THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

GENERALLY APPLICABLE LIMITED-LENGTH SEQUENTIAL PERMUTATION TESTS FOR ONE-WAY ANOVA

by

John E. Walsh

Technical Report No. 80
Department of Statistics THEMIS Contract

September 10, 1970

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

DEPARTMENT OF STATISTICS
Southern Methodist University

GENERALLY APPLICABLE LIMITED-LENGTH SEQUENTIAL PERMUTATION TESTS FOR ONE-WAY ANOVA

John E. Walsh*

Southern Methodist University**

ABSTRACT

The data are independent observations that, under the null hypothesis, have the same unknown distribution (which can be arbitrary). Observations are obtained in sets of specified sizes, with a maximum total number available. An overall test is conducted as a succession of subtests and is significant when at least one subtest is significant. Exact significance levels are obtainable by use of appropriate permutation models and special kinds of subtest statistics. The subtests are constructed so that the significance level of each new subtest is independent of previous subtest results. The overall test is terminated when a significant subtest occurs (thus saving time and expense). Subtests are always included wherein the second and each following set is compared with the totality of previous observations to investigate whether they continue to be from the same population. If desired, an additional subtest can be included so that the first set is investigated for the random sample hypothesis. The alternative hypotheses emphasized are determined by the subtest statistics. Unconditional tests can be obtained when ranks are used. There are indications that, by suitable choice of the subtest statistics, these tests compare favorably with tests using the maximum number of observations. One application is for quality control and some possible uses 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 one-way analysis of variance in which the (univariate) observations are independent and, under the null hypothesis, are from the same (arbitrary) population. The observations are obtained in sets of specified sizes, with a stated maximum number of sets being available. An overall test consists of a succession of subtests and a new subtest occurs for each new set of observations (after the first set). The null hypothesis for each of these subtests asserts that the observations of the new set and the observations from previous sets are all from the same (unknown) population. Also, if desired, an initial subtest can be included to investigate whether the first set constitutes a random sample. An overall test is significant if and only if at least one of the subtests is significant. Thus, the obtaining of the sets can be discontinued (with a resultant saving in time and/or expense) the first time a subtest is significant. An overall test is not significant if and only if the maximum number of sets is obtained without significance for any of the subtests.

Development of subtests that use all the data obtained through the sequential step considered seems highly desirable but can cause difficulties in the determination of significance levels for the overall test (and the second and following subtests). These significance levels are most easily determined when the significance level for a subtest is not influenced by the conditional effect of the outcomes for the preceding subtests. This can be accomplished, for each subtest, by use of a permutation model for the null situation and use of a suitable type of statistic for applying the subtest.

In addition to accurate evaluation of significance levels, this permutation approach has the desirable feature of yielding tests that are always applicable to the independent observations. Moreover, the usable test statistics include types that are appropriate for investigating whether the observations of a new set continue to be from the same population as that yielding the observations of the previous sets. The permissible subtests can have wide ranges of significance levels, can be one-sided or two-sided, and can emphasize many kinds of alternative hypotheses.

The data for the usual subtest are the new set of observations and the totality of all previous observations. For the null case, the values of these observations can be considered fixed and probability considerations enter in only one aspect. This aspect concerns a division of all the observations (new and previous) into a set whose size is that of the new set and a set consisting of the remaining observations. Under the null hypothesis, all possible divisions are equally likely. This null property can be stated equivalently in terms of permutations of the positions of the observations in a sequence order. A subset consisting of the last positions in the sequence, with the number of these positions equal to the size of the new set of observations, provides the basis for a division (namely, the division that occurs for the permutation considered). Under the null hypothesis, all the possible permutations are equally likely. The subtests corresponding to the second and all further new sets use this type of permutation model.

If the random sample hypothesis is investigated for the first set of observations, the chronological sequence in which these observations are obtained should also be known. Again, for the null case, the observed values can be considered fixed. Under the null hypothesis, all permutations of the observations in a (chronological) sequence order are equally likely, and this is the only manner in which probability properties are considered.

Now, let us consider a way in which a subtest statistic can be constructed so that the significance level for this subtest is not affected by the results for preceding subtests. Here, the data always are a new set of observations and the (nonvacuous) totality of all previous observations. Suppose that, in the subtest statistic, the previous observations occur exclusively in one or more functions that do not include any observations of the new set. Moreover, each such function is symmetrical in the totality of the previous observations. That is, the function remains the same for all possible permutations of the identities (sequence positions) of the previous observations. Then, the value of such a function is the same for every permutation that could have (randomly) occurred for any preceding subtest. That is, each of these permutations corresponds to a subclass of the permutations for the totality of previous observations, but the function value is the same for all permutations of the subclass. Thus, the results for preceding subtests have no affect on the value for this type of subtest statistic, since the outcomes for subtests are directly determined by the corresponding permutations that randomly occur. Consequently, the significance level for a subtest using this type of statistic is not influenced by the results for the preceding subtests.

In general, the subtests are of a conditional nature, since the observations are considered to have values that are fixed at those which occurred. However, the subtests are unconditional if ranks (perhaps of the totality of new and previous observations) are used and the data are continuous or randomization is used to break ties (approximately unconditional if some other method, such as midranks, is used to break ties). Special cases of statistics using ranks are the Smirnov two-sample statistics and those based on exceedances (for example, using extreme values).

By suitable choice of the subtest statistics, and of their uses, an overall test should have a reasonably high efficiency (even for specialized alternative hypotheses) compared to "best" tests using all the obtainable subsets of data. That is, subtests that emphasize the specialized alternatives are virtually always available and the ability to use fewer observations, on the average, should at least partially offset advantages of "best" tests that require all the obtainable data. Moreover, an overall test is generally applicable while the "best" tests ordinarily are usable only for very special cases.

The limited-length sequential tests are useful for situations where a change in the population sampled may occur for the latter subsets of observations and immediate identification of when this has happened is important. They are especially useful for cases where the distribution(s) for the population(s) yielding the subsets is (are) unknown.

Quality control represents one application area, even though an overall test has a limited number of sequential steps. In fact, strictly speaking, quality control charts of the usual $\mu \pm K\sigma$ kind provide sequential tests with a limited number of steps. Otherwise, these quality control charts,which use equal-sized sets of observations, would yield overall tests with unit significance levels (since each step furnishes an independent subtest and the subtests have the same nonzero significance level). The control-chart procedure of using equal-sized sets and very small subtest significance levels often is also adopted for a sequential permutation test. However, due to the (customary) use of a small set size, and the discrete nature of the possible significance levels, the first few subtests may have larger significance levels than those which follow. Often, the first few subtests are given the smallest significance levels that are possible for them. Of course, the sequential permutation tests are most suitable for quality control situations where little is known about the distribution(s) from which the observations are obtained.

The following section contains a formal statement of the limited-length sequential permutation tests, including the relationship between the significance level for an overall test and the significance levels of its subtests. The next to last section contains a discussion on the choice of statistics and subtests. Included for consideration are t-statistics, statistics using extreme observations, and statistics directly using ranks. Some more detailed comments on quality control uses are given in the final section.

STATEMENT OF TESTS

Some notation is introduced first. The (univariate) observations, which are assumed to be statistically independent, are obtained in consecutive sets whose sizes can be unequal.

M = maximum number of sets that are obtainable

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

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

$N_i = \sum_{j=1}^{i-1} n_j$ = total number of observations in sets $j = 1,\ldots,i-1$, for $i \geq 2$, and $N_1 = 0$

S_i = statistic for the subtest where i-th set of observations is first used

T_i = subtest which is based on S_i

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

In S_i for $i \geq 2$, the N_i observations for sets preceding the i-th set occur exclusively in one or more functions that do not include data from the i-th set. Moreover, for each such function, the N_i observations occur in a symmetrical fashion.

Let us consider the permutation models that are used for the various values of i, (i=1,...,M). To perform permutation subtest T_i, the values of the totality of $N_i + n_i$ observations are fixed at the observed values. The case of $i \geq 2$ is described first. Under the null hypothesis, all possible ways of assigning identities to these $N_i + n_i$ values (equivalent to all possible ways of assigning them positions in a sequence of $N_i + n_i$ positions) are equally likely. The values in the last n_i sequence positions provide a set of the same size as the new set of observations and, for each permutation, represent a division into a set of size n_i and a set of the remaining N_i values. Since many permutations can result in the same division, statement in terms of divisions is more convenient. Stated equivalently (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 are equally likely under the null hypothesis.

For i = 1, the sequence positions used represent chronological order and, for the null situation, are unrelated to any probabilistic properties of the observations. All possible ways of assigning the n_1 fixed values to the n_1 positions are equally likely under the null hypothesis.

Now, consider use of S_1 to perform T_1 in an exact fashion. There are $n_1!$ permutations and each permutation determines a value (not necessarily unique) for S_1. Let these $n_1!$ numbers be arranged according to increasing value (arbitrary orderings within sets of ties) and consider the sequence position of the value that was actually observed for S_1.

When significance occurs if and only if the observed S_1 equals or is less than at most $\alpha_1 n_1!$ of the sequence values, T_1 is a one-sided upper-tail subtest. When significance occurs if and only if the observed S_1 equals or exceeds at most $\alpha_1 n_1!$ of the sequence values, T_1 is a one-sided lower-tail subtest. When significance occurs if and only if either the observed S_1 equals or is less than at most $\alpha_1' n_1!$, $(\alpha_1' < \alpha_1)$, of the sequence values or the observed S_1 equals or exceeds at most $(\alpha_1 - \alpha_1')n_1!$ of the sequence values, T_1 is a two-sided subtest (with null probability α_1' for the upper tail and null probability $\alpha_1 - \alpha_1'$ for the lower tail). The allowable α_1 and α_1' are such that $\alpha_1 n_1!$ and $\alpha_1' n_1!$ are integers.

Now, consider use of S_i to perform T_i in an exact fashion for $i \geq 2$. There are $(n_i + N_i)!/n_i!N_i!$ divisions and each division determines a value (not necessarily unique) for S_i. Let these $(n_i + N_i)!/n_i!N_i!$ numbers be arranged according to increasing value. 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 sequence values, T_i is a one-sided upper-tail subtest. 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 sequence values, T_i is a one-sided lower-tail subtest. When significance occurs if and only if either the observed S_i equals or is less than at most $\alpha_i'(n_i + N_i)!/n_i!N_i!$, $(\alpha_i' < \alpha_i)$, of the sequence values or equals or exceeds at most $(\alpha_i - \alpha_i') \cdot (n_i + N_i)!/n_i!N_i!$ of the sequence values, T_i is a two-sided subtest. The allowable α_i and α_i' are such that $\alpha_i(n_i + N_i)!/n_i!N_i!$ and $\alpha_i'(n_i + N_i)!/n_i!N_i!$ are integers.

In some cases involving two-sided tests, two forms of statistics could be used for S_i. One form would be used for a one-sided upper-tail test with significance level α_i' and the other for a one-sided lower-tail test with significance level $\alpha_i - \alpha_i'$, subject to the requirement that both one-sided tests cannot be significant simultaneously.

In general, these exact subtests are conditional, due to the fixing of the totality of observations at the values that actually occurred. However, an unconditional exact subtest is obtained for the case where S_i can be based exclusively on ranks and also ties in ranks occur with zero probability (for example, ties are eliminated by randomization or the data are continuous). Some statistics that may not appear to be based exclusively on ranks can be expressed in that form. For $i \geq 2$, the Smirnov statistics using the difference between the empirical cumulative distribution function (cdf) for the set of n_i new observations and the empirical cdf for the N_i previous observations are of this nature, as are the statistics based on runs, the Cramér-von Mises statistics, and statistics based on exceedances (for example, see Chapter 2 of ref. 3). For $i = 1$, statistics using runs above and below the median, statistics based on runs up and down, and statistics based on signs of differences (for example, see Chapter 5 of ref. 2) can be expressed exclusively in terms of ranks.

Except when the subtest is a suitably tabulated rank test, or

$n_i + N_i$ is rather small (so that the number of permutations or divisions is not excessive), application of an exact test can involve a prohibitive amount of effort. Also, use of randomization to break ties in ranks is sometimes considered to be undersirable and can require a moderate amount of extra effort. Consequently, approximate subtests are often used (significance levels approximately determined). This is always the case when an approximate procedure is used to evaluate a significance level. As an example, a subtest based on ranks may be approximate because ties among ranks are broken by some averaging process, such as use of midranks, but the significance level is determined as if ties do not occur. As another example, a significance level may be based on the first few terms of an expansion and is usable only when n_i and/or N_i are sufficiently large (which would impose restrictions on the sizes of n_1, n_2, etc.).

Some subtests can be simultaneously approximate in two respects. That is, they are approximate permutation tests and also approximate unconditional tests. Box and Andersen (ref. 1) use the terminology "robust" for such tests.

For one kind of overall test, significance occurs if and only if at least one of subtests $T_2, \ldots, T_M$ is significant. The significance level of this test is

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

due to the properties of the type of S_i that is used (see the preceding section). This value of α is approximate if some or all of the subtests are approximate.

For the other kind of overall test, the first set is also investigated (for the random sample hypothesis) and significance occurs if and only if at least one of subtests $T_1, \ldots, T_M$ is significant. This test has significance level

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

due to the properties of the S_i, with α approximate if at least one of $\alpha_1, \ldots, \alpha_M$ is approximate.

CHOICE OF STATISTICS AND SUBTESTS

The choice, and use, of the S_i involve many considerations besides the requirement imposed (for $i \geq 2$) on the N_i previous observations. The alternative hypotheses emphasized are one important consideration. Limitations on the sizes of M and $n_1, \ldots, n_M$ are a consideration when a small magnitude is desired for α, nearly equal values are desired for the α_i, and/or approximate subtests are used. Also, an equivalent subtest can occur for many forms of the subtest statistic, and choice of the most suitable form can be a consideration. Here, the least complicated form of statistic is ordinarily used for exact subtests while forms with approximately determined null distributions (of a convenient nature) are ordinarily used for approximate subtests. In addition, subtests that are of an unconditional nature can be desired.

Discussed first are some inequalities which call attention to restrictions on $M, n_1, \ldots, n_M$ that can be appropriate. The value of α_1 is at least $1/n_1!$ for one-sided tests and at least $2/n_1!$ for two-sided tests. Also, for $i \geq 2$,

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

for one-sided tests and α_i is at least double this value for two-sided tests. Moreover, for all M' such that $2 \leq M' \leq M$,

$$\alpha \geq 1 - \prod_{i=2}^{M'} (1 - \alpha_i)$$

for the overall test where only $T_2, \ldots, T_M$ are used, and

$$\alpha \geq 1 - \prod_{i=1}^{M''} (1 - \alpha_i),$$

$1 \leq M'' \leq M$, for the overall test where all of $T_1, \ldots, T_M$ are used.

If a small value is desired for α, these inequalities imply that n_1 and n_2 should not be too small. Also, unless the α_i are to have values that decrease fast enough (a case not considered here), the allowable values for M have an upper limit (for given α). However, M can be very large when the α_i are all very small. When small values are desired for the n_i and also nearly equal values are desired for the α_i, a compromise may be needed in which n_1 and perhaps n_2 are larger than the other n_i. Also, larger values may be needed for n_1 and n_2 when the first one or two subtests are approximate. If the n_i are required to be equal and small, and also nearly equal values are desired for the α_i, a compromise may be needed in which the significance levels for the first one or two subtests are larger than the α_i for the other subtests.

Choice of S_i is not so difficult when T_i is to have the smallest possible significance level. Then (given the alternative hypothesis emphasized), significance occurs for a one-sided test if and only if an identified permutation, or division, occurs for the $n_i + N_i$ values. Likewise, significance occurs for a two-sided test if an only if one of two identified permutations, or divisions, occurs. Any S_i that yields a subtest with this property is satisfactory. As an example, consider $i = 1$, unequal observations, and a two-tail test that emphasizes upward or downward trend in expected value (with respect to time). Then, significance occurs if and only if, when arranged chronologically, the observed values are all increasing or are all decreasing. As another example, consider $i \geq 2$, unequal observations, and a one-tail test that emphasizes larger expected value for the n_i new observations than for the N_i previous observations. Significance occurs if and only if all the new observations have values that exceed all the values of the previous observations.

Now, consider choice of the S_i under general circumstances. A large number of statistics that could be S_i, with various uses and properties, have been developed. Also, some general methods for development of statistics that are eligible to be S_i have been devised. A moderately thorough listing of basic results for $i = 1$ is given in Chapter 5 of ref. 2 (all univariate results are applicable) and for $i \geq 2$ in Chapter 2 of ref. 3 (all nonsequential univariate tests are applicable). Discussions of the basis and uses for the statistics considered are given in these references. Nearly all of these statistics yield subtests that are unconditional (when ties do not occur or they are broken, as needed, by suitable use of randomization). Specifically, only the optimum permutation tests on page 76 of ref. 2 are conditional, although the robust tests on pages 126-127 of ref. 3 are only approximately unconditional.

Finally, let us consider some examples of statistics and subtests. A type of development for T_1 that involves extreme order statistics (and does not occur in Chapter 5 of ref. 2) is given first. The data, in chronological sequence, are denoted by $x(1),\ldots,x(n_1)$, where $x(1)$ is the first observation obtained, etc. Extreme value considerations can be useful when a strong time-wise trend in expected value is the alternative emphasized. For example, consider a one-sided subtest for upward trend where significance occurs if and only if $x(n_1)$ is the largest observation, $x(n_1-1)$ is the next to largest observation,..., $x(n_1-k)$ is the k-th from the largest observation, with $k \leq n_1 - 1$. This subtest has significance level $[n_1(n_1 - 1)\ldots(n_1 - k)]^{-1}$ when the k + 1 largest observations are unequal (or ties among them have been broken by randomization). The significance level that is smallest possible for any T_1 occurs for the special case of $k = n_1 - 1$. Analogous considerations apply to emphasis of downward trend. Significance occurs for this one-sided subtest if and only if $x(n_1)$ is the smallest observation, $x(n_1 - 1)$ is the next to smallest observation, ..., $x(n_1 - k)$ is the k-th from the smallest observation. Two-sided tests are obtained as a combination of a one-sided test for upward trend and a one-sided test for downward trend, with significance if either one-sided test is significant (significance level is the sum of those for the two tests). These tests are unconditional, since they can be expressed in terms of ranks that do not involve ties.

Now, consider some test statistics for T_i with $i \geq 2$. These examples should indicate why the nonsequential univariate tests in Chapter 2 of ref. 3 are based on eligible S_i. Let $y(1),\ldots,y(N_i)$ denote the previous observations while $z(1),\ldots,z(n_i)$ are the new observations. Also, let $y[1] \leq \ldots \leq y[N_i]$ and $z[1] \leq \ldots \leq z[n_i]$ be order statistics of the y's and z's, respectively, while $R(1),\ldots,R(N_i)$ are the rank numbers (among $1,\ldots,n_i + N_i$) received by the previous observations in a ranking of the totality of the observations. Either ties do not occur or they are broken, as needed, by randomization.

Location subtests based on exceedances are considered first. The statistic used is of the form $y[u] - z[v]$, where one or both of $y[u]$ and $z[v]$ may be extreme order statistics. It is to be noted that $y[u]$ is a symmetrical function of the previous observations. Tests are based on the sign of $y[u] - z[v]$ and are unconditional (expressible in terms of the ranks for the totality of observations). Properties of tests based on this form of statistic are considered, for example, on page 150 of ref.2.

Subtests for location (more generally, stochastic relationship) can be obtained using the Wilcoxon-Mann-Whitney-Festinger statistic, which can be expressed as

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

This statistic is a symmetrical function of the previous observations and is expressed in terms of ranks (so that the tests are unconditional). Properties of tests based on this statistic are considered, for example, on page 61 of ref. 2.

Lastly, consider robust use of a t-statistic for investigating location. These results were developed by Box and Andersen in ref. 1. Let

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

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

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

Under the null hypothesis, the distribution of t is approximately that for a t-statistic with $(n_i + N_i - 2)d$ degrees of freedom, where d is determined by the observations in such a way that the totality of all observations occur in a symmetrical fashion (so that d has the same value for all permutation models that have been considered, including the one used for this test). The previous observations occur symmetrically in t (in $\bar{y}$ and in s). A subtest obtained by use of t is has an approximate significance level and also is approximately unconditional. Some restrictions on values for α_i, n_i, and N_i are stated on pages 126-127 of ref. 2. Robust t-tests tend to be very efficient when the new observations and the previous observations are from two normal populations that are the same except possibly in mean value. However, several rank tests (including the Wilcoxon-Mann-Whitney-Festinger test) also have high efficiencies when this normality situation occurs.

QUALITY CONTROL USES

Customarily, successive tests for quality control use have the two characteristics: (1) All sets of new observations, except possibly the first set, are of the same size and this set size is small. (2) The significance levels for the successive tests are equal (or almost equal) and very small. Overall permutation tests whose subtests have these characteristics and which emphasize alternatives of interest can nearly always be developed, at least for cases where a "small set size" does not require less than four or five observations in a set.

In some respects, n_1 should be made as large as possible. The second characteristic is most easily satisfied when n_1 is large. However, for a given total number of observations, substantial information can be lost by having n_1 large. Nearly always, n_1 should have the smallest value such that, in combination with the value desired for $n_2 = n_3 = \ldots$, the second characteristic is satisfied. Ordinarily, subject to satisfying the two characteristics and using an acceptably small amount of data, n_1 should be as small as possible and $n_2 = n_3 = \ldots$ should be as large as possible. Of course, if suitable past data are available, these observations could be used as the first set. Then, the larger n_1 the better for the ensuing quality control investigation of equal-sized sets of new observations.

The overall permutation tests are most appropriate for quality control situations where not much is already known about the populations yielding the observations. As more observations are obtained, information is accumulated. Thus, the subtest for the (i + 1)-th new set tends to be more efficient than that for the i-th new set of observations, although a plateau on efficiency should be reached before long. For example, consider use of the robust t-statistic described in the preceding section. As long as the null hypothesis holds, increases in N_i result in decreasing variation in the functions of the previous observations until they are approximate constants in the t-statistics. Then, for the case of normality, a subtest becomes approximately the same as for the situation where accurate null values are available for the population mean and standard deviation. That is, consider the usual $\mu \pm K\sigma$ control chart tests that are based on the assumption of normality and specified null values for μ and σ. Corre-ponding permutation subtests (equal-tail and same significance level) using the robust t-statistic become approximately equivalent to these control chart tests as i increases, for the special case where the normality assumption holds and accurate null values are used for μ and σ. However, the permutation subtests always have valid significance levels but the control chart tests may not be even roughly valid when the normality assumption is violated or the null values used for μ and σ are inaccurate.

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 Nonparatetric Statistics, II, D. Van Nostrand Co., Inc., Princeton, N. J., 1965, 712 pp.