α_{ij} must be very small. For i = 1, however, α_{ij} is at least $1/n_{ij}!$ for a one-sided testing and at least $2/n_{ij}!$ for a two-sided testing. For i ≥ 2, the value of α_{ij} is at least $n_{ij}!N_{ij}!/(n_{ij}+N_{ij})!$ for a one-sided testing and at least twice this value for a two-sided testing. These are all sharp lower bounds for α_{ij}. The implication is that the n_{ij}, n_{2j}, and N_{2j} (perhaps also the n_{3j} and N_{3j}) should all be of at least moderate size, especially when M is not small. The additional considerations for the case of a separate testing for each source are outlined in ref. 2.

The remaining part of this section is devoted to the subtests that are not based on separate testings for the sources. A value of α_i as small as $1/W_i$ can occur for one-sided T_i and as small as $2/W_i$ for two-sided T_i. Often, however, when i ≥ 2, the smallest value for α_i is $1/W_i{}'$ for one-sided T_i and $2/W_i{}'$ for two-sided T_i.

Values of the n_{1j} that are of at least moderate size may be needed

17

if α_1 is to be as small as desired. When M is at least 3 or 4 and the $c_j(k)/N_{i(k)j}$ are not too small, the smallest possible values for the α_i with $i \geq 2$ are frequently small enough for applications. This is often the case even when the n_{ij} are as small as 5 or 6.

Given α, allowable sizes for the α_i tend to decrease as G becomes larger. On the other hand, for given α and desired magnitudes for the α_i, the value for G has an upper bound. However, when all α_i are required to be very small, the upper bound on G can be very large.

Suppose that the α_i are required to be small and very nearly equal. Also, for fixed j, small and equal values are desired for the n_{ij}. Then, a compromise may be needed wherein, for fixed j, the value of n_{1j} is much larger than the n_{ij} for $i \geq 2$ (which can be equal or almost equal). A similar case is that where nearly equal small values are desired for the α_i and, for fixed j, the n_{ij} are required to be equal and small. Here, a compromise may be necessary in which α_1, and perhaps α_2, are larger than the other α_i (which can be very nearly equal). Of course, special care can be needed when, for fixed j, the $c_j(k)$ have about the same value or tend to decrease.

The alternative hypotheses emphasized are discussed next. Separate consideration of the alternative hypothesis for each source can be helpful. Whether the average values of the observations from a source tend to be nonincreasing from group to group, with an increase for the new group, frequently is of interest. Whether the average values tend to be nonincreasing, with decrease for the new group, can also be of interest.

Alternative hypotheses where the groups can be arranged so that,

18

for every source, the average values follow a trend in the same direction
(with a change in average for at least one group of one source) are
emphasized for most of the subtests considered. Hence, coordination in
data use should occur with regard to the direction of the trends emphasized
for the sources. As an example, consider the case where nondecreasing
average values are emphasized for some sources and nonincreasing averages
for the remaining sources. Modification of the observations for the re-
maining sources through multiplication by -1 results in a data situation
wherein nondecreasing average values are emphasized for all the sources.

As another example, consider the case where two kinds of alternative
hypotheses are emphasized. For one kind, the combination of nondecreasing
average values for the sources of one set and nonincreasing averages for
the remaining set of sources is of interest. For the other kind, the
combination of nonincreasing average values for the given set and non-
decreasing averages for the set of remaining sources is of interest. Here,
multiplication of the observations for the remaining set by -1 furnishes
data such that, for the two alternatives emphasized, the groups can be
arranged so that the average values follow a trend in the same direction
for every source.

Now, consider selection of S_i, with the major interest in results
that have already been developed. Correspondence with two-way ANOVA
results which are stated in terms of blocks and treatment levels (as for
randomized block designs) can be obtained by using sources to represent
blocks and groups to represent the treatment levels. Under the null
hypothesis, the treatment levels are equivalent. That is, within each block

19

the observations have the same null distribution for all treatment levels. Correspondence with two-way ANOVA results stated in terms of rows and columns can be obtained by identifying sources with rows and groups with columns. Under the null hypothesis, the observations in a row all have the same distribution.

The requirement that all the observations are expressed in the same unit can occur for some S_i. This can be an important restriction, since the type of unit for a source frequently differs in a basic manner from the type of unit for at least one other source (time, distance, volume, etc.).

Many kinds of statistics that could be used as S_i can be developed. These include statistics using extreme observations. In fact, virtually any statistic which satisfies the requirement on symmetrical use of past data could be the basis for a subtest.

Actually, when the sources are not investigated separately, the S_i for which results have been developed are limited. Moreover, some of these S_i require that all the observations are expressed in the same unit. Also, some S_i explicitly consider replication, which occurs within each group for every source, and others do not. Some S_i provide unconditional subtests while others do not. In addition, conditions are imposed on the n_{ij} and N_{ij} for some S_i, such as, for fixed i, the n_{ij} have the same value for all j and the N_{ij} have the same value for all j.

Statistics that do not explicitly allow for replication can still be used for a subtest when, for fixed i, the n_{ij} are the same for all j.

The method is to coordinate the observations from different sources by the sequence order in which they are obtained. This can be used to divide the observations into sets such that each set contains a single observation from each source. The first set consists of the first observations from the sources, etc. These sets, each containing M observations, have a chronological order and are used as if they are the groups considered. When S_i is stated in terms of blocks and treatment levels, the sources represent blocks and the sets represent treatment levels. When S_i stated in terms of rows and columns, the sources are the rows and the sets are the columns.

SOME SPECIFIC SUBTESTS

This section is devoted to identification and discussion of several kinds of subtests. Nearly all of these subtests are of an approximate nature or are applied in an approximate manner, so that extra conditions are imposed on the n_{ij} and N_{ij}.

Considered first are some robust subtests for the randomized block design with one treatment. Here, all observations must be expressed in the same unit, the subtests are of an approximate nature, (also, approximately unconditional) and, for fixed i, and equality is required for the n_{ij} and for the N_{ij}. The results of Box and Andersen (ref. 1) yield subtests that do not explicitly allow for replication. These results, plus some restrictions on the n_{ij} and N_{ij} for their use, are also stated on pages 324-325 of ref. 3. The results of Wilk in ref. 4 provide

subtests that do explicitly allow for replication but assume that some

special conditions hold. These results, plus some restrictions on the

n_{ij} and N_{ij} for their application, are also given on pages 325-326 of

ref. 3.

Considered next are some unconditional subtests for the randomized

block design with one treatment. Occurrence of the permutation model is

sufficient for use of these unconditional results. The basis for a sub-

test is the formation of a suitable function of the observations separate-

ly for each source. Under the null hypothesis, these independent functions

have distributions that are symmetrical about zero, and a subtest is ob-

tained by use of an appropriate test for symmetry around zero. More

specifically, order statistics of the observed values of the functions

are used for a subtest. Subtests are obtainable from the results of ref.

5 in which all the observations are involved in the null hypothesis. Sub-

tests are also obtainable from the results on pages 383-385 of ref. 3,

where Case (I) and the situation of the e''' equal to zero is considered;

also, for fixed i, equal values are required for the n_{ij} and for the N_{ij}.

Most of these subtests have exactly determined significance levels.

One or more pairs of order statistics for the functions provide

the basis for a subtest (both order statistics of a pair can be

the same). Ordinarily, for meaningful interpretation, all observations

must be expressed in the same unit. This requirement is not needed, how-

ever, when a one-sided test is based on a single one of these order

statistics, or when a two-sided test consists of two nonoverlapping one-

sided tests each of which is based on a single order statistic. That is

22

use of the values of the functions as if they were all expressed in the same unit is meaningful for this case. This is possible because the outcomes for subtests of this nature are entirely determined by the signs of the values for the functions.

Now, consider subtests whose statistics can be directly and entirely expressed in terms of ranks. These subtests are unconditional and the observations from different sources need not be expressed in the same unit. Exact and approximate results are both considered. The exact subtests do not explicitly allow for replication and, for fixed i, require the n_{ij}, also the N_{ij}, to have the same value. Approximate subtests occur that explicitly allow for replication and do not require equal n_{ij} or N_{ij} for fixed i. Chapter 11 of ref. 3 contains several rank tests (only those with both factors fixed are of interest). Included are restrictions on the n_{ij} and N_{ij} for use of approximate results.

Finally, some subtests are considered that have median ANOVA as a basis. Observations from different sources need not be expressed in the same unit and the results are unconditional. Explicit allowance is made for replication and, for fixed i, equality is required for the n_{ij} and for the N_{ij}. As applied, the subtests considered are approximate. These results are stated on pages 553-555 of ref. 3 (case of both factors fixed), with conditions on the n_{ij} and N_{ij} included. Other results of ref. 3 that have a permutation basis and are categorical could, with suitable inter-pretation, often also be used.

REMARKS ON QUALITY CONTROL USE

First, consider some properties of groups (observations from two or more sources) and corresponding tests that occur for quality control investigations of the control chart type. The successive groups usually satisfy: (1) The number of observations from a source is small for all groups except possibly the first group. (2) All groups, except possibly the first group, have the same composition. (3) Within each group, the number of observations from a source is the same for all sources. Now, consider the successive tests, one for each group. These tests satisfy: (4) The significance levels are small and equal (or very nearly equal).

Suitable sequential permutation tests which satisfy (4) and apply to groups that satisfy (1) - (3) almost always can be developed. An additional property is that, for fixed i, the N_{ij} are equal. Also, for efficiency reasons, the $c_j(k)/N_{i(k)j}$ should be near unity (say, at least 9/10).

Development considerations for the case where a separate testing occurs for each source are discussed in ref. 2. Another consideration, which can be important, is the way α_i depends on the α_{ij}. The remaining discussion of this section is concerned with the other case, where separate testings for the sources do not occur (except for comments about the sizes of the n_{ij}).

From properties (1) - (3), the n_{1j} are equal. Also, for $i \geq 2$, the n_{ij} are all equal and their value (say, denoted by n_{2j}) is small. The value of n_{1j} need not be small but ordinarily is the smallest value for

which (4) is satisfied. When the total number of observations from each source is fixed, use of too small a value for n_{1j} can entail substantial loss of information. Usually, n_{1j} should be as small as possible and n_{2j} should be as large as possible, subject to satisfying (1) - (4). When suitable past data that could be used for the first group are available, however, n_{1j} should be as large as possible.

Restrictions on sizes of the n_{ij} and the N_{ij} that are imposed for use of approximate tests sometimes violate some of the desired properties. This could, however, occur for small i but not occur for all larger i. When nothing suitable is available for direct use, appropriate subdivision of the observations almost always results in a situation with usable subtests. This division is like that described for using subtests that do not explicitly allow for replication.

Often, not all of the observations for a quality control situation can be expressed in the same unit. This limits the class of eligible subtests. Rank subtests have desirable characteristics for quality control use but other results can be more desirable for special situations.

The general applicability of these sequential permutation tests implies that their use is most appropriate when there is little prior information about the populations providing the observations. Such situations are common in quality control. As more groups are obtained, however, information is accumulated. Hence, the subtests tend to become more efficient, although an efficiency plateau should be reached before long. One motivation for emphasizing the more recent data is that these data are more pertinent and also little efficiency is lost by not using

all of the previous data.

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, Always Applicable Sequential Randomization Tests for One-Way ANOVA that Emphasize the More Recent Data, Themis Report 83, Statistics Department, Southern Methodist University, October 1970.

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.