THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

ALWAYS APPLICABLE SEQUENTIAL RANDOMIZATION TESTS FOR ONE-WAY ANOVA THAT EMPHASIZE THE MORE RECENT DATA

by

John E. Walsh

Technical Report No. 83
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

INTRODUCTION AND DISCUSSION

Limited-length sequential significance tests are considered for a one-way analysis of variance in which the observations are univariate, independent, and, under the null hypothesis, from the same (arbitrary) distribution. The observations are obtained in sets of stated sizes, with a specified number of sets available. An overall test is made up of a succession of subtests. A new subtest occurs for each new set of observations (after the first set, and sometimes also for the first set). For each subtest, the null hypothesis requires that the observations of previous sets and the observations of the new set are all from the same (unknown) distribution. If desired, an initial subtest can also be included, to investigate whether the observations of the first set constitute a random sample. Significance occurs for an overall test if and only if at least one subtest is significant. The obtaining of new sets can be stopped the first time a subtest is significant (with a saving in time and expense). An overall test fails to be significant if and only if the maximum number of sets occur without significance for any subtest.

Development of subtests that take into consideration the data from all previous sets seems highly desirable. For many situations, however, data from more recently obtained sets should receive greater emphasis. That is, the pertinence of any given set of previous data decreases as the number of sets increases. The subtests of this paper have the property of emphasizing the more recent data.

The taking into consideration of the data used for the preceding subtests can cause difficulties in determination of the significance levels for the subtests (and the overall test). That is, allowance needs to be made for the conditional effect of the outcomes for previous subtests. These significance levels are most easily evaluated when the development is such that the outcomes for preceding subtests have no conditional effect on the significance level for a subtest. This property, and the property of emphasizing the more recent data, can be accomplished by a combined use of suitable kinds of statistics for applying the subtests, permutation models for the null case, and randomization for deciding which previous observations continue to be used.

In addition to satisfying the two properties, the permutation-randomization approach yields subtests that are generally applicable. Moreover, the usable test statistics include types that are appropriate for this kind of investigation. The allowable subtests can be one-sided or two-sided, have wide ranges of significance levels, and can be oriented toward many kinds of alternative hypotheses.

All previous observations are used in a subtest until a specified total number of observations are obtained (correspond to a stated number of sets). Following the subtest using the last of these sets, a specified number are chosen from this totality of observations by randomization (all possible ways equally likely). Under the null hypothesis, the observations selected are a random sample from the population yielding the totality from which they were chosen. The previous data for the next subtest consist of the observations chosen by randomization. The previous observations for the following subtest are the new set obtained at the immediately preceding step and those chosen by randomization, etc. These previous data, under the null hypothesis, constitute a random sample.

New sets are now taken until the totality of available observations (those from the new sets and those obtained by randomization) reaches a stated number. Then, following the subtest using the last of these sets, a specified number of observations are selected from this totality by randomization. The procedure just described is now repeated, with new randomizations, until significance occurs or the maximum number of sets is obtained.

Ordinarily, the number of observations selected by randomization continually increases. That is, the number selected for the second utilization of randomization exceeds the number for the first use, etc. However, this is not necessarily the case.

The permutation models used are considered next. Subtests in which the new set is the second or a following set are examined first. The data are the new set and the previous observations being used at this stage. For the null case, the possible values for this totality of observations (new and previous) are conditionally fixed at those which occurred for these observations. Probability occurs only in regard to a division of the totality of possible values into a set whose size is that of the new set and a set consisting of the remaining values. All possible divisions are equally likely under the null hypothesis.

The properties for divisions can be stated equivalently in terms of permutations of the values in an arbitrary but definite sequence order

This is the case, for example, when the previous observations occur exclusively in functions that do not include any new observations and also are symmetrical in the previous observations. A statistic of this nature is independent of the outcomes for the prior subtests (see the next section for verification).

The fixing of possible values at those observed implies that, in general, all subtests are of a conditional nature. However, some situations and statistics combinations are such that this conditional fixing of possible values is not needed. That is, the statistic has the same distribution for all values that can occur for the observations. This is the case when a statistic is based entirely on ranks of the observations and either the data are continuous or randomization is used to break ties. Two-sample statistics based on exceedances (perhaps using extreme values) are an example of statistics of this nature. In fact, many of the statistics used for the two-sample problem are of this nature.

The limited-length sequential permutation tests are especially useful for situations where little or nothing is known about the distribution(s) for the observations. The principal interest is in situations where a change in the distribution may occur and rapid recognition of this change is desired.

One application area is in quality control. The limitation on number of sets does not disqualify an overall test for quality control use. In fact, the usual kind of quality control charts furnish overall sequential tests that must have a limited number of steps if their significance levels are to be acceptably small (values are unity for an unlimited number of steps).

The control chart method of using very small subtest significance levels and small equal-sized sets can be adopted for quality control use of these sequential permutation tests. However, some complications due to the restricted nature of the possible significance levels for a subtest can affect the choice of the first one or two subtests. That is, for small set sizes, none of the possible significance levels may be of the magnitude desired for all subtests. Consequently, the size used for the first set may be much larger than the sizes of the other sets. Also, the first one or two subtests may be given the smallest significance levels that are possible for the set size(s) used.

The next section contains a verification of the independence between a subtest statistic, of the type considered, and the results for the prior subtests. This is followed by a section that contains a descriptive statement of these limited-length sequential permutation tests. The next to last section is devoted to considerations in the selection of statistics and subtests. Almost all of the statistics mentioned are based exclusively on ranks (including statistics that involve extreme observations). The final section contains some remarks about quality control uses.

INDEPENDENCE BETWEEN STATISTIC AND PRIOR SUBTESTS

Considered is a statistic that, for any fixed values of the new is independent of results for the subtests that happened prior to the last randomization that was used.

Finally, consider the subtest or subtests that occur after the last randomization used (if any). This situation is essentially the same as for the case where none of the previous data was obtained by randomization. That is, a permutation that can occur for any of these subtests corresponds to subclass of the permutations for the values of the previous data, and the statistic value is the same for all the permutations of all such subclasses.

DESCRIPTION OF TESTS

The univariate observations, which are assumed to be statistically independent, are obtained in consecutive sets whose sizes can be unequal. Let

M = maximum number of sets obtainable

i = designation index for i-th set obtained ($i=1,\ldots,M$).

n_i = number of observations in i-th set obtained. $i \geq 1$ and usually is at least 3 or 4

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

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

N_i = number of past observations being used when the i-th set of observations is obtained, with $N_1 = 0$.

$$= c(k-1) + \sum_{j=i(k-1)}^{i} n_j, \quad \text{for } i(k-1) \le i < i(k)$$

S_i = statistic for the subtest where the i-th set of observations is first used (may represent two statistics if associated subtest is two-sided)

T_i = the subtest that is based on S_i

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

For fixed values of the n_i new observations, S_i is symmetrical in the N_i previous observations that are being used.

The permutation models used depend on the value of i, $(i=1,\ldots,M)$. Always, however, the totality of $n_i + N_i$ values whose permutations are considered consist of the values of the n_i new observations and of the values of the N_i previous observations that are used. This totality of values is conditionally fixed at those which are observed.

For i = 1, chronological order provides the sequence positions on which permutations are based. All possible ways of assigning the n_1 values to the n_1 positions are equally likely under the null hypothesis. The order of the sequence positions provide a basis for alternative hypotheses.

For $i \ge 2$, and under the null hypothesis, all possible ways of assigning the $n_i + N_i$ values to sequence positions are equally likely.

Here, say, the last n_i of the $n_i + N_i$ sequence positions furnish a division into a set of size n_i and a set of size N_i. Statement in terms of divisions is more convenient, however, since many permutations provide the same division. In terms of divisions, all possible ways to divide the values into a set of size n_i and a set of size N_i have the same probability when the null hypothesis holds.

Now, consider development of permutation subtests that have exactly determined significance levels. For i = 1, there are $n_1!$ permutations and a value of S_1 (not necessarily unique) occurs for each of these permutations. Let these $n_1!$ numbers be ordered according to increasing value (an arbitrary ordering within a set of tied numbers) and consider the place in this ordering of the value actually observed for S_1. Subtest T_1 is one-sided upper-tail when significance occurs if and only if S_1 equals or is less than at most $\alpha_1 n_1!$ of the numbers in the ordering. T_1 is one-sided lower-tail when significance occurs if and only if the observed S_1 equals or exceeds at most $\alpha_1 n_1!$ of the numbers in the ordering. Now consider an exact two-sided subtest with null probability α_1' for the upper tail and $\alpha_1 - \alpha_1'$ for the lower tail, where $0 < \alpha_1' < \alpha_1$. Significance occurs for this subtest if and only if either the observed S_1 equals or is less than at most $\alpha_1' n_1!$ of the numbers in the ordering, or the observed S_1 equals or exceeds at most $(\alpha_1 - \alpha_1')n_1!$ of the numbers in the ordering. Allowable α_1 and α_1' are such that $\alpha_1 n_1!$ and $\alpha_1' n_1!$ are integers. Also α_1, α_1', and $\alpha_1 - \alpha_1'$ are attainable significance levels for the corresponding one-sided subtests.

Next, consider exact performance of T_i for $i \geq 2$. Each of the $(n_i + N_i)!/n_i!N_i!$ divisions determines a value (not necessarily unique) for S_i. Let these $(n_i + N_i)!/n_i!N_i!$ numbers be ordered according to increasing value (arbitrary orderings within sets of tied numbers). 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 $\alpha_i(n_i + N_i)!/n_i!N_i!$ of the numbers 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 $\alpha_i(n_i + N_i)!/n_i!N_i!$ of the numbers in the ordering. An exact two-sided subtest with null probability α_i' in the upper tail and $\alpha_i - \alpha_i'$ in the lower tail, $0 < \alpha_i' < \alpha_i$, is considered next. For this T_i, significance occurs if and only if either the observed S_i equals or is less than at most $\alpha_i'(n_i + N_i)!/n_i!N_i!$ of the numbers in the ordering, or the observed S_i equals or exceeds at most $(\alpha_i - \alpha_i')(n_i + N_i)!/n_i!N_i!$ of the numbers in the ordering. The allowable α_i and α_i' are such that α_i, α_i', and $\alpha_i - \alpha_i'$ are achievable significance levels for the corresponding one-sided subtests.

Two forms of statistics could be used for S_i in some kinds of two-sided subtests. One form is used for a one-sided upper-tail test with significance level α_i'. The other form yields a separate ordering of $(n_i + N_i)!/n_i!N_i!$ numbers and is used for a one-sided lower-tail test with significance level $\alpha_i - \alpha_i'$. These one-sided tests are subject to the requirement that they are not significant simultaneously.

Unconditional exact subtests occur for situations where S_i can be based exclusively on ranks and also ties in ranks have zero probability (for example, the data are continuous or ties are eliminated by randomization). Many statistics that do not appear to be based exclusively on ranks can be expressed in that form (discussed in the next section).

Application of an exact subtest can entail an excessive amount of work unless the subtest is an appropriately tabulated rank test or $n_i + N_i$ is so small that the number of permutations or divisions is not overly large. Thus, subtests are often used whose significance levels are only approximately determined. This happens when a significance level is evaluated by an approximate procedure. As an example, the significance level value may be obtained from the first few terms of an expansion, and only usable when n_i and N_i are large enough. This would impose conditions on the n_i and $c(k)$. As another example, a test using ranks can be approximate because the midrank method is used for ties but the significance level evaluation assumes that ties do not occur. Some subtests that directly use the observation values (without conversion to ranks, etc.) are approximate in two respects. They are approximate permutation tests and also approximately unconditional. Box and Andersen call such tests robust (ref.1).

Significance occurs for one of the two kinds of overall tests if and only if at least one of $T_2,\ldots,T_M$ is significant. This overall test has significance level

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

The value of α is approximate when some or all of $\alpha_2,\ldots,\alpha_M$ are approximate.

Significance occurs for the other kind of overall test if and only if at least one of $T_1,\ldots,T_M$ is significant, and this overall test has

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

for its significance level. Here, α is approximate if at least one of $T_1,\ldots,T_M$ is approximate.

CONSIDERATIONS IN CHOICE OF STATISTICS AND SUBTESTS

The selection and use of the S_i involve many considerations in addition to the requirement that S_i is symmetrical in the previous observations that are used. Also, equivalent subtests can occur for many forms of the subtest statistic. Ordinarily, the least complicated form of statistic is used for exact subtests while forms that have approximately determined null distributions of a convenient nature are used for approximate subtests.

The alternative hypotheses to be emphasized furnish one consideraton. Also, limitations on the values of M, the n_i, and the N_i can be important when approximate subtests are used and/or nearly equal values are desired for the α_i and/or a small magnitude is desired for α. Moreover, subtests that are of an unconditional nature can be desired.

Choice of the i(k) and c(k) determines the relative emphasis placed on the various sets of past data. Consider the fraction of the observations from past set v that are still being used, on the average, in statistic S_i, where k is determined by $i(k-1) \le i < i(k)$. Given that the subtest based on S_i occurs, this fraction is 1 for $i(k-1) \le v < i$ and is

$$\prod_{u=U}^{k-1} [c(u)/N_{i(u)}], \quad \text{for} \quad i(U-1) \le v < i(U), \quad 1 \le U \le k - 1.$$

The values of these fractions provide a basis for selecting the i(k) and c(k) so that the desired relative emphasis is placed on sets of past data. Ordinarily, $i(k) - i(k-1) \ge 3$ but this is not necessarily the case.

The i(k) and c(k) could be chosen so that the amount of past data used stays about the same or even decreases. Usually, however, continually increasing values for c(k) are desirable. In fact, use of c(k) such that $c(k)/N_{i(k)}$ is near unity, say at least 9/10, can be desirable. Even then, on the average and for i large, the fraction of the observations (in S_i) from any of the first few sets can be small.

Some sharp lower bounds on the values of the α_i are examined next. For one sided subtests, $\alpha_1 \ge 1/n_1!$ and for two-sided subtests $\alpha_1 \ge 2/n_1!$. When $i \ge 2$,

$$\alpha_i \ge n_i!N_i!/(n_i+N_i)!$$

for one-sided subtests and is at least twice this value for two-sided subtests.

Now, consider some relationships among the α_i, the n_i, the N_i, α, and M. When a small value is desired for α, the value of n_1 and perhaps also that of n_2 should not be overly small. When there is a positive lower limit on the values of the α_i and the value of α is given, the allowable values for M have an upper limit. However, this upper limit is very large when the α_i are all very small.

Suppose that very nearly equal values are desired for the α_i and also small values are desired for the n_i. Then, a compromise may be needed in which n_1 and perhaps n_2 are much larger than the other n_i. Larger values for n_1 and n_2 may also be needed when the first one or two subtests are approximate. When the n_i must be equal and small, also very nearly equal values are desired for the α_i, a compromise may be needed. That is, the significance levels for the first one or two subtests are definitely larger than those for the remaining subtests.

Selection of S_i is relatively easy when α_i is to have its smallest possible value. Then significance occurs for a one-sided subtest if and only if an identified (from the alternative emphasized) permutation, or division, occurs for the $n_i + N_i$ values. Similarly, significance occurs for a two-sided subtest if and only if one of two identified permutations, or divisions, occurs. Ordinarily, consideration of the alternative emphasized readily identifies the permutation(s) or division(s) used for this purpose. Any S_i yielding a subtest with this property is satisfactory.

Selection of S_i under general circumstances is considered next. Many statistics that could be used as S_i have been developed. Moreover, general methods that yield statistics which could be S_i have been developed. For $i = 1$, a moderately comprehensive listing of basic results is given in Chapter 5 of ref. 2. All the univariate results are usable and, except for the first test on page 76, are unconditional (when ties are broken by randomization). For $i \geq 2$, a rather thorough listing is given in Chapter 2 of ref. 3. All nonsequential tests based on univariate observations are usable and all of these results are unconditional (although the robust tests on pages 126-127 are only approximately unconditional).

Many of the categorical data tests in Chapter 3 that have a permutation basis can also be used when the data are converted to categorical form.

Finally, for $i \geq 2$, consider some examples that should indicate why the nonsequential univariate tests of Chapter 2 in ref. 3 are unconditional and use eligible S_i. Some notation is introduced first. Let $x(1),\ldots,x(n_i)$ denote the new observations while $y(1),\ldots,y(N_i)$ are the previous observations being used. Order statistics of the x's and y's are denoted by $x[1] \leq \ldots \leq x[n_i]$ and $y[1] \leq \ldots \leq y[N_i]$, respectively, while $r(1),\ldots,r(N_i)$ are the ranks received by the previous observations in a ranking of the totality of $n_i + N_i$ observations. Any ties that occur are broken by randomization.

Investigation of a location difference by use of a robust t-statistic is discussed first. Let

$$\bar{x} = \sum_{j=1}^{n_i} x(j)/n_i, \qquad \bar{y} = \sum_{j=1}^{N_i} y(j)/N_i,$$

$$s^2 = \sum_{j=1}^{n_i}[x(j)-\bar{x}]^2 + \sum_{j=1}^{N_i}[y(j)-\bar{y}]^2,$$

$$t = (\bar{x}-\bar{y})s^{-1}[n_iN_i(n_i+N_i-2)/(n_i+N_i)]^{1/2} \quad .$$

The distribution of t is approximately that of a t-statistic with $(n_i+N_i-2)d$ degrees of freedom, where d depends on the observations but has a fixed value for the permutation models used. A subtest based on t has an approximate significance level and is approximately unconditional.

The N_i previous observations occur symmetrically in t (are in two functions not involving the new data). These results were developed by Box and Andersen (ref. 1). Some conditions on allowable values for α_i, n_i, N_i are given on pages 126-127 in ref. 3.

Some subtests for location differences that use order statistics (perhaps extreme order statistics) are described next. The test statistic is of the form x[a] - y[b], and a different statistic is used for each tail of a two-sided test. These tests can be stated in terms of ranks and are unconditional. They are eligible for use since y[b] is a symmetrical function of the past data used for any given value of b. Properties of tests of this kind are considered, for example, on page 150 of ref. 3.

Lastly, consider a test for location that is directly based on ranks. This is commonly known as the Wilcoxon two-sample test or as the Mann-Whitney test. The statistic can be stated as

$$\sum_{j=1}^{N_i} r(j) - N_i(n_i+N_i+1)/2.$$

Besides being unconditional, this test is eligible for use since, for any fixed values of the new observations, the test statistic is symmetrical in the N_i previous observations. Properties of these tests are considered, for example, on page 61 of ref. 3.

COMMENTS ON QUALITY CONTROL USES

Two characteristics that seem to be customary for successive tests of a control chart nature are: (1) The sets of observations, except possibly the first, have the same small size. (2) The significance levels are equal (or very nearly equal) and very small. These characteristics can be satisfied for overall permutation tests that emphasize alternatives of interest.

Characteristic (2) is most readily satisfied when n_1 is large. However, having n_1 large can be undesirable when the total number of obtainable observations is fixed. Then, n_1 should almost always have the smallest value such that, in combination with the common value for the other n_i, characteristic (2) is satisfied. More specifically, subject to satisfying both characteristics, n_1 should be as small as possible and $n_2 = n_3 = \ldots$ should be as large as possible. On the other hand, when the observations for n_1 do not reduce the number available for the other n_i, the value of n_1 should be as large as possible. For example, this is the case when n_1 is based entirely on data obtained prior to the start of obtaining the successive sets of observations.

These sequential permutation tests are, of course, most suitable for quality control situations where very little is known about the distributions from which the observations are obtained. Information is accumulated, however, as more and more observations become available. This accumulation has direct influence in subtests when the $N_{i(k)}$ are increasing functions of k. Actually, for efficiency reasons, the $c(k)/N_{i(k)}$ should be near unity (say, at least 9/10).

The influence of increasing amounts of past data can be directly identified in subtests based on the robust t-statistic described in the preceding section. When the null hypothesis holds, and at least the first few moments exist for the distribution sampled, increases in N_i imply decreasing variation in the two functions of previous observations. That is, these functions become approximate constants as N_i becomes large. The only important statistical variation is then due to the new observations. A subtest becomes approximately the same as for the case where accurate null values are available for the population mean and standard deviation.

REFERENCES

1. G.E.P. Box and S. L. Andersen, "Permutation theory in the derivation of robust criteria and the study of departures from assumptions," Journal of the Royal Statistical Society, Series B, Vol. 17 (1955), pp. 1 - 34.
2. John E. Walsh, Handbook of Nonparametric Statistics, D. Van Nostrand Co., Inc., Princeton, N. J., 1962, 575 pp.
3. John E. Walsh, Handbook of Nonparametric Statistics, II, D. Van Nostrand Co., Inc., Princeton, N. J., 1965, 712 pp.